Studies in Linguistics and Philosophy 95

Elin McCready
Katsuhiko Yabushita
Kei Yoshimoto *Editors*

Formal Approaches to Semantics and Pragmatics

Japanese and Beyond

Springer

Formal Approaches to Semantics and Pragmatics

Studies in Linguistics and Philosophy

For further volumes:
http://www.springer.com/series/6556

Elin McCready · Katsuhiko Yabushita
Kei Yoshimoto
Editors

Formal Approaches to Semantics and Pragmatics

Japanese and Beyond

Springer

Editors
Elin McCready
Department of English
Aoyama Gakuin University
Tokyo
Japan

Katsuhiko Yabushita
Department of English
Naruto University of Education
Naruto
Tokushima
Japan

Kei Yoshimoto
Center for the Advancement of Higher Education
Tohoku University
Sendai
Miyagi
Japan

ISSN 0924-4662 ISSN 2215-034X (electronic)
ISBN 978-94-017-8812-0 ISBN 978-94-017-8813-7 (eBook)
DOI 10.1007/978-94-017-8813-7
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2014937722

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

The original version of the book was revised: Volume Editor's name has been corrected to "Elin McCready". The correction to the book is available at DOI: 10.1007/978-94-017-8813-7_17

Contents

Contributors

Nicholas Asher IRIT, CNRS, Université Paul Sabatier, Toulouse, France

Daisuke Bekki Ochanomizu University, Graduate School of Humanities and Sciences, Tokyo, Bunkyo-ku, Japan; National Institute of Informatics, Tokyo, Chiyoda-ku, Japan; CREST, Japan Science and Technology Agency, Saitama, Kawaguchi, Japan

Alastair Butler PRESTO, Japan Science and Technology Agency, Center for the Advancement of Higher Education, Tohoku University, Miyagi, Japan

Gerhard Jäger Institute of Linguistics, University of Tübingen, Tübingen, Germany

Oleg Kiselyov University of Tsukuba, Tsukuba, Japan

Masahiro Kobayashi Education Center, Tottori University, Tottori, Japan

Emar Maier Department of Philosophy, University of Groningen, Groningen, The Netherlands

Moe Masuko Ochanomizu University, Graduate School of Humanities and Sciences, Tokyo, Bunkyo-ku, Japan; Research Fellow of the Japan Society for the Promotion of Science, Tokyo, Japan

Elin McCready Department of English, Aoyama Gakuin University, Tokyo, Japan

Hiroaki Nakamura International Education Center, Hirosaki University, Hirosaki, Aomori, Japan

Rick Nouwen Department of Language, Literature and Communication, Utrecht University, Utrecht, The Netherlands

Norihiro Ogata Department of Language and Information Science, School of Language and Culture, Osaka University, Osaka, Japan

David Y. Oshima Department of International Communication, Nagoya University, Nagoya, Japan

Katrin Schulz Institute for Logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, The Netherlands

Chung-chieh Shan School of Informatics and Computing, Indiana University, Bloomington, IN, USA

Satoru Suzuki Faculty of Arts and Sciences, Komazawa University, Tokyo, Japan

Robert van Rooij Institute for Logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, The Netherlands

Katsuhiko Yabushita Department of English, Naruto University of Education, Takashima, Naruto, Japan

Kei Yoshimoto Center for the Advancement of Higher Education and Graduate School of International Cultural Studies, Tohoku University, Sendai, Japan

Introduction

Elin McCready, Katsuhiko Yabushita, and Kei Yoshimoto

Abstract This chapter is an introduction to the volume. It briefly traces the history of the LENLS workshop and discusses some themes of the volume, as well as providing brief summaries of the papers.

Keywords Introduction · Summary

This volume is a collection of selected papers mostly presented during the first 5 years of the conference Logic and Engineering of Natural Language Semantics (LENLS), held annually in Japan since 2003; some of the papers are (revised versions of) papers presented at a session of the conference, while the others were specifically written for this volume in lieu of papers presented at the conference.

Before introducing the papers themselves, we will give a bit of background on the workshop itself. Logic and Engineering of Natural Language Semantics (LENLS) was originally envisaged by the late Norihiro "Norry" Ogata around the turn of the century, when he joined the faculty of Osaka University. Before LENLS had there been no regularly scheduled international workshops or conferences dedicated to natural-language formal semantics/pragmatics in Japan. Ogata and McCready later discussed the possibility of realizing such a workshop, and in the end Ogata organized the first instantiation in 2004. This instance of the workshop was held

E. McCready (✉)
Department of English, Aoyama Gakuin University, Tokyo, Japan
e-mail: mccready@cl.aoyama.ac.jp

K. Yabushita
Department of English, Naruto University of Education, Naruto, Japan
e-mail: yabuchan@naruto-u.ac.jp

K. Yoshimoto
Center for the Advancement of Higher Education and Graduate School of International Cultural Studies, Tohoku University, Sendai, Japan
e-mail: kei@compling.jp

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*, Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_1,

in Kanazawa, Japan, together with the annual meeting of the Japanese Society for Artificial Intelligence; this connection has remained through the first ten years of the workshop.

LENLS initially was one of the international workshops collocated with the Annual Conference of the Japanese Society for Artificial Intelligence (JSAI). Since 2009, it has been held as a workshop of the JSAI International Symposia on AI. When it started eight years ago in Kanazawa, it had only twelve participants. Since then, the workshop has developed into a reasonably well-known international academic meeting on formal semantics and pragmatics. At this point it seems fair to say that it is the highest profile such meeting held regularly in Asia; we hope that its success will encourage others to begin such workshops themselves to further the development of the semantics/pragmatics(/philosophy) community in Asia. Norry Ogata obviously played a crucial role in the success of LENLS until his untimely death in 2008. His positive and broad-minded attitudes towards scholarship and people have been inherited by the workshop (we hope). We wish Norry were with us now; this volume is dedicated to his memory.

Let us now turn to a brief discussion of the themes of the volume and the papers. The volume includes several papers relating to proof theory, substructural logics, and monads. The first of these is by Daisuke Bekki and Moe Masuko and is entitled 'Meta-Lambda Calculus: Syntax and Semantics'. As the name suggests, the meta-lambda calculus is a lambda calculus with terms and types for meta-level operations. The types represent judgements on base-level types; operations on these types correspond to functions relating judgements. The meta-lambda calculus has been presented by Bekki in a series of papers (see the chapter for references), but each of these presentations have their problems (as Bekki and Masuko note in their Sect. 1.2); the paper in the present volume provides a solution to these problems in the form of a new syntax and a categorial semantics for the meta-lambda calculus, together with an equational theory. Bekki and Masuko prove the soundness of the operations of the theory using these new formulations. Most of the paper is occupied with these tasks, making its contribution largely foundational. However, a lengthy appendix considers how monads have been used in linguistic analysis and shows that the meta-lambda calculus is well-suited to representing monads of the kind needed for dealing with natural language. It then proceeds to indicate how monads, in particular monads represented in the meta-lambda calculus, can be used to analyze three linguistic phenomena: non-determinism of the kind encountered in ambiguity and underspecification, contextual parameters including those standardly used in the analysis of demonstratives and indexicals (Kaplan 1989), and, finally, continuations, as utilized in the analysis of focus movement and inverse scope.

A second foundational paper is Norry Ogata's 'Towards Computational Non-Associative Lambek Lambda-Calculi for Formal Pragmatics'. This paper takes as a starting point resource logical views of Lambek calculi and the categorial semantics associated with them. Ogata notes that such calculi, in particular the non-associative Lambek (NLC) lambda calculus which is useful for modeling tree structures. However, this calculus disallows certain kinds of variable bindings which seem to be necessary for analyzing natural language semantics (and pragmatics). Ogata therefore generalizes this calculus to a class of calculi lying between linear and NLC calculi, in which associativity can be controlled, and in which bindable positions can be made available. He then enriches these calculi further with various monads, based on the work of Moggi (1991) and others (with substantial detail about the required background on categories and monads in an appendix); finally, the resulting systems are shown to allow treatments of focus, topicalization, word order, and discourse anaphora. This paper is the last paper Ogata wrote before his untimely death, and we are very pleased to have the opportunity to publish it in this volume.

The third paper on these issues is ‘Continuation Hierarchy and Quantifier Scope’ by Chung-Chieh Shan and Oleg Kiselyov. Shan and Kiselyov aim to give an account of scope ambiguities on the basis of a well-understood and relatively simple formalism, without resorting to standard measures such as nondeterminism (QR). The resulting account is directly compositional and has the advantage of keeping lexical entries for non-scope-taking terms simple. Shan and Kiselyov begin by introducing continuation-based approaches to natural language quantification, developing a series of fragments with more and more complex elements. By the end of this explication, they have set up a continuation-passing style (CPS) semantics for quantification, meaning that each expression receives as argument a function representing its context; this semantics is implemented in the direct style, so that terms which do not manipulate continuations need not reference them in their semantics. This is admirably simple. This setup enables Shan and Kiselyov to give a semantics capable of deriving inverse scope readings of quantifiers via the continuation hierarchy, which is well-studied in computer science. This hierarchy is constructed by iterating the change from standard lexical entries to CPS expressions. The transformation to CPS can be performed multiple times; doing so leaves continuation-insensitive terms equivalent to lower versions, but continuation-sensitive terms such as quantifiers exhibit a changed behavior. Shan and Kiselyov exploit this effect by making quantifiers polysemous, with denotations corresponding to distinct scoping behaviors. Once the choice of level is made, the semantic computation is entirely deterministic, a feature not present in earlier approaches. They then use versions of this semantics to analyze scope islands, wide-scope indefinites, and inverse linking.

The collection also includes several papers which take a game-theoretic approach to issues in linguistic analysis. These two papers take radically different angles on how game theory can be applied to linguistic issues.

The first is concerned with how game theory, and evolutionary game theory in particular, can be used to deepen our understanding of linguistic universals. This contribution is Gerhard Jaeger’s ‘What is a Universal? On the Explanatory Potential of Evolutionary Game Theory in Linguistics’. Jaeger begins by discussing possible explanations for linguistic universals, separating them into nativist and functionalist views. The former assume that the basis of universals is something internal to the human organism; the latter take their basis to lie in adaptive considerations on communication and thus fitness. Jaeger then claims that evolutionary game theory allows a unification of the two kinds of approaches, in that it is able to analyze behavior at the individual and social levels simultaneously. He proceeds to show how evolutionary game theory can (and has) been used to model language evolution. The basic idea is to consider the case of signaling games in an evolutionary setting. Here, it can be shown that (as usual) signaling games have multiple equilibria, but that most of the undesirable equilibria are not evolutionarily stable, meaning that from most initial settings the system will not alight in such a state. Still, equilibria which are (neutrally) stable but not evolutionarily stable are always available in the sense that they cannot be ruled out by initial settings. He concludes that it is exceedingly hard to derive true universals, though statistical universals are available given the right set of dynamics.

The other paper on game theory is Nicholas Asher's 'The Noncooperative Basis of Implicatures'. This paper takes a very different tack from that of Jaeger: the game theory used here is of a kind which takes players to be rational agents as opposed to (as it were) instruments of evolutionary processes. This is the usual perspective taken by scholars who implement Gricean reasoning in terms of game theory, but Asher points out a problem with the standard view: it explicitly relies on assumptions about cooperation which are not always realized. Implicature generation on the Gricean picture requires that reasoners assume that their interlocutors are being cooperative in their speech. What happens in contexts where it is clear that cooperation is not happening, such as in the courtroom, or when being asked for money, or trying to avoid a fight with a white lie? Here the Gricean picture does not clearly apply. Asher proposes a replacement. He begins with the observation that in question-answering contexts it is best to simply answer the question if doing so is not harmful to the answerer's interests; if it is harmful, one can at least try to be polite, which has the minimum benefit of not producing an interaction with negative consequences for 'face'. With this in place, conversational participants can engage in default reasoning about likely moves given an observed piece of semantic content. In particular, given full information about the game, one can conclude that one's interlocutor is maximizing payoffs by not answering, which will in turn lead to certain inferences: implicatures. This system then does not require cooperativity as a basic assumption, but instead can derive implicature generation from reasoning about the behavior of rational agents.

In 'Coordinating and Subordinating Binding Dependencies', Alastair Butler attempts to account for parallel pronominal binding dependencies observed for coordination and subordination. He suggests that the similarity derives from the same mechanism underlying the two types of dependency. By developing a formal system called Scope Control Theory, he successfully simulates the opening of a variable binding by quantification, its linking to the subject or object argument of a predicate, and the handover to the pronominal sequence.

The volume also contains a number of papers which analyze Japanese linguistic expressions or constructions or to propose an analysis for a linguistic phenomenon drawing crucially on Japanese data, which is to be expected given the origin and the location of LENLS.

In 'A Categorial Grammar Account of Information Packaging in Japanese', Hiroaki Nakamura applies a Categorial Grammar approach to truth-conditional effects brought about by the choice between the topic marker *wa* and the nominative case particle *ga* in Japanese. Based on Combinatory Categorial Grammar (e.g. Steedman 2000), a lexical semantics is specified for *wa* to produce a tripartite information structure, i.e., the topic operator, the restrictor, and the nuclear scope. A sentence-internal topic is also dealt with by extending the framework. He also suggests that the contrastive interpretation of sentence-internal topics is induced by its heavier complexity profile load obtained from the semantic trip of proof nets.

In ‘Floating Quantifiers in Japanese as Adverbial Anaphora’, Kei Yoshimoto and Masahiro Kobayashi propose a new perspective on floating quantifiers (FQ) in Japanese by treating them as adverbial phrases which stand in an anaphoric relation to their hosts as quantified NPs. By hypothesizing real-time, incremental processing of the sentence and its information structure, the authors give an account of the grammaticality of subject-object asymmetry in the FQ position that differs considerably depending on the context. They also give an explanation on a construction in which an FQ and its host stand in a whole-part relationship in a consistent manner with the majority of FQs.

David Oshima’s ‘On the Functions of the Japanese Particle *Yo* in Declaratives’ is concerned with the function of the Japanese sentence-final particle *yo*. He reviews three influential analyses and demonstrates via the introduction of some new data that they do not derive the full picture of the meaning and use of the particle, though each is shown to partially do so. He presents a novel analysis of two central uses of *yo*, i.e. the guide to action (Davis 2009) and correction (McCready 2008, 2009) uses and shows that they do not fully account for his new data.

In ‘What is Evidence in Natural Language?’, Elin McCready addresses a foundational issue in the semantic analysis of evidentials, the question of what evidence is, or, speaking more strictly, what evidence is at work in the use and understanding of natural-language evidentials. In order to characterize such evidence, McCready examined what is to be considered justification for the propositional content in an evidential under skeptical scenarios and Gettier cases based on data involving Japanese evidentials. On the basis of this data, he (tentatively) concluded that the relevant notion of evidence is essentially a *de se* ascription of an increase in the probability of the target on the basis of the putative evidence. While it will be seen in the future whether his characterization is ultimately correct, the paper at least may spark interest in this subject.

In ‘Japanese Reported Speech’, Emar Maier argued that the traditional approach to quoted speech, i.e. the rigid dichotomy of direct and indirect quotations is not satisfactory in the face of data from, e.g. Japanese, where there are instances of quoted speech some parts of which are directly quoted and the other parts of which are indirectly quoted–“mixed” examples. He proposed a unified approach that takes every instance of quoted speech to be an indirect quotation; direct quotation will be just a special case where all the parts are mixed quoted.

Some of the papers in this collection do not fall neatly into any of the above categories. This set of papers is quite various in character, focusing on logical issues or on other problems in the semantics-pragmatics interface.

In his paper ‘Measurement-Theoretic Foundations of Dynamic Epistemic Preference Logic’, Satoru Suzuki proposes a new version of Dynamic Epistemic Logic (DEPL). DEPL can deal with dynamic interactions between knowledge and preferences in decision making under certainty, risk, uncertainty, and ignorance. It can be put to wider use than other existing theories like Dynamic Epistemic Upgrade Logic by van Benthem and Fenrong Liu (2007), since, according to Suzuki, this system gives an account of knowledge/preference interactions only under certainty. The author proves the completeness and soundness of the logic and enhances it with measurement-theoretic semantics.

'A Question of Priority' by Robert van Rooij and Katrin Schulz is another reminder that the primitives in semantic theories are not determined a priori but (like any theory) can be somewhat arbitrary. In particular, they take up the cases of properties as sets of individuals versus features, worlds versus propositions, individuals versus properties, time-points versus events, preference versus choice, and natural-kinds versus similarity. They not only demonstrate that each pair is such that either member can be taken as primitive and the other can be constructed from it, but also show the constructions have something in common. They conclude the paper by indicating that a notion of "naturalness" can be used to determine which direction of construction is more intuitive or reasonable.

Rick Nouwen's paper 'A Note on the Projection of Appositives' works to explain the contrast between the wide-scope and narrow-scope interpretations of nominal appositives. He argues that appositives are open propositions whose subject is a pronoun anaphoric to the anchor. Adopting Schlenker's (2010a, b) flexible attachment of appositives, it is assumed that appositions can be attached to any node of propositional type dominating the anchor. While some linguistic data still remain unaccounted for, Nouwen emphasizes the need for a more fine-grained classification of appositives with indefinite anchors to investigate the heterogeneous relations of appositives to their anchors.

Finally, the paper 'A Modal Scalar-Presuppositional Analysis of *Only*' by Katsuhiko Yabushita is another addition to the already numerous analyses of the meaning of *only* in the literature, This paper is specifically concerned with a particular feature of *only*-sentences, namely the asymmetry between positive and negative *only*-sentences in the cancellability of the prejacent. The feature was originally noted and given an analysis by Ippolito's (2008), who crucially attributed a scalar presupposition to the meaning of *only*. Yabushita argues that the scalar presupposition involved is in fact not a presupposition *simpliciter*, but rather is one restricted to the speaker, i.e. a modal presupposition. Yabushita reaches this conclusion via an argument which works to show that Ippolitos proposed presupposition is not completely adequate and does not lead to a plausible analysis of the asymmetry in question. He concludes that the proper analysis involves a revision of Schulz and van Rooij's (2006) view of *only* which incorporates a modal scalar presupposition.

We hope that this collection is not just informative and useful for its readers. It should also give a picture of the state of formal semantics and pragmatics in Asia, or at least in Japan; we hope that the volume, not to mention the conference from which it springs, will play a role in the continued development of these fields in the region.

References

Davis, C. (2009). Decisions, dynamics and the Japanese particle *yo*. *Journal of Semantics*, *26*, 329–366.

Ippolito, M. (2008). On the meaning of *only*. *Journal of Semantics*, *25*, 45–91.

Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–566) (Manuscript version from 1977). London: Oxford University Press.

McCready, E. (2008). Evidentials, knowledge and belief. In Y. Nakayama (Ed.), *Proceedings of LENLS 5*.

McCready, E. (2009). Particles: dynamics vs. utility. In Y. Takubo, T. Kinuhata, S. Grzelak & K. Nagai (Eds.), *Japanese/Korean Linguistics 16*, (pp. 466–480). CSLI.

Moggi, E. (1991). Notions of computation and monads. *Information and Computation*, *93*(1), 55–92.

Schlenker, P. (2010a). Supplements within a unidimensional semantics I: scope. In *Proceedings of Amsterdam Colloquium 2009*.

Schlenker, P. (2010b). Supplements within a unidimensional semantics II: Epistemic status and projection. In *Proceedings of NELS 40*.

Schulz, K., & van Rooij, R. (2006). Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy*, *29*, 205–250.

Steedman, M. (2000). *The Syntactic Process*. Cambridge: MIT Press.

van Benthem, J., & Liu, F. (2007). Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logic*, *17*(2), 157–182.

The Non Cooperative Basis of Implicatures

Nicholas Asher

Abstract This chapter presents a model according to which implicatures, which are traditionally analyzed in terms of cooperative principles, remain rational in strongly non cooperative settings.

Keywords Implicatures · Politeness · Game theory · Gricean cooperativity · Pragmatics · Discourse structure

1 Introduction

According to (Grice 1975), conversation is a biproduct of rational behavior, to be analyzed in terms of beliefs, desires, and intentions. In addition, Grice makes specialized cognitive hypotheses about conversational agents—in particular that they are highly cooperative. Grice's conversational maxims of quantity quality and relevance encode this cooperativity in a highly informal fashion, but since the work of (Cohen and Perrault 1979; Allen and Litman 1987; Grosz and Sidner 1990; Lochbaum 1998) and others, researchers have formalized these principles in terms of BDI (belief, desire, intention) logics.

There are two problems with this sort of formalization. The first is that propositional attitudes like belief, desire and intention are *private* attitudes, not common knowledge or even part of the mutual beliefs of dialogue agents. The link between what agents say or dialogue content and their private beliefs, preferences and intentions is much less robust than what many Griceans and Neo-Griceans have postulated

---

Thanks to LACL 2012 reviewers and reviewers for the Springer volume for helpful comments. This research was supported by STAC, ERC Grant 269427

N. Asher (✉)
IRIT, CNRS, Université Paul Sabatier, Toulouse, France
e-mail: nicholas.asher@irit.fr

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*,
Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_2,

for content cooperative conversation. Any model of dialogue must infer information about mental states from the observed dialogue actions and vice versa. So we must *interpret* those actions—in other words, we must provide a representation of dialogue content and a procedure for constructing it during dialogue processing. The current mentalist approaches to dialogue content, couched within BDI logics, all equate dialogue interpretation with updating mental states: for instance interpreting an assertion that *p* and updating the model of the speaker's mental state to one that includes a belief in *p* are treated as equivalent. But they clearly are not equivalent in even in cooperative dialogue. If I am having a bad day, then my wife may say something to make me feel better even though she does not believe it.

With dialogue content separated from the agents' mental states, we need a term for what a speaker engages in when he or she makes a conversational move: following (Hamblin 1987), say that a speaker makes a *public commitment* to some content—exactly what content he commits to depends on the nature of the speech act he's performed. In fact, (Lascarides and Asher 2009) make speakers publicly commit to the *illocutionary effects* of their speech acts and not just to the locutionary content so as to accurately predict implicit agreement and denial.

In cooperative conversation, most of the time people say or commit to what they believe. So we can offer one informal precisification of some of Grice's maxims by appealing to defeasible generalizations like the following.

- Sincerity: Normally agents who commit to ϕ believe ϕ.
- Quantity: say as much as you can say to achieve conversational goals.
- Competence: Normally if B believes that A believes that ϕ, then B should believe that ϕ.
- Strong Competence: Normally if B believes that A doesn't believe ϕ, then B should not believe that ϕ.
- Sincerity about Intentions: Normally if A publicly commits to the intention that ϕ, then A intends that ϕ.
- Strong cooperativity: Normally if A publicly commits to the intention that ϕ, then B should intend that ϕ.

Defeasible rules link various sorts of speech acts to intentions, beliefs and actions of their agents; for instance, if an agent asks a question, then he normally intends to know the answer to it.

2 Implicatures and the Problem

Such rules provide the basis of an account of implicatures, and inter alia scalar implicatures. Implicatures are defeasible inferences that involve the following problem: under what conditions can one reasonably infer from a speaker's not committing to ϕ that he commits to $\neg\phi$? Consider (1).

(1) a. A: Did all of the students pass?
b. B: Some passed.

In (1) *A* does not commit to the claim that all of the students passed, and most speakers would reasonably infer that *A* in fact commits to the claim that not all the students passed. Here is a sketch of the kind of reasoning that one can adduce as a Gricean in favor of such an inference. Suppose a set of alternative moves, that the move chosen normally conforms to all the constraints above, and those that do not deviate from one of the constraints. Suppose also as given, either by discourse structure or by the lexicon, a set of alternatives for *some*, {some, all}. We can now sketch an informal derivation of the scalar implicature that B believes that not all the students passed.

- Sincerity: implies B believes his response to A's question. Competence implies that A should believe it.
- Strong Cooperativity: B wants A to know an answer to his question—that either all the students passed or they didn't.
- So B's response should provide A an answer. (rationality)
- He didn't say all the students passed, which would have implied an answer.
- Choosing the alternative would not have violated Cooperativity (since it clearly provides an answer), so it must violate Sincerity.
- So B doesn't believe the all the students passed. And by Strong Competence, A shouldn't believe it either.

Rather than fully formalize this reasoning in a particular nonmonotonic logic,[1] let's step back and consider what this kind of approach does and doesn't do. First it requires a strong form of cooperativity to derive scalar implicatures—that interlocutors defeasibly adopt each other's conversational goals and that speakers tell the truth. While it's not clear that Grice ever committed himself to something like Strong Cooperativity as I have formulated it, it isn't clear how we can derive scalar implicatures otherwise. Second, it doesn't account for why B provides an "over-answer" to the question; with the implicature, B's answer not only provides a direct answer to A's question but tells him more by picking out a subset of the worlds that constitute the direct, negative answer to the question. Neither these axioms nor any Gricean account of which I am aware provides an account of why didn't B just give a direct answer to A's question. It's clearly not a matter of the Gricean maxim of Quantity, since the over-answer provides more information and is longer and more complex than a simple "No" answer! Yet over answers are very common in dialogue; for instance in the Verbmobil corpus (Wahlster 2000), for instance, there is a far higher proportion of over answers than direct answers to questions. Are we following a maxim of being as informative as possible? If so, then people would never shut up (of course some people don't)!

[1] See Asher (2012) or Schulz (2007) for details.

None of these reflections show that the Gricean account of the implicature is wrong, only that it is incomplete. But the real problem is that when these defeasible generalizations don't apply, the account fails to generate any implicatures; the machinery is silent on what happens when these defeasible generalizations don't apply. Let me explain. Real conversations can have many purposes, not just information exchange. People talk to bargain, to bluff, to mislead, to show off or promote themselves, to put others down, to persuade others what they want them to do regardless of the facts. They often misdirect or conceal crucial information. In other words, conversation is often, even largely, non cooperative in the Gricean sense.

Consider the cross-examination in (2) of a defendant by a prosecutor, from Solan and Tiersma (2005) (thanks to Chris Potts for this example):

(2) a. Prosecutor: Do you have any bank accounts in Swiss banks, Mr. Bronston?
b. Bronston: No, sir.
c. Prosecutor: Have you ever?
d. Bronston: The company had an account there for about six months, in Zurich.

The locutionary content of (2d) is true. But Bronston succeeds in deflecting the prosecutor's enquiry by exploiting a misleading implicature, or what one might call a *misdirection*: (2d) implicates that Bronston never had any Swiss bank account and this is false.

Misdirections can happen outside the courtroom too. Dialogue (3) occurred in a context where Janet and Justin are a couple, Justin is the jealous type, and Valentino is Janet's former boyfriend (from Chris Potts and Matthew Stone (pc)).

(3) a. Justin: Have you been seeing Valentino this past week?
b. Janet: Valentino has mononucleosis.

Janet's response implicates that she hasn't seen Valentino, whereas in fact Valentino has mononucleosis but she has seen him.

Clearly, neither Janet nor Bronston are abiding by Gricean principles as I have formulated them; they're not trying to help their interlocutors achieve the intention behind their questions—to know an answer. They are not cooperative at the level of intentions, which is required to generate implicatures à la Grice. However, they *are* relying on their interlocutors to draw these implicatures. Why would Janet bring out a random fact about Valentino, unless she intended Justin to draw the implicature that she didn't see Valentino? Why would Bronston announce a random fact about his bank, unless he hoped the prosecutor would draw the implicature that *Bronston* didn't have a bank account and find an answer to his question?

This much Griceans can readily admit to. The Gricean principles are defaults that are believed by the interpreters of messages, and the implicatures that interpreters draw can be false while the at issue content of the speaker's contribution is true.

The real problem arises when we push the reasoning one step further: it is reasonable to assume the prosecutor in (2) doesn't believe that Bronston is abiding by principles like Strong Cooperativity, so the prosecutor *shouldn t* derive the implicature intended by Bronston. Nevertheless, the prosecutor *does* derive the implicature, because he takes Bronston's response in (2d) to answer his question. In fact, he used this response to convict Bronston of perjury. But on the Gricean account, our prosecutor appears to be irrational—mutatis mutandis for Justin in (3): he knows that the relevant defeasible generalizations for drawing the scalar implicature needed to make a response to a prior question an answer don't apply, and yet he draws the implicatures anyway.

Misdirection is quite different from another form of conversation that is known as *opting out*. Gricean maxims also don't apply when a speaker simply opts out of quite basic conversational requirements. Consider dialogue (4) (from Chris Potts (pc)):

(4) a. Reporter: On a different subject is there a reason that the Senator won't say whether or not someone else bought some suits for him?
b. Sheehan: Rachel, the Senator has reported every gift he has ever received.
c. Reporter: That wasn't my question, Cullen.
d. Sheehan: The Senator has reported every gift he has ever received.
e. We are not going to respond to unnamed sources on a blog.
f. Reporter: So Senator Coleman's friend has not bought these suits for him? Is that correct?
g. Sheehan: The Senator has reported every gift he has ever received. (Sheehan says "The Senator has reported every gift he has ever received" seven more times in two minutes. http://www.youtube.com/watch?v=VySnpLoaUrI)

This is different from misdirection. Sheehan's utterances cannot be interpreted as implying an answer, and so contrary to Bronston's utterance (2d) this *exposes* that Sheehan hasn't adopted the reporter's intention.

Dialogue (5) is another real life example of an 'opting out' move that happened the author in New York City:

(5) a. N: Excuse me. Could you tell me the time please?
b. B: Fuck you!

In opting out, the speaker doesn't intend for his or her interlocutors to assume any sort of cooperativity is in play. Opting out occurs when an answer to a question isn't provided, or when an appropriate response to another's speech act isn't provided. In misdirection the response is intended to thwart the asker's goals, though the response *appears* cooperative. In opting out, no cooperative response is given. Notice that opting out moves are thus a way of quickly ending the conversation; if you can't attach someone's response to the rest of the discourse context in a coherent fashion, then you're probably not going to continue talking to that person.

With respect to opting out, the Gricean account fares better. The defeasible generalization of Strong Cooperativity doesn't apply, and so Griceans predict that no implicature is drawn—which is what the facts show. However, misdirections like that in (2) pose severe problems for extant, Gricean accounts of scalar implicature that are based on Strong Cooperativity. To investigate this in detail, we need some some background. We need to set out a minimal level of cooperativity that distinguishes misdirection and normal cooperative conversation on the one hand, from opting out on the other. I call this level *rhetorical cooperativity*. Rhetorical cooperativity has to do with a cooperativity at level of speech acts. Some examples will clarify; when someone greets you, you greet him or her back or make at least some recognition of the greeting. When someone asks a question, you respond by either giving a direct answer to the question, an indirect answer, which relies on an implicature, or you say that you can't answer the question.

We can make this notion of rhetorical cooperativity precise by appealing to a theory of discourse structure like SDRT (Asher and Lascarides 2003).[2] Such theories postulate that a text or a conversation is coherent, just in case each contribution to the discourse can be linked to some other element via a relation that makes clear the contribution's rhetorical function in the conversation. A speaker is *rhetorically cooperative* if and only if her contribution to a conversation can be linked to the conversational context via a rhetorical relation. While previously many argued that the inference to a discourse relation was often a matter of extralinguistic competence (for example myself in Asher 1993), matters have changed somewhat with the advent of powerful machine learning methods that show one can go quite a long way towards reliably labelling discourse relations in text using only linguistic information (Subba and Eugenio 2009; duVerle and Prendinger 2009; Muller et al. 2012). It now seems that grammar encodes in a subtle way a lot of information about rhetorical structure. When conversational agents tailor their contributions so that the grammar allows their interlocutors to conclude a rhetorical connection, they are being rhetorically cooperative.

Let's now take a closer look at a particular discourse configuration that concerns our present examples. It is a matter of a question by A and some sort of response by B. SDRT postulates different sorts of rhetorical responses to questions. One is labelled *QAP*, or *Question-Answer-Pair*. *QAP*(π_1, π_2) entails K_{π_2} is a true direct answer to the question K_{π_1} according to the compositional semantics of questions and answers. So when Bronston answers *No, Sir* to the prosecutor's first question in (2a), the response would have been linked to the question via *QAP*.

(2) a. Prosecutor: Do you have any bank accounts in Swiss banks, Mr. Bronston?
b. Bronston: No, sir.

Another is called *IQAP* or *Indirect Question Answer Pair*. $IQAP(\pi_1, \pi_2)$ entails K_{π_2} defeasibly implies, via default rules that the questioner and respondent both believe, a direct answer to the question K_{π_1}. Moreover, *IQAP* entails that the answer is true.[3] This is the relation that holds between Bronston's response and the prosecutor's second question. Bronston's response implies a direct answer via a quantity implicature.

There are other ways of responding to questions. One way is with another question, which may be connected to the first question in a variety of ways (Asher and Lascarides 2003). One is using a question to get more details about what sort of response the first question requires, a sort of follow-up question, which in SDRT is called Q-elab.

[2] Griceans can think of SDRT roughly as a large-scale development of the principle of relevance.

[3] In SDRT terms, *IQAP* is right veridical.

(6) a. A: How do I solve this problem?
b. B: Do you know how to do derivatives?

From a discourse theory point of view, calculating the relevant scalar implicature here is part and parcel of calculating the discourse connection between B's contribution and the discourse context. The scalar implicature is required to link Bronston's answer in (2d) with IQAP; it's the scalar implicature that links (2d) to a direct answer to P's question in (2c). Without the scalar implicature, (2d) is no better a response to (2c) than some random assertion about anything. And because rhetorical cooperativity is a basic form needed for the conversation to continue, we will infer rhetorical cooperativity unless it's clearly at odds with the data. Thus B *banks on* P's interpreting his response as an IQAP.

We now come back to the real problem for Gricean accounts. In fact, we all attach Bronston's answer with IQAP, regardless of assumptions about cooperativity. The derivation of IQAP in SDRT is triggered simply by sentence mood, as can be gleaned from the axiom in SDRT's logic GL for computing discourse relations.

- SDRT's GL axiom for IQAP:

- $(\lambda :?(\alpha, \beta) \wedge int(\alpha)) > \lambda : IQAP(\alpha, \beta)$

In words the axiom says that if β is to be attached to α and α is in interrogative mood, then normally β attaches with *IQAP*. In other words, sentence mood alone triggers the inference to *IQAP*. But the soundness of the rule as explained in Asher and Lascarides (2003) and the quantity implicature the IQAP is based on in cases of misdirection rely on cooperativity principles as we saw above that are not sound in this scenario. Clearly, Bronston does not share the prosecutor's goal of finding out whether Bronston had an illegal bank account in Switzerland, and the prosecutor believes this. Probably the audience believes it too. But then how do we conclude IQAP? Are we all irrational? Or perhaps there is another type of derivation of the implicature given by IQAP.

There are several possible strategies to rescue the situation. First, Griceans can attempt to maintain that implicatures depend on Gricean maxims and strong cooperativity, but that the example under discussion poses no problem for the view. Here's how a Gricean might put it:

> Ps question establishes the immediate public goal of the conversation as being to decide whether B has ever had Swiss bank accounts or not. Bs contribution can be taken to decide the issue and thus to be cooperative in achieving this goal under the assumption that he answers cooperatively a more general question, namely, who, among the relevant alternatives, had such accounts, as far as B knows. By mentioning his company but not himself, B conversationally implicates that he did not have such accounts, which settles the initial question negatively. B can be taken to be publicly committed to the claim that he didnt have such accounts, a claim that he implicated but was not semantically entailed by what his utterance said. B behaves as if his decision to answer the higher level question and implicate the answer that decides the immediate question is driven by a desire to be even more informative than Ps question requires: B is offering information on who *did* have such accounts.

The key weak spot in this response is the "assumption that he answers cooperatively a more general question". Why on earth should we or P assume this sort of cooperativity? But without this cooperativity, no implicatures should be drawn. So this "solution" in fact labels us all as irrational. I suspect that this response is even worse because cooperativity of intentions is the only way a Gricean has of providing implicatures like the following relevant to the interpretation of direct answers like (2b).

- X can be taken as an answer to Y, and so interpreters take X as an answer to Y.

Without cooperativity of intentions for the Gricean, there aren't implicatures of any kind. So there isn't even rhetorical cooperativity. But clearly the facts show that there can be rhetorical cooperativity without cooperativity at the level of intentions.

One might argue that implicatures can arise from other sources besides cooperativity of intentions. One such source could be an external constraint on conversation like the oath to tell the truth, the whole truth and nothing but the truth in a courtroom, or the threat of perjury. These constraints haven't stopped witnesses from lying in court under oath, but they do make it more costly to lie and so provide grounds for supposing that speakers under oath are not lying. It's unclear how one can support the conclusions of scalar implicatures from the oath, however. The oath doesn't force the implicature, as far as I can see, unless the Gricean stipulates that nothing but the truth entails all scalar implicatures. This is far too rigid an interpretation considering what happens in actual conversation. If the oath did force all implicatures to be entailed, then the Supreme Court would have had no business overturning the conviction, which in this case it did. Thus empowering the oath would also vitiate the distinction between what is said and what is implicated, something dear to Griceans. Thinking for a bit about how to force scalar implicatures to hold, we might take "the whole truth" to introduce something like an exhaustivity operator into Bronston's response. If that's right, we have as an entailment that Bronston never had a bank account. That is, it follows on this move that Bronston in fact *said* he didn't have a bank account, not only that he implicated it. Entrapment would be an easy matter for a prosecutor if this is what the oath actually did! And without such an explicit argument about the oath, we haven't gotten anywhere. Furthermore, in the Justin and Janet example of a misleading implicature (3), Janet is under no obvious external constraint like an oath when talking to her boyfriend.

One could argue that implicatures rely on Gricean cooperativity but have become fossilized. We might try to account for them as an "evolutionary adaptation": over repeated interactions where cooperativity is present, implicatures become automatic and thus are calculated even when the conditions of cooperativity that validate the implicatures are not present. While this is an appealing possibility to some, it is not so easy to provide a formal framework in which this intuition is borne out. Asher et al. (2001) attempt to model strongly cooperative principles of the sort mentioned above using evolutionary game theory. They show, however, that strong Gricean cooperative principles do not form an evolutionarily stable strategy unless rather strong initial assumptions are made.

The same problem bedevils a Gricean analysis of the Janet and Justin misdirection example in (3). First, I summarize the assumptions for the misdirection example (3). Let's assume that Justin does not believe or is not confident that Janet shares his intentions to get a true and informative answer to his question. That is, we're in a non cooperative or strategic situation.

Now what are the facts?

1. Justin, and we, take Janet's response e as an indirect answer to his question. To do this Justin must engage in some non monotonic or defeasible reasoning to connect the response with the question. Note that this doesn't entail that Justin accepts the answer or finds it credible. We are interested just in what information he extracts from it.
2. Janet's not lying but she is trying to mislead, to get, say, Justin off her back. She is committed to only to the factual content of her claim; but as a competent conversationalist, she realizes that it is naturally interpreted as an indirect answer. She realizes that the interpretation of her response as an answer involves some defensible reasoning on Justin's part, and she has the option of denying that that reasoning was sound in the present case or that she was completely responsible for it.

For the Gricean, the problem is that without Cooperativity, Griceans have no way to run through the defeasible reasoning that turns the response into an indirect answer. So Janet's response in the actual context is no different than an assertion of some random fact.

The Gricean might suggest that the difference between an assertion of just any random fact and Janet's actual response is the counterfactual claim that had they been in a cooperative situation, Janet's response would have been an indirect answer—an assertion of some random fact would not. Nevertheless, it's hard to see what this counterfactual claim does for the interpretation of Janet's response in the actual context. Clearly one difference between the counterfactual context and the actual one is that Justin probably doesn't believe the indirect answer or may be wary of it, as you say. But that's a matter of credibility and belief about the information imparted, not a matter of what information is imparted by the response. But in both contexts, the same information is imparted; that's why Justin will be justified in being mad *in the actual context*, when he finds out that Janet has been seeing Valentino. Griceans have no way of explaining why this is the case. Justin knew or suspected he wasn't in a cooperative environment; it would be irrational on Gricean grounds to draw the implicature.

The conclusion: People have diverging interests in many cases; any time someone wants to bargain for a car, invest in stocks, play a competitive game, get his parents to do something for him, his interests are not aligned or may not be with those of his conversational partner or partners. Nevertheless, people do interact, draw implicatures and provide indirect answers all the time. In fact, indirect answering is a pervasive discourse move in such situations. Grice's model doesn't adapt well to such strategic situations, since the implicatures required to link a response to a question as an indirect answer are based on Cooperativity. But why try to force all of this into

Cooperativity? Cooperativity of intentions is a special case of a much more general conversational situation. I have argued that implicatures are inferred when beliefs about cooperativity at the level of intentions are lacking. The Gricean makes a false prediction if he takes implicatures to be generated solely by cooperative principles associated with the maxims. What emerges is the lack of any immediately plausible alternative foundation of implicatures. In the next sections, I will propose such an alternative foundation.

3 The Model

I propose to look at our interpretation of Bronston's response from the perspective of game theory. Saying and interpreting what is said are both actions. Assuming that conversationalists are rational, what they say and how they interpret what is said should follow as actions that maximize their interests given what they believe. Conversation involves moves that are calculated via an estimation of best return given what other participants say, and this is a natural setting for game theoretic analyses.

Game theory has had several applications in pragmatics (Parikh 1991, 2000, 2001; Benz et al. 2005; Franke 2008; Franke et al. 2009; van Rooy 2003; van Rooij 2004). Much of this literature uses the notion of a signaling game, which is a sequential (dynamic) game in which one player with a knowledge of the actual state sends a signal and the other player who has no knowledge of the state. The games I will examine here are different, though they involve two players in a sequential game. I assume that the meanings of all signals are fixed and thus that we have a more orthodox game of strategy involving a player that makes one kind of conversational move and another player that responds with another type of move. I will concentrate on an analysis of the payoffs for different conversational strategies. A crucial feature of the model is that payoffs are fixed, not by coordination on meanings or interpretations (as is the case in signaling games) but by effects of politeness, broadly speaking. I take the view in this chapter that an important aspect of language and linguistic usage is not directly related to truth conditional content but to relationships of power between conversational participants. According to Brown and Levinson (1978)'s strategic theory of politeness, language does not have the role merely to convey or ask for propositional content. Language also serves a second role in negotiating the relationships between speakers and hearers, in particular what they call their "positive" and "negative" face. Positive face involves an agent's reputation and image from the perspective of his interlocutors, while negative face involves the agent's "distance" from his interlocutors, his freedom from constraints imposed by them on his possible actions. While these terms aren't precisely defined, they define relatively intuitive dimensions of an agent's social status in a community. Face is the medium through which conversational participants recognize and negotiate their partner's potential status their needs and their autonomy.

Following Asher and Quinley (2011) and Quinley (2011), I use the notion of an exchange game, which is a formal model of two or more agents sending goods to

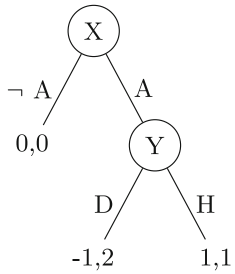


Fig. 1 Extensive form. Trust Games in Extensive Form: Player X has the option to Ask (*A*) Player Y for Help. Y can Help (*H*) or Defect (*D*)

		Player Y	
		H	*D*
Player X	*A*	1;1	-1;2
	¬*A*	0;0	0;0

Fig. 2 Normal form. Trust Games in Normal Form: Player X has the option to Ask (*A*) Player Y for Help. Y can Help (*H*) or Defect (*D*)

one another. Moves are dialogue speech acts, and information and face are the goods exchanged. Asher and Quinley (2011)'s model is asymmetric because the speaker places his fate in the hands of the hearer when making a request, or asking a question. Such conversational moves place one participant in the position of asking another to do something for him—this something is the *speech act related goal* or SARG of the speaker's move. All conversational moves have SARGs (Asher and Lascarides 2003). For instance, the SARG of someone's asking a question is normally to get an answer to the question and perhaps to get answers to follow up questions as well; other SARGs, however, are possible, as when for example a speaker asks a biased or rhetorical question (Asher and Reese 2005). However, to keep things simple here, I'll assume that the SARG of a question is the normal one of getting an answer.

The exchange game I use is a variant of a trust game (McCabe et al. 2003). Trust games depict a scenario where Player X has an initial option to defer to Player Y for a potentially larger payoff for both. Similar to the Prisoner's Dilemma, Player Y could defect on Player X and get a reward while X fares badly. For a one-shot game, this act of deference will not occur for a rational Player X. However, reputation and observation effects and the possibility of repeated games make deference rational (Asher and Quinley 2011) (Figs. 1 and 2).

The question is whether a conversation as I have conceived it is just a one shot two move game, one by each player, or is a conversational game more open ended with many possible continuations. I believe that conversations are not just one shot games, though this is seldom recognized in approaches that use signaling games. Conversational games are extended and dynamic, with an open ended sequence of conversational moves (though exactly the same move is almost never an option). Discourse theories like SDRT model this flexibility of conversation: one can always attach to the discourse structure with new information. There are natural endings to conversations, but they have to do with a mutual agreement on facts, an exchange or that a disagreement exists with no resolution. It's not clear when this mutual agreement will take place. As conversational games are not just one shot, but may involve several, and even many, actions by each player, reputation effects are *always* an issue in conversation. I propose to capitalize on this fact.

Conversational games also involve many possible moves, perhaps an in principle unbounded number, as one can almost always say anything in a conversation. However, discourse theories like SDRT provide us with a typology of conversational moves with different effects on content. These are the so called *discourse relations* or types of relational speech acts by which we attach one contribution to a conversation to the discourse context. I take these to constitute the moves or actions in the game. A strategy is a function from a finite sequence of such moves to another sequence of moves. Because games are in principle unbounded, I shall consider sub games in which utilities are assigned to (possibly) intermediate nodes in the game tree. To keep things simple, I will not introduce considerations of player types and assume the games are ones of perfect information.

While in principle any conversation may always be continued with further discourse moves, these moves have costs. They induce commitments by the speaker in the case of assertions; a speaker who asserts that p incurs the cost of potentially being challenged and having to defend his assertion. Not to do so leads to a loss of positive face. For questions and requests, the cost involves both a threat to the other's face (being too forward) and inviting a retaliatory attack on the speaker's reputation. Politeness theory following Brown and Levinson (1978) has studied the relative politeness of various types of speech acts, but these speech acts only characterize individual sentences. My proposal here is to look at the costs of relational speech acts, discourse moves that not only characterize the current utterance but affect the structure of the discourse context. A choice of a particular discourse move at stage m by participant i of an extensive game modeling a dialogue may make it very costly for a move of a certain type by participant j at $m + 1$, effectively ending the conversation or turning it in a new direction. The reason for this has to do with already incurred costs. Suppose a speaker i makes a move that involves a particular SARG with a certain cost. Costs of turns by i that continue to develop or help realize that SARG, once such a development is started but not completed, are intuitively lower than the cost of turns that incur a new SARG, ceteris paribus. This will be a key feature in accounting for implicatures.

3.1 Questions and Their Responses in the Model

The next thing to specify is how to model questions and their answers. I understand questions as a dynamic operation on an information state, following the outlines of SDRT. The input information state for a question is a set of sets of possibilities, and a question's semantic effect on this set of possibilities is to introduce further structure to this set of sets by regrouping the elements of those sets into possibly overlapping subsets, where each one of the subsets corresponds to a direct answer of the question. The linguistically encoded continuations are: eliminate some of the subsets by providing a direct answer or indirect answer (which implicates a direct answer), leave the structure as it is either by doing nothing or with a statement to the effect that the addressee is not in a position to provide any information, or ask a follow up question.

Let's now look at the costs of questions and their responses, in particular the face threatening or face saving nature of responses to questions. To make it concrete let us investigate the details of the conversation between Bronston (B) and the prosecutor (P). Let us assume that B does not wish to converse with P and does not, in particular, want to dwell on the topic of his bank accounts. If B gives an obvious non answer, he doesn't even commit to the question or address P's SARG to get an answer to his question. He affronts P's face, with potential retaliation and an unpleasant discourse move in subsequent turns, perhaps forcing him under oath to perjure himself or to admit damaging information. This would be rational if B were playing a one shot game (this is akin to the defect move in the Prisoner's Dilemma). But B is not playing a one shot game; if he defects, he will pay for it in the subsequent moves by P. B could also respond with a direct answer to P's question; in this case his response links to the question with the SDRT relation *Question Answer Pair* or *QAP*. If B responds with *QAP*, he does address P's SARG, at least as P has so far developed it. But B opens himself up to an explicit admission of guilt or explicit commitment to something perjurable. An *IQAP* answer that supplies additional information besides a direct answer, i.e. an IQAP that is an over answer, is more polite and increases the positive face of P. As such it is a lower cost move for B. More importantly, *IQAP* also increases the probability of no further negotiations on P's SARG, as the added information supplied in the *IQAP* anticipates follow up questions, answering them and so providing a more complete closure with respect to the questioner's SARG. This also increases the positive face of the interlocutor, making the move less costly. But, and importantly in this case, *IQAP* makes a continuation on the same topic by *P* more costly, because it forces him to introduce a new SARG or take the costly step of saying that his interlocutor hasn't answered the question (this is a direct attack on B's face and carries with it reputation effects). As it is in B's interest to avoid further questioning on this topic in particular, *IQAP* is the dominating strategy for him. If B answers *IQAP*, he avoids the potential face-threat and the politeness looks good to a judge and jury too. For P, *IQAP* is also an acceptable move by B to his question, and reputation effects make it less costly for him to accept it. Notice though that *QAP* is also an acceptable move for *P*: B gives him the information he was seeking and in a way that attends to P's reputation. B also gives additional information anticipating follow up questions, and this is information that could be of value to P. So it is in the interest of all to take a discourse move that is not a direct answer to be an indirect answer.

What about outside of the courtroom situation, say in the case of Janet and Justin in (3)? It seems that *IQAP* here too is a preferred move. With it Janet addresses Justin's SARG but also provides a justification for her indirect answer, for why she wouldn't have seen him. An argument for one's answer is a priori a way of making it more convincing, and of making the message more credible for Justin. This aspect of her *IQAP* reveals another reason why it might be preferable for her. A simple inspection of the trust game model for conversation implies dominance of IQAP over QAP in most situations, whether cooperative or non cooperative. It predicts that a question/IQAP strategy is the best strategy in a question response game. This is the prediction we were looking for.

3.2 Complex Structures in Discourse and Costs of Discourse Moves

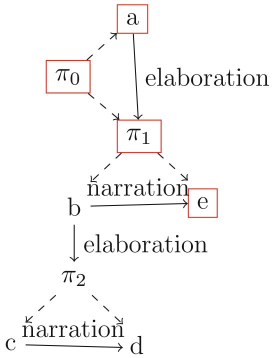
There is a close connection between SARG satisfaction and discourse structure in dialogue. Roughly, a move that satisfies a previously unsatisfied SARG forces a discourse "pop"; new material is no longer attached locally but to some higher constituent. This is a familiar principle for questions and their answers in theories that posit "questions under discussion" as a discourse structure organized around a stack of open questions (Roberts 1996; Ginzburg 1996); once the topmost question on the stack has been answered, the question is removed from the stack and the dialogue proceeds by answering the next question on the stack. SDRT has a much more general notion of discourse structure in which not only questions and their answers figure as constituents but also other assertions and relations between them. SDRT allows questions to be related to other questions via various discourse relations, and it allows assertions to attach to questions by other relations than simply answerhood, or in this case *IQAP*. Nevertheless, SDRT also incorporates a notion of discourse pop in its theory of where to attach new information and follows the intuition laid out in theories using questions under discussion in its account of attachment of new material to questions and their answers. Higher attachments incur new SARGs and in general incur higher costs, unless they are discourse closing moves or acknowledgments of a previous move or moves. By looking at discourse structure, we can examine in more detail how *IQAP*, other discourse moves and their possible continuations have different costs.

To give some more detail, I need to say more about discourse structure as it's described in a theory like SDRT. Discourse structures are graphs, where the nodes are discourse units and the arcs represent links between discourse units that are labelled with discourse relations (Asher 1993; Asher and Lascarides 2003). Discourse constituents may be elementary or complex. Elementary discourse units (EDUs), the atomic elements of a discourse structure, which correspond typically to clauses but also sub sentential constituents like appositions, non restrictive relative clauses inter alia (Afantenos et al. 2012), may be linked together via (one or more) discourse relations and form complex discourse units (CDUs) that are themselves arguments of discourse relations. CDUs in the ANNODIS corpus come in all sizes but the majority are relatively small (less than 10 EDUs in total (Nicholas et al. 2011)); in the few corpora of discourse annotated dialogues (Cadilhac et al. 2012), the CDUs that exist are very short, as typically they occur within one conversational turn.

In SDRT discourse graphs have a recursive structure with two sorts of edges in order to represent CDUs, one for the discourse relations and one to encode the relation between CDUs and their constituents. Consider the figure below for (7), an example familiar to those who have read about SDRT.

(7) a. Max had a great evening last night.
b. He had a great meal.
c. He ate salmon.
d. He devoured lots of cheese.
e. He then won a dancing competition.

Here is the discourse structure:



Discourse structure for texts, in particular the presence of CDUs, has interpretive effects. For example, the event described in (7e) comes after the events in (7b, c, d), as detailed in (Asher and Lascarides 2003). The presence of CDUs in a discourse structure is essential because they allow us to give a discourse relation scope over several EDUs, which is especially useful in cases where the relation cannot be "factored" or distributed over the constituents inside the CDU. This occurs for right arguments with relations like Explanation:

(8) James is sick. [He drank too much last night and he smoked too much.]

The part in brackets describes a CDU with two EDUs both of which contribute to an explanation of why James is sick. But we cannot distribute this explanation across the constituent EDUs; both EDUs contribute to cause James's sickness but neither one might be sufficient to cause the sickness on its own.

SDRT distinguishes between two types of discourse relations: subordinating and coordinating discourse relations. Relations like *Elaboration*, *Explanation* and *IQAP* are subordinating relations, while relations like Narration are coordinating. This gives SDRT graphs the 2 dimensional structure seen in the diagram above. If a CDU is closed off and an attachment happens either to the CDU itself or a discourse unit that dominates it (in the sense that there is some sequence of subordinating discourse relations from the unit to the CDU), then a discourse pop occurs. In this structure we see

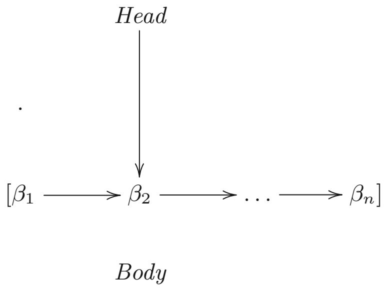


Fig. 3 A subordinate structure with a head

a discourse pop, even though no question answer structure is present: the EDU introduced by (7e) attaches to that introduced by (7b). This higher attachment is needed because the spatio temporal constraints introduced by the relations of Elaboration and Narration would entail that the dancing competition was part of the meal, if (7e) were attached to (7d), which is not plausible. In fact, this example points to a two way dependence between SARGs and discourse structure. Often SARGs for discourse moves are underspecified, especially with descriptive indicative sentences, though underspecification can also arise in the case of interrogatives or interrogatives (Asher and Lascarides 2003). The discourse pop mandated by considerations of plausibility here tells us that the speaker has finished with the SARG associated with (7d), which was to continue the Elaboration of the meal in (7b).

CDUs are also important for dialogue. The opening and closure of CDUs, or their boundaries, here too have to do with the SARGs of conversational turns. All discourse and dialogue moves, like asking a question for example, are defeasibly associated with a SARG–for instance, the asking of a question is associated with a goal of knowing the answer of the question. However, a SARG for a question may involve more information than just getting a direct answer to an explicit question; it may also include getting answers to certain follow up questions, demands for justification and so on. A SARG can develop and this development provides the grounds of a single local dialogue structure. Such a structure consists of a head or superordinate element, which gives rise to the general SARG together with a subordinate part, which develops the SARG. Such a structure is depicted in Fig. 3.

A question is typically an opening move in a CDU. The closure of that CDU will occur when an answer goal of that question is either satisfied or known not to be satisfiable and related follow up questions have similarly been answered or are known not to be answerable.

Let's now go back to our examples. An *IQAP* response to a question like that in (2) answers the question but also typically provides an "over answer". We've seen two kinds of over answers: one that anticipates follow up questions, another that provides a justification for the response, providing a priori grounds for rendering it more credible. It's more difficult to continue a SARG development once an indirect answer has been given. For instance (2e') is an example of an elaboration move on the answer.

(2e) Can you elaborate?

(2e) sounds silly, given the extended answer.

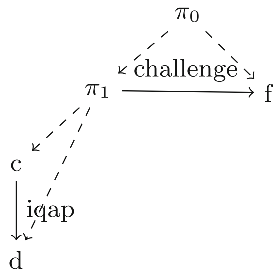
An *IAQP* answer thus provides a higher probability closing off of the local discourse structure: the SARG underlying the original question is satisfied and follow up questions are anticipated or justifications are provided. Continuations in this situation are more likely to be on the whole structure, which I hypothesize is a higher code move. For example, if P pursues this line of questioning in (2) it will most likely be on the whole structure—e.g. "challenge" to B's indirect answer. A challenge would be something like (2f) or (2f'):

(2f) Would you please answer the question, yes or no.

(2f') That wasn't my question. I'm not interested in whether the company had an account. I want to know whether you ever had an account.

The challenge takes as its left argument the response to the question, and the relation it bears to the question itself, thus the entire CDU. Such a challenge move is a higher cost move. The higher cost comes from a threat to face, generally perceived as very aggressive. IQAP thus raises the probability of a move by P to another topic or to exit the conversation. This is precisely what would suit B best.

Here's a picture of the IQAP scenario:



Once again, the situation is similar for the exchange between Justin and Janet. Justin can challenge the indirect answer, but it will be a higher cost move. Janet is rational in calculating that an IQAP will lead Justin to move to a different topic or to stop the conversation.

Let's now consider the alternative moves to *IQAP*, either *QAP* or ¬*Ans*. The *QAP* response is dispreferred for strategic reasons by B. It's also not a move in Janet's interest; either she admits to seeing Valentino or explicitly lies. A short answer QAP invites follow up questions, and so makes it easy for P to stay on this topic and get more information. For B or Janet, a non answer, which I label here with ¬ *Ans*, is a good one shot move, but in an extended game with further moves, it invites a low cost restatement of the question, since the SARG is not satisfied. It also invites a retaliation since it does not address the SARG of the questioner and so is attack on his positive face. The low cost move by P is depicted in Fig. 4.

The game tree in Fig. 5 for the exchange between P and B, where the costs of the different moves are motivated by the discussion above. I abstract from details

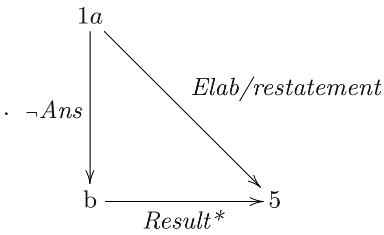


Fig. 4 The structure of a low cost move

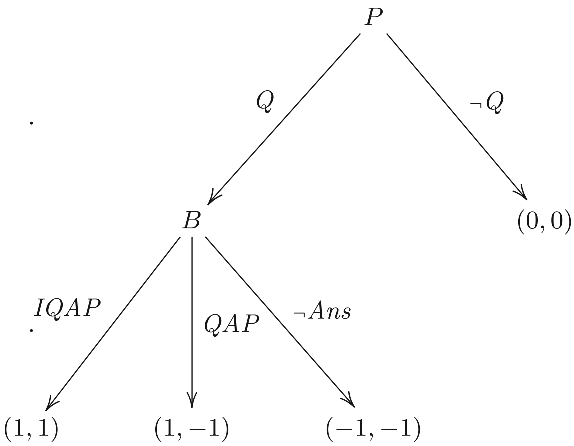


Fig. 5 The game tree for the Prosecutor and Bronston

and hypothesize three different discourse actions of B. The utilities provided, where P's utility is the first element of the pair and B's utility is the second element, are motivated by the preceding discussion:

The game over responses to questions is sub-tree compatible with a sequential trust game (McCabe et al. 2003) and its solution concept. While IQAP and Not Answer are equally rational for B in a one shot game, Not Answer puts P at a disadvantage. This disadvantage may lead to an unpleasant conversational turn, and reiteration of the question; the utilities on the ¬ *Ans* reflect this. In this case QAP is dispreferred by Bronston for reasons having to do with extra linguistic issues (the problem of explicit perjury).

Notice that P is indifferent between *QAP* and *IQAP*, since they both satisfy the SARG underlying the question. In the game that I have been using to model the situation, I have taken the contribution of each participant to be unambiguous. For the purposes of evaluating the rationality of *IQAP* responses, the assumption of non-ambiguity is harmless. However, a more detailed analysis of the moves by B reveals a sub game of an interpretive sort, i.e. a type of signaling game: B emits a signal, which we can assume has a fixed lexical and compositional semantics, and

P must infer from this semantic content what sort of move *P* is making. From the words themselves uttered by *B*, *P* has in fact two options: *IQAP* and ¬*Ans*. Though he is indifferent between *IQAP* and *QAP*, he may be suspicious of *B*'s desire to avoid the *QAP* move. The satisfaction of the questioner's SARG in the *IQAP* case depends on some defeasible reasoning about discourse attachment, reasoning which does not entail the answer. This reasoning is not guaranteed to be sound, and so the commitment by *B* to the answer might be challenged, if *P* takes the answer on board as the move. *B* might try to deny that he was answering *P*'s question. A full analysis of the signaling game as in Asher and Lascarides (2012) would introduce different player types for *B*, one where he is deceptive and one where he is not. Asher and Lascarides (2012) argue that *P* should be indifferent between QAP and IQAP only if IQAP remains an equilibrium in a larger game, in which facts about *B*'s player type that might distinguish between these IQAP and QAP are taken into account in the interpretation of *B*'s response to the question. But I will not go into the lengthy discussion that this engenders here, as it is not relevant for demonstrating the utility of *IQAP* moves.

4 Back to Implicatures

So far, I've developed a game theoretic model based on politeness and on assumptions of costs of continuations of certain discourse structures. I've shown that it's reasonable to suppose that certain kinds of responses to questions are preferred in non cooperative conversations—in effect over answers to questions and *IQAP* moves are preferred for strategic reasons. Thus, the account fills a gap in the Gricean account. But what about our puzzle about implicatures in non cooperative contexts?

Given the model, *IQAP* is strategically favored as a response. What I do is turn the problem on its head. For Griceans, it's only the presence of the implicature that allows us to treat the contribution in a misdirection as an *IQAP*. Here it's the choice of discourse relation itself, inferred on independent grounds and justified on the basis of game-theoretic and prudential grounds, that generates the implicature. The way to the implicature is relatively straightforward, once the discourse relation is fixed. When the move doesn't entail a direct answer, we have to engage in defeasible reasoning to get a direct answer. Sometimes this reasoning depends on a set of alternatives generated lexically or by the discourse context (see (Asher 2012) for a discussion of this issue and a proposal). The counterfactual reasoning goes as follows for *P*. *B* would know whether he had a bank account and so, given this presumption, would have said so; this would have been a natural and relevant issue to include in an *IQAP*. The *IQAP* move is designed to anticipate follow up questions, and a natural one in this case would be the question of whether Bronston himself had a bank account. In fact, it's the question that *P* asked! *P* can reasonably assume that since *B* doesn't want the questioning to go on, he says all that is relevant to *P*'s question—he is anticipating follow-up questions. Since *B* didn't say that he had a bank account, he commits to not having one, given the type of discourse turn. So in this case

$\neg$Commit(ϕ), *B*'s not committing to having a bank account, leads defeasibly to Commit$\neg\phi$, the commitment to the implicature that *B* didn't have a bank account. The scalar inference to the conclusion that Bronston doesn't have a bank account can be justified without appealing to any theses about cooperativity. Here I've substituted utility of *IQAP* and its semantics for cooperativity to get generate the implicature. The more general perspective, developed in Asher (2012) is that it is inferences about discourse structure that drives most if not all implicatures.

Other cases of misdirection have a similar analysis. Once again, we infer *IQAP* from the presence of an interrogative sentence mood to which is attached a contribution in indicative mood. The reasoning to *IQAP* is once again justified on prudential and game-theoretic grounds. The implicature generated by Janet's response to Justin in (3) is itself triggered by the search for a link between what is said and an answer to Justin's question; Janet's response in fact explains why she hasn't been seeing Valentino and this discourse configuration entails a negative answer to the question.

Nevertheless, misdirection is still possible. People who misdirect say the truth but exploit discourse structure to generate incorrect implicatures. The inference to *IQAP* remains consistent with the facts, despite the fact that the implicature generated is incorrect. And in so doing, the implicature remains as well. This contrasts with the Gricean approach, on which arguably the implicature simply doesn't arise. Our assumption of *IQAP*, though consistent and even reasonable, is fragile. The problem lies, as I intimated above, in the interpretation of *B*'s signal. Is it really an *IQAP* or is it in fact an evasion? Normal speakers do anticipate follow up questions, especially ones directly relevant to the issue. Were Bronston a normal uncooperative speaker, the inference to IQAP and the commitment to not having a bank account would follow. That is, *B* would in fact commit to not having a bank account (note that the question as to whether he commits to not having a bank account is a *very* different question from whether this information is credible, as I said before). But *B* in fact did claim that he did not *say* that he had a bank account. He was just giving some background information about the bank and his firm. So he argued that *P* in fact misinterpreted what he actually said; it wasn't a normal case. *B*'s argument, however, doesn't challenge the rationality of *IQAP* but rather the reasoning involved in the signaling game. Of course the prosecutor P is also at fault as Asher and Lascarides (2012) argue. He should have realized that Bronston's commitment here is less strong than one based just on compositional semantics; it relies on defeasible reasoning about discourse moves, which are ambiguously signaled. He should have realized that Bronston might try to weasel out of his commitment, and the attendant charge of perjury. In fact, this is what happened.

5 Conclusions

I have proposed in this chapter a model and an argument for supporting implicatures without Gricean assumptions about general beliefs in cooperativity. I've argued that certain discourse moves like over answers or *IQAP* moves are equilibrium points in a question response game, and that these over-answers generate the implicatures. The implicatures are drawn, even in the presence of misdirection. The model also explains why over answers are so frequent. The model makes clear on a hidden reputation effect that is constant in extended conversation. The foundation of this model rests on ideas from politeness theory, regimented within the framework of game theory. And thus face and reputation emerge as important factors in the evolution of discourse structure for conversation. Which perhaps points to a new and important role for expressive meaning.

References

Afantenos, S., Asher, N., Benamara, F., Bras, M., Fabre, C., Ho-dac, M., et al. (2012). An empirical resource for discovering cognitive principles of discourse organisation: the annodis corpus. In *Proceedings of LREC 2012*.

Allen, J., & Litman, D. (1987). A plan recognition model for subdialogues in conversations. *Cognitive Science*, *11*(2), 163–200.

Asher, N. (1993). *Reference to abstract objects in discourse*. Dordrecht: Kluwer Academic Publishers.

Asher, N. (2012). Implicatures and discourse structure. *Lingua 132*, 29–50. http://dx.doi.org/10.1016/j.lingua.2012.10.001

Asher, N., & Lascarides, A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Asher, N., & Lascarides, A. (2012). A cognitive model of conversation. In *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue (Seinedial), Paris, 2012*.

Asher, N., & Quinley, J. (2011). Begging questions, their answers and basic cooperativity. In *Proceedings of the 8th International Conference on Logic and Engineering of Natural Language Semantics (LENLS), Japan, 2011*.

Asher, N., & Reese, B. (2005). Negative bias in polar questions. In E. Maier, C. Bary, & J. Huitink (Eds.), *Proceedings of SuB9*, (pp. 30–43). http://www.ru.nl/ncs/sub9

Asher, N., Sher, I., & Williams, M. (2001). Game theoretic foundations for pragmatic defaults. In *Amsterdam Formal Semantics Colloquium, Amsterdam, December 2001*.

Asher, N., Venant, A., Muller, P., & Afantenos, S. D. (2011). Complex discourse units and their semantics. In *Contstraints in Discourse (CID 2011), Agay-Roches Rouges, France, 2011*.

Benz, A., Jäger, G., & van Rooij, R. (Eds.). (2005). *Game theory and pragmatics*. Basingstoke: Palgrave Macmillan.

Brown, P., & Levinson, S. C. (1978). *Politeness: Some universals and language usage*. Cambridge: Cambridge University Press.

Cadilhac, A., Asher, N., & Benamara, F. (2012). Annotating preferences in negotiation dialogues. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics*, (pp. 105–113).

Cohen, P. R., & Perrault, C. R. (1979). Elements of a plan-based theory of speech acts. *Cognitive Science*, *3*, 177–212.

duVerle, D., & Prendinger, H. (2009). A novel discourse parser based on support vector machine classification. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, August 2009*, (pp. 665–673). Suntec, Singapore: Association for Computational Linguistics. http://www.aclweb.org/anthology/P/P09/P09-1075

Franke, M. (2008). Meaning and inference in case of conflict. In K. Balogh (Ed.), *Proceedings of the 13th ESSLLI Student Session* (pp. 65–74).

Franke, M., de Jager, T., & van Rooij, R. (2009). Relevance in cooperation and conflict. *Journal of Logic and Language*.

Ginzburg, J. (1996). Dynamics and the semantics of dialogue. In Seligman, J., & Westerst ahl, D. (Eds.), *Logic, language and computation* (vol. 1).

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts*, (vol. 3, pp. 41–58). New York: Academic Press.

Grosz, B., & Sidner, C. (1990). Plans for discourse. In J. Morgan, P. R. Cohen, & M. Pollack (Eds.), *Intentions in communication*, (pp. 417–444). Cambridge: MIT Press.

Hamblin, C. (1987). *Imperatives*. New York: Blackwells.

Lascarides, A., & Asher, N. (2009). Agreement, disputes and commitment in dialogue. *Journal of Semantics*, *26*(2), 109–158.

Lochbaum, K. E. (1998). A collaborative planning model of intentional structure. *Computational Linguistics*, *24*(4), 525–572.

McCabe, K., Rigdon, M., & Smith, V. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior and Organization*, *52*(2), 267–275.

Muller, P., Afantenos, S., Denis, P., & Asher, N. (2012). Constrained decoding for text-level discourse parsing. In *COLING - 24th International Conference on Computational Linguistics, Mumbai, Inde, 2012*. http://hal.inria.fr/hal-00750611

Parikh, P. (1991). Communication and strategic inference. *Linguistics and Philosophy*, *14*(5), 473–514.

Parikh, P. (2000). Communication, meaning and interpretation. *Linguistics and Philosophy*, *25*, 185–212.

Parikh, P. (2001). *The use of language*. Stanford, California: CSLI Publications.

Quinley, J. (2011). Politeness and Trust Games, Student Papers Session, *Proceedings of ESSLLI 2011*.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. In J.H. Yoon, & A. Kathol (Eds.), *OSU Working Papers in Linguistics 49: Papers in Semantics*, 91–136. The Ohio State University Department of Linguistics.

Schulz, K. (2007). Minimal Models in Semantics and Pragmatics: Free Choice, Exhaustivity, and Conditionals (PhD thesis, University of Amsterdam).

Solan, L. M., & Tiersma, P. M. (2005). *Speaking of crime: The language of criminal justice*. Chicago, IL: University of Chicago Press.

Subba, R., & Di Eugenio, B. (2009). An effective discourse parser that uses rich linguistic information. In *Proceedings of HLT-NAACL* (pp. 566–574). ACL, 2009. http://www.aclweb.org/anthology/N/N09/N09-1064

van Rooij, R. (2004). Signalling games select horn strategies. *Linguistics and Philosophy*, *27*, 493–527.

van Rooy, R. (2003). Being polite is a handicap: towards a game theoretical analysis of polite linguistic behavior. In *TARK* (pp. 45–58).

Wahlster, W. (Ed.). (2000). *Verbmobil: Foundations of speech-to-speech translation*. Berlin: Springer.

Meta-Lambda Calculus and Linguistic Monads

Daisuke Bekki and Moe Masuko

Abstract *Meta-lambda calculus* (MLC) is a two-level typed lambda calculus with meta-level types and terms. MLC has been adopted in the analyses of natural language semantics and pragmatics by means of *monads* and *monadic translation* (Bekki 2009; Bekki and Asai 2010), however, the soundness of the equational theory in Bekki (2009) has not been fully proven with respect to the categorical semantics in Bekki (2009). In this article, we introduce a revised syntax and an equational theory of MLC with base-level/meta-level substitution and $\alpha/\beta/\eta$-conversions, and prove their soundness with respect to a revised categorical semantics of MLC.

Keywords Meta-lambda calculus · Type theoretic semantics · Category theory · Monads · Non-determinism · Ostension · Continuations

Our sincere thanks to the participants at LENLS5 and LENLS6, especially Elin McCready, Robert van Rooij, Kei Yoshimoto, Masahiko Sato, and the late Norry Ogata, and also to the participants at FLOPS2010, especially Olivier Danvy, Ian Zerney, Chung-chieh Shan, Oleg Kiselyov, Yukiyoshi Kameyama for their valuable comments. Our special thanks go to Kenichi Asai for many discussions in our research group. We also thank the anonymousss reviewers for their insightful comments. Daisuke Bekki is partially supported by a Grant-in-Aid for Young Scientists (A) (No. 22680013) from the Ministry of Education, Science, Sports and Culture

D. Bekki (✉) · M. Masuko
Ochanomizu University, Graduate School of Humanities and Sciences, 2-1-1 Ohtsuka, Tokyo, Bunkyo-ku 112-8610, Japan
e-mail: bekki@is.ocha.ac.jp

D. Bekki
National Institute of Informatics, 2-1-2 Hitotsubashi, Tokyo, Chiyoda-ku 101-8430, Japan

D. Bekki
CREST, Japan Science and Technology Agency, 4-1-8 Honcho, Saitama, Kawaguchi 332-0012, Japan

M. Masuko
Research Fellow of the Japan Society for the Promotion of Science, Tokyo, Japan
e-mail: masuko.moe@is.ocha.ac.jp

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*,
Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_3,

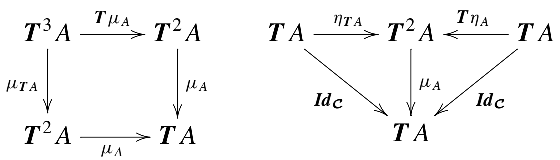
1 Introduction

1.1 Monads in Category Theory and Programming Language

The notion of monads originates in homological algebra and category theory: a monad in a category $\mathcal{C}$ is a triple $\langle \boldsymbol{T}, \eta, \mu \rangle$ that consists of a functor $\boldsymbol{T} : \mathcal{C} \longrightarrow \mathcal{C}$ and two natural transformations:

$$\eta : \boldsymbol{Id}_{\mathcal{C}} \stackrel{\cdot}{\longrightarrow} \boldsymbol{T}, \qquad \mu : \boldsymbol{T}^2 \stackrel{\cdot}{\longrightarrow} \boldsymbol{T}$$

such that the following diagrams commute for any object A in $\mathcal{C}$.



Lambek (1980) established categorical semantics of simply-typed lambda calculi (hereafter STLC), showing that STLC are equivalent to Cartesian closed categories (CCC), in which STLC terms are interpreted as morphisms.

These studies converged to the *monadic* categorical semantics of STLC in Moggi's seminal work (Moggi 1989), where each lambda term is interpreted as a morphism in the Kleisli category generated by a certain monad. This setting is intended to uniformly encapsulate "impure" aspects of functional programming languages, such as side-effects, exceptions and continuations, within the enhanced data types specified by the monad, and hide them within "pure" structures of STLC. The method requires, however, some tangled notions such as tensorial strength (Kock 1970) for the definition of lambda abstraction, evaluation and products.

This complexity motivated Wadler (1990) to propose a simplified model, known as *monad comprehension*, which generalizes the notion of list comprehension. Results from this study were incorporated into the programming language Haskell, and this showed that the monadic analyses can treat a wider range of computational concepts than those enumerated in Moggi (1989), such as state readers, array update, non-determinism, inputs/outputs, and even parsers and interpreters.

1.2 Monads in Linguistics

It was first suggested in Shan (2001) that the notion of monad can be imported to the field of natural language semantics, where various semantic, pragmatic, or computational phenomena such as non-determinism, focus, intensionality, variable binding, continuation and quantification, can be represented as monads, just as the "impure" aspects in programming languages. A few works such as Ogata (2008) and Unger (2011) have pursued this view recently.

The general idea is that a monad expands the data structure of semantic representations in such a way that certain phenomena become representable. Many theories in formal semantics have employed their own extended type-theoretical languages in order to explain certain linguistic phenomena, however, monads achieve this in a uniform way: for semantic representations, we use a simple language of STLC, except that it may contain *control operators* in the sense of Danvy and Filinski (1990). Those control operators are interpreted by translating the simple language (called a *direct-style*) into the expanded language (called a *monadic-style*) by the map defined by the given monad.

This enterprise, encapsulation of "impure" aspects of computation by monads, seems to be an attractive prospect, especially given the lack of standard formal models for the interfaces between semantics/pragmatics or semantics/computation. On the other hand, monadic analyses, as they have drifted among different fields, seem to have become gradually dissociated from the original monad concept.

In Bekki (2009) we aim to restore the relation between monadic analyses and categorical monads. In other words, we aim to combine the recent advances in monadic analyses with the categorical semantics of STLC along the lines of Lambek (1980). This is realized through *Meta-Lambda Calculus* (henceforth MLC; Bekki (2009)) (for details, see Sects. 2, 3 and 5) and a framework of *monadic translations* (for details, see Sect. 4). In our subsequent works, we have proposed a unified analysis for phenomena including non-determinism and contextual parameters (Bekki 2009), focus, *only* and inverse scope (Bekki and Asai (2010)), and conjoined nominals in Japanese (Hayashishita and Bekki 2011) by means of MLC and corresponding monads, which are otherwise problematic for a pure type-theoretical treatment.

1.3 Meta-Lambda Calculus

MLC is a two-level typed lambda calculus with finite products that has *meta-level types* and *meta-level terms*. Each meta-level type represents a *judgment* of a base-level term, and each meta-level lambda abstraction and product corresponds to a function relating judgments and a tuple of judgments.

While a judgment in STLC *á la* Curry is in the form of (1a), it is represented in MLC in the form of (1b), where M is a *meta-level term* and $\Gamma \vdash \alpha$ is a *meta-level type*.

(1) a. $\Gamma \vdash M : \alpha$
b. $\Vdash M : (\Gamma \vdash \alpha)$

A judgement of a term of functional type such as (2a) in STLC is represented as shown in (2b) in MLC.

(2) a. $\Gamma \vdash M : \alpha \to \beta$
b. $\Vdash M : (\Gamma \vdash \alpha \to \beta)$

This setting of MLC enables us to represent a *meta*-function type, which is not representable in STLC, as follows.

(3) $\Vdash M : (\Gamma \vdash \alpha) \Rightarrow (\Gamma' \vdash \alpha')$

As for β-conversion, a pair consisting of a base-level abstraction (notation: $\lambda x.M$) and an application (notation: MN) and a pair consisting of a meta-level abstraction (notation: $\zeta X.M$) and an application (notation: $M \between N$) behave in a parallel way:

$$\begin{aligned} (\lambda x.x)c &=_{\beta} \quad c \\ (\zeta X.X) \between c &=_{M\beta} c \end{aligned}$$

where β and $M\beta$ signify base-level and meta-level β-conversion, respectively.

The difference between base-level/meta-level terms emerges when free/bound variables are considered. In the following base-level β-conversion, y is not allowed as a substitute for x within the scope of λy.

$$\begin{aligned} (\lambda x.\lambda y.x)y &=_{\beta} (\lambda y.x)[y/x] \\ &=_{\alpha} (\lambda z.x)[y/x] \\ &\equiv \lambda z.x[y/x] \\ &\equiv \lambda z.y \end{aligned}$$

However, in the following meta-level β-conversion, y *is* allowed as a substitute for X within the scope of λy since X is a meta-level variable.

$$\begin{aligned} (\zeta X.\lambda y.X) \between y &=_{M\beta} (\lambda y.X)[y/X] \\ &\equiv \quad \lambda y.X[y/X] \\ &\equiv \quad \lambda y.y \end{aligned}$$

This occurs because the "input" of the meta-level abstraction ($\zeta X.\lambda y.X$) is not a simple value y but rather a base-level judgment $y : \alpha \vdash y : \alpha$ (namely, $\Vdash y : (y : \alpha \vdash \alpha)$ in MLC), which is illustrated in the following type inference diagram for the term $(\zeta X.\lambda y.X) \between y$.

(4)

$$\dfrac{\dfrac{\dfrac{\dfrac{}{X : (y : \alpha \vdash \alpha) \Vdash X : (y : \alpha \vdash \alpha)}(MVAR)}{X : (y : \alpha \vdash \alpha) \Vdash \lambda y.X : (\vdash \alpha \to \alpha)}(LAM)}{\Vdash \zeta X.\lambda y.X : (y : \alpha \vdash \alpha) \Rightarrow (\vdash \alpha \to \alpha)}(MLAM) \quad \dfrac{}{\Vdash y : (y : \alpha \vdash \alpha)}(VAR)}{\Vdash (\zeta X.\lambda y.X) \between y : (\vdash \alpha \to \alpha)}(MAPP)$$

At first glance, this behavior of MLC might seem peculiar enough to make one suspect whether it is a reliable calculus in the first place. Thus, it is of paramount importance for us to ensure that MLC is a sound formal system.

This article presents a revised syntax of MLC and an equational theory including base-level/meta-level substitutions and $\alpha/\beta/\eta$-conversions, together with a revised categorical semantics with respect to which we give a proof that the equational theory is sound. It gives us a firm foundation in order to pursue linguistic analyses based on various monads and monadic translations defined by means of MLC, which were presented in our previous work (Bekki 2009; Bekki and Asai 2010; Hayashishita and Bekki 2011).

1.4 Remarks on Previous Formulations of MLC

Several remarks should be made on the different formulations of MLC that we adopted in our previous works. We first presented the framework of MLC in Bekki (2009) which contains a syntax of MLC and an equational theory with meta-level β-conversion and base-level β-conversion (that we term "normal" β-conversion), which we proved to be sound with respect to a categorical semantics of MLC.

At that time, substitution was treated as a distinct syntactic structure, and in terms such as $M[L/x]$, M was implicitly assumed to be a term of base-level type following the definition of the categorical semantics.

One problem was that this setting did not provide a way to calculate base-level substitutions in a syntactic way. For example, suppose that $M : \tau_1 \Rightarrow \tau_2$ and $N : \tau_1$, where τ_2 is a base-level type. Then, $M \lightning N : \tau_2$ is a term of a base-level type, so $(M \lightning N)[L/x] : \tau_2$ must be defined, and we want to calculate it by using an equation, such as $(M \lightning N)[L/x] = (M[L/x]) \lightning (N[L/x]) : \tau_2$. However, $M[L/x]$ is not defined in Bekki (2009) since M is of meta-level type $\tau_1 \Rightarrow \tau_2$, and base-level substitution is defined only for a term of a base-level type.

This problem led us to define substitution by a set of rewriting rules in Bekki and Asai (2010), and not as an independent syntactic structure as in Bekki (2009). On the other hand, this means that base-level/meta-level β-conversions are *no longer sound* with respect to the categorical semantics in Bekki (2009) since β-conversions refer to substitutions.

This problem was fixed in Masuko and Bekki (2011) by extending the categorical semantics of substitution so that it can interpret the form $M[L/x]$, where M is a term of any type.

Another problem that was fixed in Masuko and Bekki (2011) is the relation between base-level/meta-level terms and base-level/meta-level types. In Bekki (2009), base-level terms are assumed to be of base-level types, and meta-level terms are assumed to be of meta-level types, while in fact this is not the case. It is true that a base-level term is always of base-level type, but it also means that it is of meta-level type since any base-level type is a meta-level type.

A meta-level term may be of base-level type or meta-level type. For example, in the judgement $X : (x : \alpha \vdash \alpha) \Vdash X : (x : \alpha \vdash \alpha)$, meta-level variable X is of base-level type $x : \alpha \vdash \alpha$. If $M : \tau \Rightarrow (x : \alpha \vdash \alpha)$ and $N : \tau$, then meta-level application $M \lightning N$ is of base-level type $(x : \alpha \vdash \alpha)$. Meta-level abstractions and meta-level products must be of meta-level type. Thus, the correspondence between base-level/meta-level terms and base-level/meta-level types is not as straightforward as assumed in Bekki (2009).

The notational variance of MLC is shown in Table 1.

Table 1 Notational variance of MLC

	Bekki (2009)	Bekki and Asai (2010)	This article
Meta-level functional type	$\tau_1 \mapsto \tau_2$	$\tau_1 \Rightarrow \tau_2$	$\tau_1 \Rightarrow \tau_2$
Base-level functional type	$\tau_1 \tau_2$	$\tau_1 \rightarrow \tau_2$	$\tau_1 \rightarrow \tau_2$
Meta-level functional application	$M\lceil N \rceil$	MN or $M \lhd N$	$M \between N$

2 Syntax of MLC

2.1 Base-level/Meta-level Types

The syntax of MLC is specified by the following definitions.

Definition 1 *(Alphabet for MLC)* An *alphabet* for MLC is a sextuple $\langle \mathcal{GT}, \mathcal{C}on, \mathcal{M}con, \mathcal{V}ar, \mathcal{M}var, \mathfrak{S} : \mathcal{M}var \rightarrow \mathcal{P}ow(\mathcal{V}ar)\rangle$, the elements of which respectively represent a finite collection of ground types, base-level constant symbols, *meta-level* constant symbols, base-level variables, *meta-level* variables and an assignment function of *free base-level variables* for each meta-level variable.

Definition 2 *(Base-level types)* A *collection of base-level types* (notation: $\mathcal{T}yp$) for an alphabet $\langle \mathcal{GT}, \mathcal{C}on, \mathcal{M}con, \mathcal{V}ar, \mathcal{M}var, \mathfrak{S}\rangle$ is recursively defined by the following BNF grammar (where $\gamma \in \mathcal{GT}$).

$$\mathcal{T}yp := \gamma \mid unit \mid \mathcal{T}yp \times \cdots \times \mathcal{T}yp \mid \mathcal{T}yp \rightarrow \mathcal{T}yp$$

Definition 3 *(Base-level contexts)* A *base-level context* is a finite list of pairs that are members of $\mathcal{V}ar \times \mathcal{T}yp$ (notation: $\Gamma = x_1 : \alpha_1, \ldots, x_n : \alpha_n$).

Definition 4 *(Meta-level types)* A *collection of meta-level types* (notation: $\mathcal{M}typ$) for an alphabet $\langle \mathcal{GT}, \mathcal{C}on, \mathcal{M}con, \mathcal{V}ar, \mathcal{M}var, \mathfrak{S}\rangle$ is recursively defined by the following BNF grammar (where Γ is a base-level context and $\alpha \in \mathcal{T}yp$).

$$\mathcal{M}typ \ := \ \Gamma \vdash \alpha \mid munit \mid \mathcal{M}typ \otimes \cdots \otimes \mathcal{M}typ \mid \mathcal{M}typ \Rightarrow \mathcal{M}typ$$

Definition 5 *(Meta-level contexts)* A *meta-level context* is a finite list of pairs that are members of $\mathcal{M}var \times \mathcal{M}typ$ (notation: $\Delta = X_1 : \tau_1, \ldots, X_n : \tau_n$).

2.2 Raw Terms

Definition 6 *(Raw terms)* A collection of *raw terms* (notation: Λ) for an alphabet $\langle \mathcal{GT}, \mathcal{C}on, \mathcal{M}con, \mathcal{V}ar, \mathcal{M}var, \mathfrak{S}\rangle$ is recursively defined by the following BNF notation, where $x \in \mathcal{V}ar$, $c \in \mathcal{C}on$, $X \in \mathcal{M}var$, and $C \in \mathcal{M}con$.[1]

[1]This definition of raw terms is a slightly revised version of that in Bekki (2009) and Bekki and Asai (2010).

$$\begin{array}{lll} \overline{\mathcal{X}} & ::= X & \mid \overline{\mathcal{X}}[\Lambda/x] \\ \overline{\mathcal{C}} & ::= C & \mid \overline{\mathcal{C}}[\Lambda/x] \\ \Lambda & ::= x & \mid c \mid \langle\, \rangle \mid \langle \Lambda, \ldots, \Lambda \rangle \mid \pi_i(\Lambda) \mid \lambda x.\Lambda \mid \Lambda\Lambda \mid (\Lambda) \\ & \mid\ \overline{\mathcal{X}} & \mid \overline{\mathcal{C}} \mid \langle\!\langle\, \rangle\!\rangle \mid \langle\!\langle \Lambda, \ldots, \Lambda \rangle\!\rangle \mid p_i(\Lambda) \mid \zeta X.\Lambda \mid \Lambda \lightning \Lambda \mid \end{array}$$

where $1 \le i \le$ (*length of* Λ).[2]

In Definition 6, $\overline{\mathcal{X}}$ and $\overline{\mathcal{C}}$ are sets of *meta-level variables with substitutions* and *meta-level constants with substitutions*, respectively.

Among MLC terms, those of the forms listed in the first row in the definition of Λ are called *base-level terms*, and those in the second row are called *meta-level terms*. Each row lists a form of a base-level/meta-level variable, a constant, a unit, a finite product, a projection, a lambda abstraction, and a function application, respectively. (Λ) is a bracketed term. The concept of size of terms is defined as follows:

Definition 7 *(Size of Terms)*

$$\begin{array}{ll} size(x \mid c \mid \langle\,\rangle \mid \overline{\mathcal{X}} \mid \overline{\mathcal{C}} \mid \langle\!\langle\,\rangle\!\rangle) & \stackrel{def}{\equiv} 1 \\ size(\langle M_1, \ldots, M_n \rangle \mid \langle\!\langle M_1, \ldots, M_n \rangle\!\rangle) & \stackrel{def}{\equiv} max(size(M_1), \ldots, size(M_n)) + 1 \\ size(\pi_i(M) \mid \lambda x.M \mid p_i(M) \mid \zeta X.M) & \stackrel{def}{\equiv} size(M) + 1 \\ size(M_1 M_2 \mid M_1 \lightning M_2) & \stackrel{def}{\equiv} max(size(M_1), size(M_2)) + 1 \end{array}$$

2.3 Free Base-level/Meta-level Variables

The sets of *free base-level variables* and *free meta-level variables* are defined respectively by the following two sets of rules.

Definition 8 *(Free Base-level Variables)*

$$\begin{array}{rl} fv(x) & \stackrel{def}{\equiv} \{x\} \\ fv(c) & \stackrel{def}{\equiv} \{\} \\ fv(\langle\,\rangle) & \stackrel{def}{\equiv} \{\} \\ fv(\langle M_1, \ldots, M_n \rangle) & \stackrel{def}{\equiv} \bigcup_{1 \le i \le n} fv(M_i) \\ fv(\pi_i(M)) & \stackrel{def}{\equiv} fv(M) \\ fv(\lambda x.M) & \stackrel{def}{\equiv} fv(M) - \{x\} \\ fv(M_1 M_2) & \stackrel{def}{\equiv} fv(M_1) \cup fv(M_2) \end{array}$$

$$\begin{array}{rl} fv(X) & \stackrel{def}{\equiv} \mathfrak{S}(X) \\ fv(\overline{\mathcal{X}}[M/x]) & \stackrel{def}{\equiv} (fv(\overline{\mathcal{X}}) - \{x\}) \cup fv(M) \\ fv(C) & \stackrel{def}{\equiv} \{\} \\ fv(\overline{\mathcal{C}}[M/x]) & \stackrel{def}{\equiv} (fv(\overline{\mathcal{C}}) - \{x\}) \cup fv(M) \\ fv(\langle\!\langle\,\rangle\!\rangle) & \stackrel{def}{\equiv} \{\} \\ fv(\langle\!\langle M_1, \ldots, M_n \rangle\!\rangle) & \stackrel{def}{\equiv} \bigcup_{1 \le i \le n} fv(M_i) \\ fv(p_i(M)) & \stackrel{def}{\equiv} fv(M) \\ fv(\zeta X.M) & \stackrel{def}{\equiv} fv(M) \\ fv(M_1 \lightning M_2) & \stackrel{def}{\equiv} fv(M_1) \end{array}$$

---

[2] The *length* of a base-level unit $\langle\rangle$ and a meta-level unit $\langle\!\langle\rangle\!\rangle$ is defined to be 0, and a base-level finite product of the form $\langle M_1, \ldots, M_n \rangle$ and a meta-level finite product of the form $\langle\!\langle M_1, \ldots, M_n \rangle\!\rangle$ is defined to be n.

Definition 9 *(Free Meta-level Variables)*

$$\begin{aligned}
fmv(x) &\stackrel{def}{\equiv} \{\} \\
fmv(c) &\stackrel{def}{\equiv} \{\} \\
fmv(\langle\,\rangle) &\stackrel{def}{\equiv} \{\} \\
fmv(\langle M_1,\ldots,M_n\rangle) &\stackrel{def}{\equiv} \textstyle\bigcup_{1\le i\le n} fmv(M_i) \\
fmv(\pi_i(M)) &\stackrel{def}{\equiv} fmv(M) \\
fmv(\lambda x.M) &\stackrel{def}{\equiv} fmv(M) \\
fmv(M_1M_2) &\stackrel{def}{\equiv} fmv(M_1)\cup fmv(M_2)
\end{aligned}$$

$$\begin{aligned}
fmv(X) &\stackrel{def}{\equiv} \{X\} \\
fmv(\overline{\mathcal{X}}[M/x]) &\stackrel{def}{\equiv} fmv(\overline{\mathcal{X}})\cup fmv(M) \\
fmv(C) &\stackrel{def}{\equiv} \{\} \\
fmv(\overline{\mathcal{C}}[M/x]) &\stackrel{def}{\equiv} fmv(\overline{\mathcal{C}})\cup fmv(M) \\
fmv(\langle\!\langle\,\rangle\!\rangle) &\stackrel{def}{\equiv} \{\} \\
fmv(\langle\!\langle M_1,\ldots,M_n\rangle\!\rangle) &\stackrel{def}{\equiv} \textstyle\bigcup_{1\le i\le n} fmv(M_i) \\
fmv(p_i(M)) &\stackrel{def}{\equiv} fmv(M) \\
fmv(\zeta X.M) &\stackrel{def}{\equiv} fmv(M)-\{X\} \\
fmv(M_1 \between M_2) &\stackrel{def}{\equiv} fmv(M_1)\cup fmv(M_2)
\end{aligned}$$

2.4 Judgment

A *judgment* in MLC is of the following form:

$$\Delta \Vdash M : \tau$$

where Δ is a meta-level context, M is a raw term, and τ is a meta-level type recursively derived by the following set of rules.

Definition 10 *(Base-level Structural Rules)*

$$\textbf{\textit{Weakening}}\quad (w)\frac{\Delta \Vdash M : (\Gamma \vdash \alpha_1) \qquad x \notin fv(M)}{\Delta \Vdash M : (\Gamma, x : \alpha_2 \vdash \alpha_1)}$$

$$\textbf{\textit{Exchange}}\quad (e)\frac{\Delta \Vdash M : (\Gamma,\ x : \alpha_1,\ y : \alpha_2,\ \Gamma' \vdash \alpha)}{\Delta \Vdash M : (\Gamma,\ y : \alpha_2,\ x : \alpha_1,\ \Gamma' \vdash \alpha)}$$

$$\textbf{\textit{Contraction}}\quad (c)\frac{\Delta \Vdash M : (\Gamma, x : \alpha_2, x : \alpha_2 \vdash \alpha_1)}{\Delta \Vdash M : (\Gamma, x : \alpha_2 \vdash \alpha_1)}$$

Definition 11 *(Meta-level Structural Rules)*

$$\textbf{\textit{Weakening}}\quad (Mw)\frac{\Delta \Vdash M : \tau_1 \qquad X \notin fmv(M)}{\Delta, X : \tau_2 \Vdash M : \tau_1}$$

$$\textbf{\textit{Exchange}}\quad (Me)\frac{\Delta,\ X : \tau_1,\ Y : \tau_2,\ \Delta' \Vdash M : \tau}{\Delta,\ Y : \tau_2,\ X : \tau_1,\ \Delta' \Vdash M : \tau}$$

$$\textbf{\textit{Contraction}}\quad (Mc)\frac{\Delta, X : \tau_2, X : \tau_2 \Vdash M : \tau_1}{\Delta, X : \tau_2 \Vdash M : \tau_1}$$

Definition 12 *(Typing Rules for Base-level Terms)*

Variables $(VAR)\dfrac{}{\Vdash x : (x : \alpha \vdash \alpha)}$

Constants $(CON)\dfrac{}{\Vdash c : (\vdash \alpha)}$

Products $(UNIT)\dfrac{}{\Vdash \langle\ \rangle : (\Gamma \vdash unit)}$

$$(PRD)\dfrac{\Delta \Vdash M_1 : (\Gamma \vdash \alpha_1)\ \cdots\ \Delta \Vdash M_n : (\Gamma \vdash \alpha_n)}{\Delta \Vdash \langle M_1, \ldots, M_n\rangle : (\Gamma \vdash \alpha_1 \times \cdots \times \alpha_n)}\ (1 \leq n)$$

Projections $(PJ)\dfrac{\Delta \Vdash M : (\Gamma \vdash \alpha_1 \times \cdots \times \alpha_n)}{\Delta \Vdash \pi_i(M) : (\Gamma \vdash \alpha_i)}\ (1 \leq i \leq n)$

Abstractions $(LAM)\dfrac{\Delta \Vdash M : (\Gamma, x : \alpha_2 \vdash \alpha_1)}{\Delta \Vdash \lambda x.M : (\Gamma \vdash \alpha_2 \rightarrow \alpha_1)}$

Applications $(APP)\dfrac{\Delta \Vdash M_1 : (\Gamma \vdash \alpha_2 \rightarrow \alpha_1) \qquad \Delta \Vdash M_2 : (\Gamma \vdash \alpha_2)}{\Delta \Vdash M_1 M_2 : (\Gamma \vdash \alpha_1)}$

Definition 13 *(Typing Rules for Meta-level Terms)*

Variables $(MVAR)\dfrac{}{X : \tau \Vdash X : \tau}$

$$(\overline{MVAR})\dfrac{\Delta \Vdash \overline{\mathcal{X}} : S_{\Gamma,\, x:\alpha}(\tau) \quad \Delta \Vdash M : (\Gamma \vdash \alpha)}{\Delta \Vdash \overline{\mathcal{X}}[M/x] : \tau}$$

Constants $(MCON)\dfrac{}{\Vdash C : \tau}$

$$(\overline{MCON})\dfrac{\Delta \Vdash \overline{\mathcal{C}} : S_{\Gamma,\, x:\alpha}(\tau) \quad \Delta \Vdash M : (\Gamma \vdash \alpha)}{\Delta \Vdash \overline{\mathcal{C}}[M/x] : \tau}$$

Products $(MUNIT)\dfrac{}{\Vdash \langle\!\langle\ \rangle\!\rangle : munit}$

$$(MPRD)\dfrac{\Delta \Vdash M_1 : \tau_1\ \cdots\ \Delta \Vdash M_n : \tau_n}{\Delta \Vdash \langle\!\langle M_1, \ldots, M_n\rangle\!\rangle : \tau_1 \otimes \cdots \otimes \tau_n}\ (1 \leq n)$$

Projections $(MPJ)\dfrac{\Delta \Vdash M : \tau_1 \otimes \cdots \otimes \tau_n}{\Delta \Vdash p_i(M) : \tau_i}\ (1 \leq i \leq n)$

Abstractions $(MLAM)\dfrac{\Delta, X : \tau_2 \Vdash M : \tau_1}{\Delta \Vdash \zeta X.M : \tau_2 \Rightarrow \tau_1}$

Applications $(MAPP)\dfrac{\Delta \Vdash M_1 : \tau_2 \Rightarrow \tau_1 \quad \Delta \Vdash M_2 : \tau_2}{\Delta \Vdash M_1 \lightning M_2 : \tau_1}$

In the substitution rules, $S_{\Gamma,\ x:\alpha}$ (where $x \in \mathcal{V}ar$, $\alpha \in \mathcal{T}yp$ and Γ is a base-level context) is a type transformer defined as follows:

Definition 14 *(Type Transformer $S_{\Gamma,\ x:\alpha}$)*

$$S_{\Gamma,\ x:\alpha}(\Gamma \vdash \alpha') \stackrel{def}{\equiv} (\Gamma, x : \alpha \vdash \alpha')$$
$$S_{\Gamma,\ x:\alpha}(\tau_2 \Rightarrow \tau_1) \stackrel{def}{\equiv} \tau_2 \Rightarrow S_{\Gamma,\ x:\alpha}(\tau_1)$$
$$S_{\Gamma,\ x:\alpha}(munit) \stackrel{def}{\equiv} munit$$
$$S_{\Gamma,\ x:\alpha}(\tau_1 \otimes \cdots \otimes \tau_n) \stackrel{def}{\equiv} S_{\Gamma,\ x:\alpha}(\tau_1) \otimes \cdots \otimes S_{\Gamma,\ x:\alpha}(\tau_n)$$

2.5 Substitution

Definition 15 *(Substitution of Base-level Variables)*

$$\begin{array}{rcl}
x[L/x] &\stackrel{def}{\equiv}& L \\
y[L/x] &\stackrel{def}{\equiv}& y \quad \textit{where } y \not\equiv x \\
c[L/x] &\stackrel{def}{\equiv}& c \\
\langle\ \rangle[L/x] &\stackrel{def}{\equiv}& \langle\ \rangle \\
\langle M_1,\ldots,M_n\rangle[L/x] &\stackrel{def}{\equiv}& \langle M_1[L/x],\ldots,M_n[L/x]\rangle \\
(\pi_i(M))[L/x] &\stackrel{def}{\equiv}& \pi_i(M[L/x]) \\
(\lambda x.M)[L/x] &\stackrel{def}{\equiv}& \lambda x.M \\
(\lambda y.M)[L/x] &\stackrel{def}{\equiv}& \lambda y.M[L/x] \\
&& \textit{where } y \not\equiv x \wedge (x \notin fv(M) \vee y \notin fv(L)) \\
(\lambda y.M)[L/x] &\stackrel{def}{\equiv}& \lambda z.M[z/y][L/x] \\
&& \textit{where } y \not\equiv x \wedge x \in fv(M) \wedge y \in fv(L) \wedge z \notin fv(M) \cup fv(L) \\
(M_1 M_2)[L/x] &\stackrel{def}{\equiv}& (M_1[L/x])(M_2[L/x])
\end{array}$$

$$\begin{array}{rcl}
\langle\!\langle\ \rangle\!\rangle[L/x] &\stackrel{def}{\equiv}& \langle\!\langle\ \rangle\!\rangle \\
\langle\!\langle M_1,\ldots,M_n\rangle\!\rangle[L/x] &\stackrel{def}{\equiv}& \langle\!\langle M_1[L/x],\ldots,M_n[L/x]\rangle\!\rangle \\
(p_i(M))[L/x] &\stackrel{def}{\equiv}& p_i(M[L/x]) \\
(\zeta X.M)[L/x] &\stackrel{def}{\equiv}& \zeta X.M[L/x] \\
&& \textit{where } x \notin fv(M) \vee X \notin fmv(L) \\
(\zeta X.M)[L/x] &\stackrel{def}{\equiv}& \zeta Z.M[Z/X][L/x] \\
&& \textit{where } x \in fv(M) \wedge X \in fmv(L) \wedge Z \notin fmv(M) \cup fmv(L) \\
(M_1 \between M_2)[L/x] &\stackrel{def}{\equiv}& (M_1[L/x]) \between M_2
\end{array}$$

Definition 16 *(Substitution of Meta-level Variables)*

$$\begin{array}{rcl}
x[L/X] &\stackrel{def}{\equiv}& x \\
c[L/X] &\stackrel{def}{\equiv}& c \\
\langle\ \rangle[L/X] &\stackrel{def}{\equiv}& \langle\ \rangle \\
\langle M_1,\ldots,M_n\rangle[L/X] &\stackrel{def}{\equiv}& \langle M_1[L/X],\ldots,M_n[L/X]\rangle \\
(\pi_i(M))[L/X] &\stackrel{def}{\equiv}& \pi_i(M[L/X]) \\
(\lambda x.M)[L/X] &\stackrel{def}{\equiv}& \lambda x.M[L/X] \\
(M_1 M_2)[L/X] &\stackrel{def}{\equiv}& (M_1[L/X])(M_2[L/X])
\end{array}$$

$$\begin{array}{rcl}
X[L/X] &\stackrel{def}{\equiv}& L \\
Y[L/X] &\stackrel{def}{\equiv}& Y \quad \textit{where } Y \not\equiv X \\
(\overline{\mathcal{X}}[M/x])[L/X] &\stackrel{def}{\equiv}& (\overline{\mathcal{X}}[L/X])[(M[L/X])/x] \\
C[L/X] &\stackrel{def}{\equiv}& C \\
(\overline{\mathcal{C}}[M/x])[L/X] &\stackrel{def}{\equiv}& (\overline{\mathcal{C}}[L/X])[(M[L/X])/x] \\
\langle\!\langle\ \rangle\!\rangle[L/X] &\stackrel{def}{\equiv}& \langle\!\langle\ \rangle\!\rangle \\
\langle\!\langle M_1,\ldots,M_n\rangle\!\rangle[L/X] &\stackrel{def}{\equiv}& \langle\!\langle M_1[L/X],\ldots,M_n[L/X]\rangle\!\rangle \\
(p_i(M))[L/X] &\stackrel{def}{\equiv}& p_i(M[L/X]) \\
(\zeta X.M)[L/X] &\stackrel{def}{\equiv}& \zeta X.M \\
(\zeta Y.M)[L/X] &\stackrel{def}{\equiv}& \zeta Y.M[L/X] \\
&& \textit{where } X \not\equiv Y \wedge (X \notin fmv(M) \vee Y \notin fmv(L)) \\
(\zeta Y.M)[L/X] &\stackrel{def}{\equiv}& \zeta Z.M[Z/Y][L/X] \\
&& \textit{where } X \not\equiv Y \wedge X \in fmv(M) \wedge Y \in fmv(L) \wedge Z \notin fmv(M) \cup fmv(L) \\
(M_1 \between M_2)[L/X] &\stackrel{def}{\equiv}& (M_1[L/X]) \between (M_2[L/X])
\end{array}$$

3 Theory of MLC

Axiom 1 (Equivalence)

$$(=R)\frac{\Delta \Vdash M : \tau}{\Delta \Vdash M = M : \tau} \qquad (=S)\frac{\Delta \Vdash M = N : \tau}{\Delta \Vdash N = M : \tau}$$

$$(=T)\frac{\Delta \Vdash L = M : \tau \quad \Delta \Vdash M = N : \tau}{\Delta \Vdash L = N : \tau}$$

Axiom 2 (Replacement)

$$(=\lambda)\frac{\Delta \Vdash M = N : (\Gamma, x : \alpha_2 \vdash \alpha_1)}{\Delta \Vdash \lambda x.M = \lambda x.N : (\Gamma \vdash \alpha_2 \to \alpha_1)}$$

$$(=F)\frac{\Delta \Vdash M : (\Gamma \vdash \alpha_2 \to \alpha_1) \qquad \Delta \Vdash N_1 = N_2 : (\Gamma \vdash \alpha_2)}{\Delta \Vdash MN_1 = MN_2 : (\Gamma \vdash \alpha_1)}$$

$$(=A)\frac{\Delta \Vdash M_1 = M_2 : (\Gamma \vdash \alpha_2 \to \alpha_1) \qquad \Delta \Vdash N : (\Gamma \vdash \alpha_2)}{\Delta \Vdash M_1 N = M_2 N : (\Gamma \vdash \alpha_1)}$$

$$(=\langle\rangle)\frac{\Delta \Vdash M_1 = N_1 : (\Gamma \vdash \alpha_1) \ \cdots \ \Delta \Vdash M_n = N_n : (\Gamma \vdash \alpha_n)}{\Delta \Vdash \langle M_1, \ldots, M_n \rangle = \langle N_1, \ldots, N_n \rangle : (\Gamma \vdash \alpha_1 \times \cdots \times \alpha_n)} \ (1 \leq n)$$

$$(=\pi)\frac{\Delta \Vdash M = N : (\Gamma \vdash \alpha_1 \times \cdots \times \alpha_n)}{\Delta \Vdash \pi_i(M) = \pi_i(N) : (\Gamma \vdash \alpha_i)} \ (1 \leq i \leq n)$$

Axiom 3 (Meta-level Replacement)

$$(=M\lambda)\frac{\Delta, X : \tau_2 \Vdash M = N : \tau_1}{\Delta \Vdash \zeta X.M = \zeta X.N : \tau_2 \Rightarrow \tau_1}$$

$$(=MF)\frac{\Delta \Vdash M : \tau_2 \Rightarrow \tau_1 \qquad \Delta \Vdash N_1 = N_2 : \tau_2}{\Delta \Vdash M \lightning N_1 = M \lightning N_2 : \tau_1}$$

$$(=MA)\frac{\Delta \Vdash M_1 = M_2 : \tau_2 \Rightarrow \tau_1 \qquad \Delta \Vdash N : \tau_2}{\Delta \Vdash M_1 \lightning N = M_2 \lightning N : \tau_1}$$

$$(=\langle\langle\rangle\rangle)\frac{\Delta \Vdash M_1 = N_1 : \tau_1 \ \cdots \ \Delta \Vdash M_n = N_n : \tau_n}{\Delta \Vdash \langle\langle M_1, \ldots, M_n \rangle\rangle = \langle\langle N_1, \ldots, N_n \rangle\rangle : \tau_1 \otimes \cdots \otimes \tau_n} \ (1 \leq n)$$

$$(=p)\frac{\Delta \Vdash M = N : \tau_1 \otimes \cdots \otimes \tau_n}{\Delta \Vdash p_i(M) = p_i(N) : \tau_i} \ (1 \leq i \leq n)$$

Axiom 4 **(Base-level Product Equations)**

$$(\pi 0)\frac{\Delta \Vdash M : (\Gamma \vdash unit)}{\Delta \Vdash M = \langle\ \rangle : (\Gamma \vdash unit)}$$

$$(\pi 1)\frac{\Delta \Vdash M : (\Gamma \vdash \alpha_1 \times \cdots \times \alpha_n)}{\Delta \Vdash \langle \pi_1(M), \cdots, \pi_n(M)\rangle = M : (\Gamma \vdash \alpha_1 \times \cdots \times \alpha_n)}\ (1 \leq n)$$

$$(\pi 2)\frac{\Delta \Vdash M_1 : (\Gamma \vdash \alpha_1)\ \cdots\ \Delta \Vdash M_n : (\Gamma \vdash \alpha_n)}{\Delta \Vdash \pi_i(\langle M_1, \ldots, M_n\rangle) = M_i : (\Gamma \vdash \alpha_i)}\ (1 \leq i \leq n)$$

Axiom 5 **(Meta-level Product Equations)**

$$(p0)\frac{\Delta \Vdash M : munit}{\Delta \Vdash M = \langle\!\langle\ \rangle\!\rangle : munit}$$

$$(p1)\frac{\Delta \Vdash M : \tau_1 \otimes \cdots \otimes \tau_n}{\Delta \Vdash \langle\!\langle p_1(M), \ldots, p_n(M)\rangle\!\rangle = M : \tau_1 \otimes \cdots \otimes \tau_n}\ (1 \leq n)$$

$$(p2)\frac{\Delta \Vdash M_1 : \tau_1\ \cdots\ \Delta \Vdash M_n : \tau_n}{\Delta \Vdash p_i(\langle\!\langle M_1, \ldots, M_n\rangle\!\rangle) = M_i : \tau_i}\ (1 \leq i \leq n)$$

Axiom 6 **(Base-level/Meta-level Conversions)**

$$(\alpha)\frac{\Delta \Vdash M : (\Gamma, x : \alpha_2 \vdash \alpha_1) \qquad y \notin fv(M)}{\Delta \Vdash \lambda x.M = \lambda y.M[y/x] : (\Gamma \vdash \alpha_2 \to \alpha_1)} \qquad (M\alpha)\frac{\Delta, X : \tau_2 \Vdash M : \tau_1 \qquad Y \notin fmv(M)}{\Delta \Vdash \zeta X.M = \zeta Y.M[Y/X] : \tau_2 \Rightarrow \tau_1}$$

$$(\beta)\frac{\Delta \Vdash M_1 : (\Gamma, x : \alpha_2 \vdash \alpha_1) \qquad \Delta \Vdash M_2 : (\Gamma \vdash \alpha_2)}{\Delta \Vdash (\lambda x.M_1)M_2 = M_1[M_2/x] : (\Gamma \vdash \alpha_1)} \qquad (M\beta)\frac{\Delta, X : \tau_2 \Vdash M_1 : \tau_1 \qquad \Delta \Vdash M_2 : \tau_2}{\Delta \Vdash (\zeta X.M_1) \between M_2 = M_1[M_2/X] : \tau_1}$$

$$(\eta)\frac{\Delta \Vdash M : (\Gamma \vdash \alpha_2 \to \alpha_1) \qquad x \notin fv(M)}{\Delta \Vdash \lambda x.Mx = M : (\Gamma \vdash \alpha_2 \to \alpha_1)} \qquad (M\eta)\frac{\Delta \Vdash M : \tau_2 \Rightarrow \tau_1 \qquad X \notin fmv(M)}{\Delta \Vdash \zeta X.M \between X = M : \tau_2 \Rightarrow \tau_1}$$

These conversions are proved to be sound in the categorical semantics introduced in the next section. For the proof of soundness, see Sect. 6 and Sect. 7.

4 Linguistic Monad and Monadic Translations

MLC provides a natural way for representing monads (called *internal monad* in Bekki (2009) in contrast with monads in category theory), and define such translation in which internal monads serves as parameters.

Definition 17 *(Internal Monad)* An *internal monad* is a triple $\langle \boldsymbol{T}, \eta, \mu\rangle$ (where $\boldsymbol{T}$, η, μ are MLC terms) that satisfies the following four sets of equations, where $\circ$ is a composition operator: $f \circ g \stackrel{def}{\equiv} \zeta X.f \between (g \between X)$

$\boldsymbol{T}$ conditions: $\boldsymbol{T} \between (\zeta X.X) = \zeta X.X \quad (\boldsymbol{T} \between g) \circ (\boldsymbol{T} \between f) = \boldsymbol{T} \between (g \circ f)$
η and μ conditions: $(\boldsymbol{T} \between f) \circ \eta = \eta \circ f \qquad (\boldsymbol{T} \between f) \circ \mu = \mu \circ (\boldsymbol{T} \between (\boldsymbol{T} \between f))$
Square identity: $\mu \circ (\boldsymbol{T} \between \mu) = \mu \circ \mu$
Triangular identity: $\mu \circ \eta = \zeta X.X \qquad \mu \circ (\boldsymbol{T} \between \eta) = \zeta X.X$

The four sets of conditions above are sufficient for an internal monad to correspond to a categorical monad (Bekki 2009). By specifying an internal monad, a monadic translation is defined as follows:

Definition 18 *(Translation with Internal Monad: call-by-value* (Bekki 2009)

$$\begin{aligned}
[\![x]\!]_{\boldsymbol{T}} &\stackrel{def}{\equiv} \eta \between x \\
[\![c]\!]_{\boldsymbol{T}} &\stackrel{def}{\equiv} \eta \between c \\
[\![\lambda x.M]\!]_{\boldsymbol{T}} &\stackrel{def}{\equiv} \boldsymbol{T} \between (\zeta X.\lambda x.X) \between [\![M]\!]_{\boldsymbol{T}} \\
[\![MN]\!]_{\boldsymbol{T}} &\stackrel{def}{\equiv} \mu \between ((\boldsymbol{T} \between (\zeta X.((\boldsymbol{T} \between \zeta Y.(X \between Y)) \between [\![N]\!]_{\boldsymbol{T}}))) \between [\![M]\!]_{\boldsymbol{T}}) \\
[\![\langle\,\rangle]\!]_{\boldsymbol{T}} &\stackrel{def}{\equiv} \eta \between \langle\,\rangle \\
[\![\langle M, N\rangle]\!]_{\boldsymbol{T}} &\stackrel{def}{\equiv} \mu \between ((\boldsymbol{T} \between (\zeta Y.((\boldsymbol{T} \between \zeta X.\langle X, Y\rangle) \between [\![M]\!]_{\boldsymbol{T}}))) \between [\![N]\!]_{\boldsymbol{T}}) \\
[\![\pi_i(M)]\!]_{\boldsymbol{T}} &\stackrel{def}{\equiv} \boldsymbol{T} \between (\zeta X.\pi_i(X)) \qquad \textit{where } i = 1, 2
\end{aligned}$$

Bekki (2009) establishes a link between a monadic translation (by means of an internal monad) and a categorical monad. In other words, an internal monad *is* a monad in a corresponding category called Δ-indexed category.[3]

4.1 Non-determinism

The sentence (5) is ambiguous with respect to at least two factors: the antecedent of the pronoun ‘he’ and the lexical meaning of ‘a suit’ (clothing or a legal action).

(5) He brought a suit.

Suppose that there are currently two possible antecedents for the subject pronoun, say ‘John’ and ‘Bill’. Then, in the context of standard natural language processing, a parser is expected to spell out the following set of semantic representations for the input sentence (5).

(6) $\{\ brought'(suit_1)(j'), brought'(suit_2)(j'),$
$brought'(suit_1)(b'), (brought')(suit_2)(b')\ \}$

[3]See Theorem 9 in Bekki (2009).

But this kind of 'duplication' of output trees is known to bring about a combinatorial explosion in parsing complexity, which has motivated the pursuit of an 'information packing' strategy. The monadic treatment of such non-deterministic information has been discussed in Shan (2001) and Bekki (2009), for which the internal monad for non-determinism is defined as follows in Bekki (2009).

Definition 19 *(Internal Monad for Non-determinism)*

$$\begin{aligned}
\boldsymbol{T}_{nd} &\stackrel{def}{\equiv} \zeta F.\zeta X.\{F \lightning x \mid x \in X\} \\
\eta_{nd} &\stackrel{def}{\equiv} \zeta X.\{X\} \\
\mu_{nd} &\stackrel{def}{\equiv} \zeta X.\bigcup X
\end{aligned}$$

The internal monad specified in Definition 19 defines a translation $[\![-]\!]_{nd}$ as in (7) via Definition 18.

(7) Translation to Non-deterministic Monads:

$$\begin{aligned}
[\![x]\!]_{nd} &= \{x\} \\
[\![c]\!]_{nd} &= \{c\} \\
[\![\lambda x.M]\!]_{nd} &= \left\{\lambda x.m \;\middle|\; m \in [\![M]\!]_{nd}\right\} \\
[\![MN]\!]_{nd} &= \left\{mn \;\middle|\; m \in [\![M]\!]_{nd} \wedge n \in [\![N]\!]_{nd}\right\} \\
[\![\langle\,\rangle]\!]_{nd} &= \{\langle\,\rangle\} \\
[\![\langle M, N\rangle]\!]_{nd} &= \left\{\langle m, n\rangle \;\middle|\; m \in [\![M]\!]_{nd} \wedge n \in [\![N]\!]_{nd}\right\} \\
[\![\pi_i(M)]\!]_{nd} &= \left\{\pi_i(m) \;\middle|\; m \in [\![M]\!]_{nd}\right\}
\end{aligned}$$

Then, each ambiguity due to the antecedent of 'he' and the lexical ambiguity of 'a suit' can be lexically represented in the following way.

(8)
$$\begin{aligned}
[\![he']\!]_{nd} &\stackrel{def}{\equiv} \left\{j', b'\right\} \\
[\![suit']\!]_{nd} &\stackrel{def}{\equiv} \{suit_1, suit_2\}
\end{aligned}$$

Using these expressions, the set of representations (6) can be packed into the single representation (9).

(9) $(brought'(suit'))(he')$

The non-deterministic aspects in the sentence (5) are successfully encapsulated within (9). The following equations show that the monadic translation of (9) is equivalent to (6).

(10)
$$\begin{aligned}
&[\![(brought'(suit'))]\!]_{nd} \\
&= \left\{mn \;\middle|\; m \in [\![brought']\!]_{nd} \wedge n \in [\![suit']\!]_{nd}\right\} \\
&= \left\{mn \;\middle|\; m \in \left\{brought'\right\} \wedge n \in \{suit_1, suit_2\}\right\}
\end{aligned}$$

$$= \left\{brought'(suit_1), brought'(suit_2)\right\}$$

$$\begin{aligned}
&[\![(brought'(suit'))(he')]\!]_{nd} \\
&= \left\{mn \;\middle|\; m \in [\![brought'(suit')]\!]_{nd} \wedge n \in [\![he']\!]_{nd}\right\} \\
&= \left\{mn \;\middle|\; m \in \left\{brought'(suit_1), brought'(suit_2)\right\} \wedge n \in \left\{j', b'\right\}\right\} \\
&= \{brought'(suit_1)(j'), brought(suit_2)(j'), \\
&\quad brought'(suit_1)(b'), brought(suit_2)(b')\}
\end{aligned}$$

Generally, a non-deterministic monad is called a *powerset monad* and has applications other than the representation of non-deterministic information. In Hayashishita and Bekki (2011), we defined a *disjunctive monad*, which is isomorphic to the non-deterministic monad, and this allows for the interpretation of various types of conjoined nominals in Japanese.

4.2 Contextual Parameters

Contextual parameters such as speaker/hearer, topic, and point of view, can be analyzed by means of the internal monad for contextual parameters.

Definition 20 *(Internal Monad for Contextual Parameters)*

$$\begin{aligned}
\boldsymbol{T}_{cp} &\stackrel{def}{\equiv} \zeta F.\zeta X.\lambda h.\langle F(\pi_1(x)), \pi_2(x)\rangle[Xh/x] \\
\eta_{cp} &\stackrel{def}{\equiv} \zeta X.\lambda h.\langle X, h\rangle \\
\mu_{cp} &\stackrel{def}{\equiv} \zeta X.\lambda h.(\pi_1(x))(\pi_2(x))[Xh/x]
\end{aligned}$$

The internal monad specified in Definition 20 defines a translation $[\![-]\!]_{cp}$ as in (11) via Definition 18.

(11) Translation to Contextual Parameter Monads:

$$\begin{aligned}
[\![x]\!]_{cp} &= \lambda h.\langle x, h\rangle \\
[\![c]\!]_{cp} &= \lambda h.\langle c, h\rangle \\
[\![\lambda x.M]\!]_{cp} &= \lambda h.\langle \lambda x.\pi_1(m), \pi_2(m)\rangle[[\![M]\!]_{cp}h/m] \\
[\![MN]\!]_{cp} &= \lambda h.\langle \pi_1(m)\pi_1(n), \pi_2(n)\rangle[[\![N]\!]_{cp}(\pi_2(m))/n][[\![M]\!]_{cp}h/m] \\
[\![\langle\,\rangle]\!]_{cp} &= \lambda h.\langle\langle\,\rangle, h\rangle \\
[\![\langle M, N\rangle]\!]_{cp} &= \lambda h.\langle\langle \pi_1(m), \pi_1(n)\rangle, \pi_2(n)\rangle[[\![N]\!]_{cp}(\pi_2(m))/n][[\![M]\!]_{cp}h/m] \\
[\![\pi_i(M)]\!]_{cp} &= \lambda h.\langle \pi_i(\pi_1(m)), \pi_2(m)\rangle[[\![M]\!]_{cp}h/m]
\end{aligned}$$

In this case, contextual parameters can be easily referenced in semantic representations and can also be changed.

A contextual monad provides control operators to represent a contextual parameter such as a hearer. By using $set\_hearer(x)$, we can set the current hearer to x, even in the middle of a sentence. By using $hearer()$, we can refer to a current hearer.

(12)
$$[\![set\_hearer(x)]\!]_{cp} \stackrel{def}{\equiv} \lambda h.\langle \top, x\rangle$$
$$[\![hearer()]\!]_{cp} \stackrel{def}{\equiv} \lambda h.\langle h, h\rangle$$

For example, the semantic representation of the sentence (13) can be simply stated as (14).

(13) (Pointing to John) You passed, (pointing to Mary) and you passed.

(14) $set\_hearer(j') \wedge passed'(hearer()) \wedge set\_hearer(m') \wedge passed'(hearer())$

When (14) is translated by $[\![-]\!]_{cp}$, each occurrence of 'you' successfully refers to the intended individual, although the two representations for 'you passed' in (14) are exactly the same. Suppose that $A \wedge B = \wedge(\langle A, B\rangle)$.

(15)
$$\begin{aligned}
[\![\wedge]\!]_{cp} &= \lambda h.\langle \wedge, h\rangle\\
[\![passed']\!]_{cp} &= \lambda h.\langle passed', h\rangle\\
[\![passed'(hearer())]\!]_{cp} &= \lambda h.\langle passed'(h), h\rangle\\
[\![\langle set_hearer(j'), passed'(hearer())\rangle]\!]_{cp} &= \lambda h.\langle\langle \top, passed'(j')\rangle, j'\rangle\\
[\![set_hearer(j') \wedge passed'(hearer())]\!]_{cp} &= \lambda h.\langle \top \wedge passed'(j'), j'\rangle\\
&= \lambda h.\langle passed'(j'), j'\rangle
\end{aligned}$$

Therefore, the following result is obtained.

(16) $[\![set\_hearer(j') \wedge passed'(hearer()) \wedge set\_hearer(m') \wedge passed'(hearer())]\!]_{cp}$
$= \lambda h.\langle passed'(j') \wedge passed'(m'), m'\rangle$

4.3 Continuations

The effect of covert movements, involved in focus movements and inverse scope, can be simulated in the type-theoretical framework by the internal monad for delimited continuations (Bekki and Asai 2010).

Definition 21 *(Internal Monad for Deliminted Continuations)*

$$\begin{aligned}
\boldsymbol{T}_c &\stackrel{def}{\equiv} \zeta F.\zeta X.\zeta\kappa.X \between (\zeta V.(\kappa \between (F \between V)))\\
\eta_c &\stackrel{def}{\equiv} \zeta X.\zeta\kappa.(\kappa \between X)\\
\mu_c &\stackrel{def}{\equiv} \zeta X.\zeta\kappa.(X \between (\zeta V.V \between \kappa))
\end{aligned}$$

The internal monad specified in Definition 21 defines a translation $[\![-]\!]_c$ as in (17) via Definition 18.

(17) Translation to Continuation Monad:

$$\begin{aligned}
[\![x]\!]_c &= \zeta\kappa.\kappa \lightning x \\
[\![c]\!]_c &= \zeta\kappa.\kappa \lightning c \\
[\![\lambda x.M]\!]_c &= \zeta\kappa.[\![M]\!]_c \lightning (\zeta v.\kappa \lightning (\lambda x.v)) \\
[\![MN]\!]_c &= \zeta\kappa.[\![M]\!]_c \lightning (\zeta m.[\![N]\!]_c \lightning (\zeta n.\kappa \lightning (m \lightning n))) \\
[\![\langle\,\rangle]\!]_c &= \zeta\kappa.\kappa \lightning \langle\,\rangle \\
[\![\langle M, N\rangle]\!]_c &= \zeta\kappa.[\![M]\!]_c \lightning (\zeta m.[\![N]\!]_c \lightning (\zeta n.\kappa \lightning \langle m, n\rangle)) \\
[\![\pi_i(M)]\!]_c &= \zeta\kappa.[\![M]\!]_c \lightning (\zeta m.(\kappa \lightning \pi_i(m)))
\end{aligned}$$

In light of this setting, the operators for *delimited continuations*, such as *shift* and *reset* in Danvy and Filinski (1990), are defined in the following way.

(18)
$$\begin{aligned}
[\![\mathit{shift}\ \kappa.M]\!]_c &\stackrel{def}{\equiv} \zeta\kappa.([\![M]\!]_c \lightning (\zeta x.x)) \\
[\![\mathit{reset}(M)]\!]_c &\stackrel{def}{\equiv} \zeta\kappa.\kappa \lightning ([\![M]\!]_c \lightning (\zeta x.x))
\end{aligned}$$

Although the definition above may look too simple as compared with that in Danvy and Filinski (1990), it has the same effects, as indicated by the following computations (see Bekki and Asai (2010) for the details).

(19)
$$\begin{aligned}
[\![1 + \mathit{reset}(10 + \mathit{shift}\, f.(f(f(100))))]\!] &= \zeta\kappa.(\kappa \lightning 121) \\
[\![1 + \mathit{reset}(10 + \mathit{shift}\, f.(100))]\!] &= \zeta\kappa.(\kappa \lightning 101) \\
[\![1 + \mathit{reset}(10 + \mathit{shift}\, f.(f(100) + f(1000)))]\!] &= \zeta\kappa.(\kappa \lightning 1121)
\end{aligned}$$

5 Categorical Semantics of MLC

An *interpretation* of MLC terms for an alphabet $\langle \mathcal{GT}, \mathcal{C}on, \mathcal{M}con, \mathcal{V}ar, \mathcal{M}var, \mathfrak{S}\rangle$ is specified by a quadruple as follows:

$$\mathcal{M} = \langle \mathcal{C}, \mathit{valT}, \mathit{valC}, \mathit{valMC}\rangle$$

where $\mathcal{C}$ is a Cartesian closed category with small hom-sets, *valT* is a function that sends each $\gamma \in \mathcal{GT}$ to an object in $\mathcal{C}$, and *valC* and $valMC$ are functions that send each $c \in Con$ and each $C \in \mathcal{M}con$ to a global element in $\mathcal{C}$ and $\mathcal{S}et$, respectively.

5.1 Interpretation of Bas-level/Meta-level Types

Given a quadruple, a *base-level/meta-level type interpretation* $[\![-]\!]$ is defined as follows.

Definition 22 *(Interpretation of Base-level/Meta-level Types)* $[\![-]\!]$ maps each member of $\mathcal{T}yp$ to an object in $\mathcal{C}$ and $\mathcal{M}typ$ to an object in $\mathcal{S}et$ via the following rules:

$$\begin{array}{rclrcl} [\![\gamma]\!] &=& valT(\gamma) & [\![\Gamma \vdash \alpha]\!] &=& \mathcal{C}([\![\Gamma]\!], [\![\alpha]\!]) \\ [\![\alpha_2 \to \alpha_1]\!] &=& [\![\alpha_1]\!]^{[\![\alpha_2]\!]} & [\![\tau_2 \Rightarrow \tau_1]\!] &=& [\![\tau_1]\!]^{[\![\tau_2]\!]} \\ [\![unit]\!] &=& 1 & [\![munit]\!] &=& * \\ [\![\alpha_1 \times \cdots \times \alpha_n]\!] &=& [\![\alpha_1]\!] \times \cdots \times [\![\alpha_n]\!] & [\![\tau_1 \otimes \cdots \otimes \tau_n]\!] &=& [\![\tau_1]\!] \times \cdots \times [\![\tau_n]\!] \end{array}$$

where $\gamma \in \mathcal{GT}$, 1 is a selected terminal object in $\mathcal{C}$, and $*$ is a selected terminal object in $\mathcal{S}$et, namely, a singleton set.

Definition 23 *(Interpretation of Base-level/Meta-level Contexts)* Suppose that $\Gamma = x_1 : \alpha_1, \ldots, x_n : \alpha_n$ and $\Delta = X_1 : \tau_1, \ldots, X_n : \tau_n$. Then, $[\![\Gamma]\!] = [\![\alpha_1]\!] \times \cdots \times [\![\alpha_n]\!]$ and $[\![\Delta]\!] = [\![\tau_1]\!] \times \cdots \times [\![\tau_n]\!]$. When $n = 0$, $[\![\Gamma]\!] = 1$ and $[\![\Delta]\!] = *$.

In contrast to the standard categorical semantics[4] of STLC where the interpretation $[\![-]\!]$ of a lambda term of type α under a base-level context Γ is a morphism: $[\![\Gamma]\!] \longrightarrow [\![\alpha]\!]$ in a Cartesian closed category $\mathcal{C}$, an MLC term that is of base-level type α under a base-level context Γ is interpreted under a meta-level context Δ as a morphism $[\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha]\!])$ in $\mathcal{S}et$.

5.2 Interpretation of Structural Rules

Definition 24 *(Base-level Structural Rules)* Let $|\Gamma| = n$ and $|\Gamma'| = n'$.

$$(w)\frac{[\![\Delta \Vdash M : (\Gamma \vdash \alpha_1)]\!] = m : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_1]\!]) \qquad x \notin fv(M)}{[\![\Delta \Vdash M : (\Gamma, x : \alpha_2 \vdash \alpha_1)]\!] = \mathcal{C}(\langle \pi_1, \ldots, \pi_n \rangle, [\![\alpha_1]\!]) \circ m : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!] \times [\![\alpha_2]\!], [\![\alpha_1]\!])}$$

$$(e)\frac{[\![\Delta \Vdash M : (\Gamma,\ x : \alpha_1,\ y : \alpha_2,\ \Gamma' \vdash \alpha)]\!] = m : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!] \times [\![\alpha_1]\!] \times [\![\alpha_2]\!] \times [\![\Gamma']\!], [\![\alpha]\!])}{\begin{array}{r}[\![\Delta \Vdash M : (\Gamma,\ y : \alpha_2,\ x : \alpha_1,\ \Gamma' \vdash \alpha)]\!] = \mathcal{C}(\langle \pi_1, \ldots, \pi_n, \pi_{n+2}, \pi_{n+1}, \pi_{n+3}, \ldots, \pi_{n+n'+2} \rangle, [\![\alpha]\!]) \circ m \\ : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!] \times [\![\alpha_2]\!] \times [\![\alpha_1]\!] \times [\![\Gamma']\!], [\![\alpha]\!])\end{array}}$$

$$(c)\frac{[\![\Delta \Vdash M : (\Gamma, x : \alpha_2, x : \alpha_2 \vdash \alpha_1)]\!] = m : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!] \times [\![\alpha_2]\!] \times [\![\alpha_2]\!], [\![\alpha_1]\!])}{[\![\Delta \Vdash M : (\Gamma, x : \alpha_2 \vdash \alpha_1)]\!] = \mathcal{C}(\langle \pi_1, \ldots, \pi_n, \pi_{n+1}, \pi_{n+1} \rangle, [\![\alpha_1]\!]) \circ m : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!] \times [\![\alpha_2]\!], [\![\alpha_1]\!])}$$

In case $n = 0$ (namely, $[\![\Gamma]\!] = 1$), $\langle \pi_1, \ldots, \pi_n \rangle = !_{[\![\alpha_2]\!]}$.

[4] See, for example, Lambek (1980), Lambek and Scott (1986), and Crole (1993).

Definition 25 *(Meta-level Structural Rules)* Let $|\Delta| = d$ and $|\Delta'| = d'$.

$$(Mw)\ \frac{[\![\Delta \Vdash M : \tau_1]\!] = m : [\![\Delta]\!] \longrightarrow [\![\tau_1]\!] \qquad X \notin fmv(M)}{[\![\Delta,\ X : \tau_2 \Vdash M : \tau_1]\!] = m \circ \langle \pi_1, \ldots, \pi_d \rangle : [\![\Delta]\!] \times [\![\tau_2]\!] \longrightarrow [\![\tau_1]\!]}$$

$$(Me)\ \frac{[\![\Delta,\ X : \tau_1,\ Y : \tau_2,\ \Delta' \Vdash M : \tau]\!] = m : [\![\Delta]\!] \times [\![\tau_1]\!] \times [\![\tau_2]\!] \times [\![\Delta']\!] \longrightarrow [\![\tau]\!]}{\begin{array}{r}[\![\Delta,\ Y : \tau_2,\ X : \tau_1,\ \Delta' \Vdash M : \tau]\!] = m \circ \langle \pi_1, \ldots, \pi_d, \pi_{d+2}, \pi_{d+1}, \pi_{d+3}, \ldots, \pi_{d+d'+2} \rangle \\ : [\![\Delta]\!] \times [\![\tau_2]\!] \times [\![\tau_1]\!] \times [\![\Delta']\!] \longrightarrow [\![\tau]\!]\end{array}}$$

$$(Mc)\ \frac{[\![\Delta,\ X : \tau_2,\ X : \tau_2 \Vdash M : \tau_1]\!] = m : [\![\Delta]\!] \times [\![\tau_2]\!] \times [\![\tau_2]\!] \longrightarrow [\![\tau_1]\!]}{[\![\Delta,\ X : \tau_2 \Vdash M]\!] = m \circ \langle \pi_1, \ldots, \pi_d, \pi_{d+1}, \pi_{d+1} \rangle : [\![\Delta]\!] \times [\![\tau_2]\!] \longrightarrow [\![\tau_1]\!]}$$

In case $d = 0$ (namely, $[\![\Delta]\!] = *$), $\langle \pi_1, \ldots, \pi_d \rangle =!_{[\![\tau_2]\!]}$.

5.3 *Interpretation of Base-level Terms*

There is a bijection between the two hom-sets below, which is natural in B.[5]

(20)

$$\mathcal{C}(A, B) \quad \underset{\overline{tp}:\ g \longmapsto g(*)}{\overset{tp:\ f \longmapsto (* \mapsto f)}{\rightleftarrows}} \quad \mathcal{S}et(*, \mathcal{C}(A, B))$$

This bijection amounts to a following natural isomorphism:

$$\mathcal{C}(A, -) \cong \mathcal{S}et(*, \mathcal{C}(A, -))$$

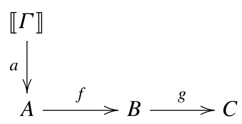
Therefore, elements of MLC with no meta-level variables (namely, Δ is empty and $[\![\Delta]\!] = *$) are in a one-to-one correspondence with elements of STLC. The covariant functor $\mathcal{C}(A, -)$ used here is a *Yoneda functor* $Y(A)$.[6]

The interpretation of (base-level) variables and constant symbols utilizes the abovementioned natural isomorphism. Any interpretation of a STLC term M of type

---

[5]See MacLane (1997), p. 60 for a proof of Proposition 2.

[6]See MacLane (1997), p. 34. $Y([\![\Gamma]\!]) = \mathcal{C}([\![\Gamma]\!], -) : \mathcal{C} \longrightarrow \mathcal{S}et$ maps a morphism $f : A \longrightarrow B$ in $\mathcal{C}$ to the morphism $\mathcal{C}([\![\Gamma]\!], f) : \mathcal{C}([\![\Gamma]\!], A) \longrightarrow \mathcal{C}([\![\Gamma]\!], B)$ in $\mathcal{S}et$. $Y([\![\Gamma]\!])(f) = \mathcal{C}([\![\Gamma]\!], f)$ is
(Footnote 6 continued)
also written as f_* and called "composition with f on the left" or "the map induced by f." It then maps a morphism $a : [\![\Gamma]\!] \longrightarrow A$ in $\mathcal{C}([\![\Gamma]\!], A)$ to $f \circ a : [\![\Gamma]\!] \longrightarrow B$ in $\mathcal{C}([\![\Gamma]\!], B)$. The two morphisms $\mathcal{C}([\![\Gamma]\!], f)$ and $\mathcal{C}([\![\Gamma]\!], g)$ induced by two composable morphisms $f : A \longrightarrow B$ and $g : B \longrightarrow C$ are also composable in $\mathcal{S}et$, as indicated in the following diagram.



α by the standard categorical semantics, which is also an element of $\mathcal{C}(\llbracket \Gamma \rrbracket, \llbracket \alpha \rrbracket)$, is mapped to the corresponding element in $\mathcal{S}et(*, \mathcal{C}(\llbracket \Gamma \rrbracket, \llbracket \alpha \rrbracket))$ via the (component $tp_{\llbracket \alpha \rrbracket}$ of) natural transformation tp.

Definition 26 *(Base-level Variables)*

$$\frac{}{\llbracket \Vdash x : (x : \alpha \vdash \alpha) \rrbracket = tp_{\llbracket \alpha \rrbracket}(id_{\llbracket \alpha \rrbracket}) : * \longrightarrow \mathcal{C}(\llbracket \alpha \rrbracket, \llbracket \alpha \rrbracket)}$$

Definition 27 *(Base-level Constant Symbols)*

$$\frac{}{\llbracket \Vdash c : (\vdash \alpha) \rrbracket = tp_{\llbracket \alpha \rrbracket}(valC(c)) : * \longrightarrow \mathcal{C}(1, \llbracket \alpha \rrbracket)}$$

Since a projection π_{n+1} in $\mathcal{C}$ is a morphism $\llbracket \Gamma \rrbracket \times \llbracket \alpha \rrbracket \times \llbracket \Gamma' \rrbracket \longrightarrow \llbracket \alpha \rrbracket$, its transpose $tp_{\llbracket \alpha \rrbracket}(\pi_{n+1})$ is a morphism $* \longrightarrow \mathcal{C}(\llbracket \Gamma \rrbracket \times \llbracket \alpha \rrbracket \times \llbracket \Gamma' \rrbracket, \llbracket \alpha \rrbracket)$ in $\mathcal{S}et$.

Theorem 7 (Base-level Variables and Constants) Suppose that $|\Gamma| = n$.

$$\frac{}{\llbracket \Delta \Vdash x : (\Gamma, x : \alpha, \Gamma' \vdash \alpha) \rrbracket = tp_{\llbracket \alpha \rrbracket}(\pi_{n+1}) \circ !_{\llbracket \Delta \rrbracket} : \llbracket \Delta \rrbracket \longrightarrow * \longrightarrow \mathcal{C}(\llbracket \Gamma \rrbracket \times \llbracket \alpha \rrbracket \times \llbracket \Gamma' \rrbracket, \llbracket \alpha \rrbracket)}$$

$$\frac{}{\llbracket \Delta \Vdash c : (\Gamma \vdash \alpha) \rrbracket = tp_{\llbracket \alpha \rrbracket}(valC(c) \circ !_{\llbracket \Gamma \rrbracket}) \circ !_{\llbracket \Delta \rrbracket} : \llbracket \Delta \rrbracket \longrightarrow * \longrightarrow \mathcal{C}(\llbracket \Gamma \rrbracket, \llbracket \alpha \rrbracket)}$$

Proof By Definition 26, Definition 27, Base-level weakening and exchange, and Meta-level weakening.

The following pair of maps $\ltimes$ and $\rtimes$ is a bijection, for a Yoneda functor $\mathcal{C}(C, -)$ preserves all finite limits.[7]

$$\text{(21)} \qquad \mathcal{C}(C, A) \times \mathcal{C}(C, B) \underset{\rtimes : h \longmapsto (\pi_1 \circ h, \pi_2 \circ h)}{\overset{\ltimes : (f,g) \longmapsto \langle f, g \rangle}{\rightleftarrows}} \mathcal{C}(C, A \times B)$$

Since this bijection is natural in C, it amounts to the following natural isomorphism that consists of the components $\ltimes_C$ and $\rtimes_C$.

$$\mathcal{C}(-, A) \times \mathcal{C}(-, B) \cong \mathcal{C}(-, A \times B)$$

This natural isomorphism extends to all finite products. We use the same symbol for its components $\ltimes_C$ and $\rtimes_C$ when no confusion seems to arise.

$$\mathcal{C}(-, A_1) \times \cdots \times \mathcal{C}(-, A_n) \cong \mathcal{C}(-, A_1 \times \cdots \times A_n)$$

[7] See MacLane (1997), p.116 for "Theorem 1" and its proof.

Definition 28 *(Base-level Products)*

$$\frac{\begin{array}{c}\overline{[\![\Vdash \langle\,\rangle : (\Gamma \vdash unit)]\!] = tp_1(!_{[\![\Gamma]\!]}) : * \longrightarrow \mathcal{C}([\![\Gamma]\!], 1)} \\ [\![\Delta \Vdash M_1 : (\Gamma \vdash \alpha_1)]\!] = m_1 : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_1]\!]) \\ \vdots \\ [\![\Delta \Vdash M_n : (\Gamma \vdash \alpha_n)]\!] = m_n : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_n]\!])\end{array}}{\begin{array}{l}[\![\Delta \Vdash \langle M_1, \ldots, M_n\rangle : (\Gamma \vdash \alpha_1 \times \cdots \times \alpha_n)]\!] = \ltimes_{[\![\Gamma]\!]} \circ \langle m_1, \ldots, m_n\rangle \\ \quad : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_1]\!]) \times \cdots \times \mathcal{C}([\![\Gamma]\!], [\![\alpha_n]\!]) \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_1]\!] \times \cdots \times [\![\alpha_n]\!])\end{array}} \; (1 \le n)$$

$$\frac{[\![\Delta \Vdash M : (\Gamma \vdash \alpha_1 \times \cdots \times \alpha_n)]\!] = m : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_1]\!] \times \cdots \times [\![\alpha_n]\!])}{[\![\Delta \Vdash \pi_i(M) : (\Gamma \vdash \alpha_i)]\!] = \mathcal{C}([\![\Gamma]\!], \pi_i) \circ m : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_i]\!])} \; (1 \le i \le n)$$

We assume that $\mathcal{C}$ is Cartesian closed and there is a bijection between the following two hom-sets, which is natural in C.

(22)
$$\mathcal{C}(C \times A, B) \; \underset{\hat{\_}\,:\, g \longmapsto ev \circ (g \times id_A)}{\overset{\lambda:\, f \longmapsto \lambda(f)}{\rightleftarrows}} \; \mathcal{C}(C, B^A)$$

This bijection amounts to the following natural isomorphism, whose components are λ_C and $\hat{\_}_C$ that are morphisms in $\mathcal{S}et$.

$$\mathcal{C}(- \times A, B) \cong \mathcal{C}(-, B^A)$$

Definition 29 *(Base-level Lambda Abstractions)*

$$\frac{[\![\Delta \Vdash M : (\Gamma, x : \alpha_2 \vdash \alpha_1)]\!] = m : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!] \times [\![\alpha_2]\!], [\![\alpha_1]\!])}{[\![\Delta \Vdash \lambda x.M : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!] = \lambda_{[\![\Gamma]\!]} \circ m : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_1]\!]^{[\![\alpha_2]\!]})}$$

Definition 30 *(Base-level Function Applications)*

$$\frac{\begin{array}{c}[\![\Delta \Vdash M_1 : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!] = m_1 : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_1]\!]^{[\![\alpha_2]\!]}) \\ [\![\Delta \Vdash M_2 : (\Gamma \vdash \alpha_2)]\!] = m_2 : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_2]\!])\end{array}}{\begin{array}{l}[\![\Delta \Vdash M_1 M_2 : (\Gamma \vdash \alpha_1)]\!] = \mathcal{C}([\![\Gamma]\!], ev) \circ \ltimes_{[\![\Gamma]\!]} \circ \langle m_1, m_2\rangle \\ : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_1]\!]^{[\![\alpha_2]\!]}) \times \mathcal{C}([\![\Gamma]\!], [\![\alpha_2]\!]) \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_1]\!]^{[\![\alpha_2]\!]} \times [\![\alpha_2]\!]) \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_1]\!])\end{array}}$$

Definition 31 *(Base-level Substitution)*

$$\frac{\Delta \Vdash M : S_{\Gamma,\, x:\alpha}(\tau) \qquad [\![\Delta \Vdash L : (\Gamma \vdash \alpha)]\!] = l : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha]\!])}{[\![\Delta \Vdash M[L/x] : \tau]\!] = \left\| \Delta \Vdash M : S_{\Gamma,\, x:\alpha}(\tau) \right\|_l^{\Gamma, x:\alpha} : [\![\Delta]\!] \longrightarrow [\![\tau]\!]}$$

$$where\ M \in \overline{\mathcal{X}} \cup \overline{\mathcal{C}}$$

The interpretation of base-level substitution given in Definition 31 is based on the following natural transformation, which is natural in C.

$$\mathcal{C}(C \times A, B) \times \mathcal{C}(C, A) \overset{\text{-}}{\longrightarrow} \mathcal{C}(C, B)$$

In this case, the substitution map is a component of this natural transformation, which is defined as follows, where $m_1 : C \times A \longrightarrow B$ and $m_2 : C \longrightarrow A$.

$$\mathcal{S}ub_C : (m_1, m_2) \longmapsto m_1 \circ \langle id_C, m_2 \rangle$$

By means of the substitution map $\mathcal{S}ub$, we can define the following set of translation rules, which is utilized in Definition 31.

Definition 32 *(Substitution Translation Rule)* Suppose that $|\Delta| = d$. For any base-level variable x of type α and any morphism $m : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha]\!])$, $\|-\|_m^{\Gamma,x:\alpha}$ is a map from a meta-level term $\Delta \Vdash M : \tau$ (of one of the following forms) to a morphism $[\![\Delta]\!] \longrightarrow [\![S_{\Gamma,\ x:\alpha}(\tau)]\!]$, defined as follows[8]:

$$\begin{aligned}
\|\Delta \Vdash M : (\Gamma, x : \alpha \vdash \alpha')\|_m^{\Gamma,x:\alpha} &\stackrel{def}{\equiv} \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash M : (\Gamma, x : \alpha \vdash \alpha')]\!], m \rangle \\
\|\Delta \Vdash M : \tau_2 \Rightarrow \tau_1\|_m^{\Gamma,x:\alpha} &\stackrel{def}{\equiv} \lambda(\|\Delta, X : \tau_2 \Vdash M \between X : \tau_1\|_{m\circ\langle \pi_1,\ldots,\pi_d \rangle}^{\Gamma,x:\alpha}) \\
\|\Delta \Vdash M : munit\|_m^{\Gamma,x:\alpha} &\stackrel{def}{\equiv} !_{[\![\Delta]\!]} \\
\|\Delta \Vdash M : \tau_1 \otimes \cdots \otimes \tau_n\|_m^{\Gamma,x:\alpha} &\stackrel{def}{\equiv} \langle \|\Delta \Vdash p_1(M) : \tau_1\|_m^{\Gamma,x:\alpha}, \ldots, \|\Delta \Vdash p_n(M) : \tau_n\|_m^{\Gamma,x:\alpha} \rangle
\end{aligned}$$

5.4 *Interpretation of Meta-level Terms*

The interpretation of meta-level terms in MLC is similar to the interpretation of (base-level) terms of STLC.

Definition 33 *(Meta-level Variables)*

$$\frac{}{[\![X : \tau \Vdash X : \tau]\!] = id_{[\![\tau]\!]} : [\![\tau]\!] \longrightarrow [\![\tau]\!]}$$

Definition 34 *(Meta-level Constant Symbols)*

$$\frac{}{[\![\Vdash C : \tau]\!] = valMC(C) : * \longrightarrow [\![\tau]\!]}$$

[8]For the definition of $S_{\Gamma,\ x:\alpha}$, see Definition 14.

Theorem 8 **(Meta-level Variables and Constants)** Suppose that $|\Delta| = d$.

$$\frac{}{[\![\Delta, X : \tau, \Delta' \Vdash X : \tau]\!] = \pi_{d+1} : [\![\Delta]\!] \times [\![\tau]\!] \times [\![\Delta']\!] \longrightarrow [\![\tau]\!]}$$

$$\frac{}{[\![\Delta \Vdash C : \tau]\!] = valMC(C) \circ !_{\Delta} : [\![\Delta]\!] \longrightarrow * \longrightarrow [\![\tau]\!]}$$

Proof By Definition 33, Definition 34, and the meta-level weakening and exchange.

In other words, a meta-level variable X_i is interpreted simply as the projection π_i in $\mathcal{S}et$, which selects the i-th member of Δ and returns its value.

Definition 35 *(Meta-level Products)*

$$\frac{}{[\![\Vdash \langle\!\langle\, \rangle\!\rangle : munit]\!] = id_* : * \longrightarrow *}$$

$$\frac{\begin{array}{c}[\![\Delta \Vdash M_1 : \tau_1]\!] = m_1 : [\![\Delta]\!] \longrightarrow [\![\tau_1]\!] \\ \vdots \\ [\![\Delta \Vdash M_n : \tau_n]\!] = m_n : [\![\Delta]\!] \longrightarrow [\![\tau_n]\!]\end{array}}{[\![\Delta \Vdash \langle\!\langle M_1, \ldots, M_n \rangle\!\rangle : \tau_1 \otimes \cdots \otimes \tau_n]\!] = \langle m_1, \ldots, m_n \rangle : [\![\Delta]\!] \longrightarrow [\![\tau_1]\!] \times \cdots \times [\![\tau_n]\!]}\ (1 \leq n)$$

$$\frac{[\![\Delta \Vdash M : \tau_1 \otimes \cdots \otimes \tau_n]\!] = m : [\![\Delta]\!] \longrightarrow [\![\tau_1]\!] \times \cdots \times [\![\tau_n]\!]}{[\![\Delta \Vdash p_i(M) : \tau_i]\!] = \pi_i \circ m : [\![\Delta]\!] \longrightarrow [\![\tau_i]\!]}\ (1 \leq i \leq n)$$

Meta-level products roughly correspond to products in $\mathcal{S}et$, whose interpretation is similar to the interpretation of (base-level) products in STLC.

Definition 36 *(Meta-level Lambda Abstractions)*

$$\frac{[\![\Delta, X : \tau_2 \Vdash M : \tau_1]\!] = m : [\![\Delta]\!] \times [\![\tau_2]\!] \longrightarrow [\![\tau_1]\!]}{[\![\Delta \Vdash \zeta X.M : \tau_2 \Rightarrow \tau_1]\!] = \lambda(m) : [\![\Delta]\!] \longrightarrow [\![\tau_1]\!]^{[\![\tau_2]\!]}}$$

Definition 37 *(Meta-level Function Applications)*

$$\frac{\begin{array}{c}[\![\Delta \Vdash M_1 : \tau_2 \Rightarrow \tau_1]\!] = m_1 : [\![\Delta]\!] \longrightarrow [\![\tau_1]\!]^{[\![\tau_2]\!]} \\ [\![\Delta \Vdash M_2 : \tau_2]\!] = m_2 : [\![\Delta]\!] \longrightarrow [\![\tau_2]\!]\end{array}}{[\![\Delta \Vdash M_1 \lightning M_2 : \tau_1]\!] = ev \circ \langle m_1, m_2 \rangle : [\![\Delta]\!] \longrightarrow [\![\tau_1]\!]}$$

6 Soundness of Base-level Substitution and Conversions

In the following two sections, we prove the soundness of base-level/meta-level substitution and conversions.[9]

The domain of interpretation of substitution given in Definition 31 is not limited to meta-level variables and meta-level constants, but is extended to encompass any other terms as well.

Lemma 1 (Base-level Substitution Lemma)

$$\frac{\Delta \Vdash M : S_{\Gamma,\, x:\alpha}(\tau) \qquad [\![\Delta \Vdash L : (\Gamma \vdash \alpha)]\!] = l : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha]\!])}{[\![\Delta \Vdash M[L/x] : \tau]\!] = \left\| \Delta \Vdash M : S_{\Gamma,\, x:\alpha}(\tau) \right\|_l^{\Gamma, x:\alpha} : [\![\Delta]\!] \longrightarrow [\![\tau]\!]}$$

By induction on the size and type of the term M. Suppose that $size(M) = s$, $|\Delta| = d$, $|\Gamma| = n$, $h_1, \ldots, h_d \in [\![\Delta]\!]$, and the morphism l maps $h_1, \ldots, h_d$ to $f : [\![\Gamma]\!] \longrightarrow [\![\alpha]\!]$.

First, let us consider all possible cases for $s = 1$ ($M = x \mid y \mid c \mid \langle\,\rangle \mid \overline{\mathcal{X}} \mid \overline{\mathcal{C}} \mid \langle\!\langle\,\rangle\!\rangle$).

– If $M = x$ and $\tau = (\Gamma \vdash \alpha)$, the left side of the equation is $[\![\Delta \Vdash L : (\Gamma \vdash \alpha)]\!] = l$. On the other hand,

$$\begin{aligned} &\|\Delta \Vdash x : (\Gamma, x : \alpha \vdash \alpha)\|_l^{\Gamma, x:\alpha} \\ &= \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash x : (\Gamma, x : \alpha \vdash \alpha)]\!], l \rangle \\ &= \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle tp_{[\![\alpha]\!]}(\pi_{n+1}) \circ !_{[\![\Delta]\!]}, l \rangle \text{ --- } (\dagger) \end{aligned}$$

Recall that tp and $\mathcal{S}ub$ map morphisms in the following way:

$$\begin{aligned} tp_{[\![\alpha]\!]}(\pi_{n+1}) \circ !_{[\![\Delta]\!]} &: h_1, \ldots, h_d \longmapsto \pi_{n+1} \\ \mathcal{S}ub_{[\![\Gamma]\!]} &: \ (\pi_{n+1}, f) \longmapsto \pi_{n+1} \circ \langle id_{[\![\Gamma]\!]}, f \rangle \end{aligned}$$

Thus, l and $(\dagger)$ are identical morphisms since $\pi_{n+1} \circ \langle id_{[\![\Gamma]\!]}, f \rangle = f$.

– If $M = y$ and $\tau = (\Gamma \vdash \alpha')$, the left side of the equation is $[\![\Delta \Vdash y : (\Gamma \vdash \alpha')]\!]$. On the other hand,

$$\begin{aligned} &\left\|\Delta \Vdash y : (\Gamma, x : \alpha \vdash \alpha')\right\|_l^{\Gamma, x:\alpha} \\ &= \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash y : (\Gamma, x : \alpha \vdash \alpha')]\!], l \rangle \\ &= \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle \mathcal{C}(\langle \pi_1, \ldots, \pi_n \rangle, [\![\alpha']\!]) \circ [\![\Delta \Vdash y : (\Gamma \vdash \alpha')]\!], l \rangle \text{ --- } (\dagger) \end{aligned}$$

Suppose that the morphism $[\![\Delta \Vdash y : (\Gamma \vdash \alpha')]\!]$ maps $h_1, \ldots, h_d$ to $g : [\![\Gamma]\!] \longrightarrow [\![\alpha']\!]$. Then, $(\dagger)$ maps $h_1, \ldots, h_d$ to

$$\begin{aligned} g \circ \langle \pi_1, \ldots, \pi_n \rangle \circ \langle id_{[\![\Gamma]\!]}, f \rangle &= g \circ \langle \pi_1 \circ \langle id_{[\![\Gamma]\!]}, f \rangle, \ldots, \pi_n \circ \langle id_{[\![\Gamma]\!]}, f \rangle \rangle \\ &= g \circ \langle \pi_1, \ldots, \pi_n \rangle \\ &= g \end{aligned}$$

[9]The proof is the revised version of what we presented in Masuko and Bekki (2011).

Therefore, $[\![\Delta \Vdash y : (\Gamma \vdash \alpha')]\!]$ and (†) are identical morphisms.
The same is true for $M = c$ and $M = \langle\,\rangle$.

– If $M = \overline{\mathcal{X}}$ or $M = \overline{\mathcal{C}}$, by Definition 31.
– If $M = \langle\!\langle\,\rangle\!\rangle$, the equation holds obviously since the morphism $!_{[\![\Delta]\!]}$ is unique.

Next, suppose the equation holds for a term whose size is less than or equals to s and show that the equation also holds for a term whose size is $s+1$ ($M = \langle M_1, \ldots, M_k\rangle \mid \pi_i(M) \mid \lambda x.M \mid M_1 M_2 \mid \langle\!\langle M_1, \ldots, M_k\rangle\!\rangle \mid p_i(M) \mid \zeta X.M \mid M_1 \between M_2$).

By induction on the structure of the type τ.

First, let us consider the cases for $\tau = (\Gamma \vdash \alpha')$. In this case, the right side of the equation is

$$\left\|\Delta \Vdash M : (\Gamma, x : \alpha \vdash \alpha')\right\|_l^{\Gamma, x:\alpha} = \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash M : (\Gamma, x : \alpha \vdash \alpha')]\!], l\rangle$$

- If $M = \langle\!\langle M_1, \ldots, M_k\rangle\!\rangle$ or $\zeta X.M$, then the equation holds vacuously.
- If $M = \langle M_1, \ldots, M_k\rangle$ and $\alpha' = \alpha_1 \times \cdots \times \alpha_k$, the left side of the equation is

$$\begin{aligned}
&[\![\Delta \Vdash \langle M_1, \ldots, M_k\rangle[L/x] : (\Gamma \vdash \alpha_1 \times \cdots \times \alpha_k)]\!] \\
&= [\![\Delta \Vdash \langle M_1[L/x], \ldots, M_k[L/x]\rangle : (\Gamma \vdash \alpha_1 \times \cdots \times \alpha_k)]\!] \\
&= \ltimes_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash M_1[L/x] : (\Gamma \vdash \alpha_1)]\!], \ldots, [\![\Delta \Vdash M_k[L/x] : (\Gamma \vdash \alpha_k)]\!]\rangle \\
&= \ltimes_{[\![\Gamma]\!]} \circ \langle \|\Delta \Vdash M_1 : (\Gamma, x : \alpha \vdash \alpha_1)\|_l^{\Gamma, x:\alpha}, \ldots, \|\Delta \Vdash M_k : (\Gamma, x : \alpha \vdash \alpha_k)\|_l^{\Gamma, x:\alpha}\rangle \\
&= \ltimes_{[\![\Gamma]\!]} \circ \langle \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash M_1 : (\Gamma, x : \alpha \vdash \alpha_1)]\!], l\rangle, \ldots, \\
&\qquad \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash M_k : (\Gamma, x : \alpha \vdash \alpha_k)]\!], l\rangle\rangle \text{ --- } (\dagger)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash \langle M_1, \ldots, M_k\rangle : (\Gamma, x : \alpha \vdash \alpha_1 \times \cdots \times \alpha_k)]\!], l\rangle \\
&= \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle \ltimes_{[\![\Gamma]\!] \times [\![\alpha]\!]} \circ \langle [\![\Delta \Vdash M_1 : (\Gamma, x : \alpha \vdash \alpha_1)]\!], \ldots, \\
&\qquad [\![\Delta \Vdash M_k : (\Gamma, x : \alpha \vdash \alpha_k)]\!]\rangle, l\rangle \text{ --- } (\ddagger)
\end{aligned}$$

Suppose that each morphism $[\![\Delta \Vdash M_i : (\Gamma, x : \alpha \vdash \alpha_i)]\!]$ maps $h_1, \ldots, h_d$ to $g_i : [\![\Gamma]\!] \times [\![\alpha]\!] \longrightarrow [\![\alpha_i]\!]$ for $(1 \le i \le k)$. Then,

$$\begin{aligned}
(\dagger) &: h_1, \ldots, h_d \longmapsto \langle g_1 \circ \langle id_{[\![\Gamma]\!]}, f\rangle, \ldots, g_k \circ \langle id_{[\![\Gamma]\!]}, f\rangle\rangle \\
(\ddagger) &: h_1, \ldots, h_d \longmapsto \langle g_1, \ldots, g_k\rangle \circ \langle id_{[\![\Gamma]\!]}, f\rangle
\end{aligned}$$

Therefore, (†) and (‡) are identical morphisms.

- If $M = \pi_i(M)$, then the left side is

$$\begin{aligned}
&[\![\Delta \Vdash (\pi_i(M))[L/x] : (\Gamma \vdash \alpha')]\!] = [\![\Delta \Vdash \pi_i(M[L/x]) : (\Gamma \vdash \alpha')]\!] \\
&= \mathcal{C}([\![\Gamma]\!], \pi_i) \circ [\![\Delta \Vdash M[L/x] : (\Gamma \vdash \alpha_1 \times \cdots \times \alpha' \times \cdots \times \alpha_k)]\!] \\
&= \mathcal{C}([\![\Gamma]\!], \pi_i) \circ \left\|\Delta \Vdash M : (\Gamma, x : \alpha \vdash \alpha_1 \times \cdots \times \alpha' \times \cdots \times \alpha_k)\right\|_l^{\Gamma, x:\alpha} \\
&= \mathcal{C}([\![\Gamma]\!], \pi_i) \circ \\
&\qquad \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash M : (\Gamma, x : \alpha \vdash \alpha_1 \times \cdots \times \alpha' \times \cdots \times \alpha_k)]\!], l\rangle \text{ --- } (\dagger)
\end{aligned}$$

On the other hand, the following equation holds.

$$\begin{aligned}
&\mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash \pi_i(M) : (\Gamma, x : \alpha \vdash \alpha')]\!], l\rangle \\
&= \mathcal{S}ub_{[\![\Gamma]\!]} \circ \\
&\qquad \langle \mathcal{C}([\![\Gamma]\!], \pi_i) \circ [\![\Delta \Vdash M : (\Gamma, x : \alpha \vdash \alpha_1 \times \cdots \times \alpha' \times \cdots \times \alpha_k)]\!], l\rangle \text{ --- } (\ddagger)
\end{aligned}$$

Suppose that $[\![\Delta \Vdash M : (\Gamma, x : \alpha \vdash \alpha_1 \times \cdots \times \alpha' \times \cdots \times \alpha_k)]\!]$ maps $h_1, \ldots, h_d$ to $g : [\![\Gamma]\!] \times [\![\alpha]\!] \longrightarrow [\![\alpha_1]\!] \times \cdots \times [\![\alpha']\!] \times \cdots \times [\![\alpha_k]\!]$. Then, since both (†) and (‡) map the d-tuple to $\pi_i \circ g \circ \langle id_{[\![\Gamma]\!]}, f\rangle$, (†) and (‡) are identical morphisms.

- If $M = \lambda x.M$, the case is the same as that for y.
- If $M = \lambda y.M$ ($x \not\equiv y \wedge (x \notin fv(M) \vee y \notin fv(L))$) and $\alpha' = \alpha_2 \to \alpha_1$, let m be a morphism $[\![\Delta \Vdash M : (\Gamma, x : \alpha, y : \alpha_2 \vdash \alpha_1)]\!]$. Then, the following equation holds.

$$\begin{aligned}
&[\![\Delta \Vdash (\lambda y.M)[L/x] : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!] \;=\; [\![\Delta \Vdash \lambda y.M[L/x] : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!]\\
&= \lambda_{[\![\Gamma]\!]} \circ [\![\Delta \Vdash M[L/x] : (\Gamma, y : \alpha_2 \vdash \alpha_1)]\!]\\
&= \lambda_{[\![\Gamma]\!]} \circ \|\Delta \Vdash M : (\Gamma, y : \alpha_2, x : \alpha \vdash \alpha_1)\|^{\Gamma, y:\alpha_2, x:\alpha}_{\mathcal{C}(\langle \pi_1,\ldots,\pi_n\rangle, [\![\alpha]\!])\circ l}\\
&= \lambda_{[\![\Gamma]\!]} \circ \mathcal{S}ub_{[\![\Gamma]\!]\times[\![\alpha_2]\!]} \circ \langle [\![\Delta \Vdash M : (\Gamma, y : \alpha_2, x : \alpha \vdash \alpha_1)]\!], \mathcal{C}(\langle \pi_1, \ldots, \pi_n\rangle, [\![\alpha]\!]) \circ l\rangle\\
&= \lambda_{[\![\Gamma]\!]} \circ \mathcal{S}ub_{[\![\Gamma]\!]\times[\![\alpha_2]\!]} \circ\\
&\quad \langle \mathcal{C}(\langle \pi_1, \ldots, \pi_n, \pi_{n+2}, \pi_{n+1}\rangle, [\![\alpha_1]\!]) \circ m, \mathcal{C}(\langle \pi_1, \ldots, \pi_n\rangle, [\![\alpha]\!]) \circ l\rangle \text{ — (†)}
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash \lambda y.M : (\Gamma, x : \alpha \vdash \alpha_2 \to \alpha_1)]\!], l\rangle\\
&= \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle \lambda_{[\![\Gamma]\!]\times[\![\alpha]\!]} \circ m, l\rangle \text{ — (‡)}
\end{aligned}$$

Suppose that m maps $h_1, \ldots, h_d$ to $g : [\![\Gamma]\!] \times [\![\alpha]\!] \times [\![\alpha_2]\!] \longrightarrow [\![\alpha_1]\!]$. Then, (†) maps the d-tuple to

$$\begin{aligned}
&\lambda(g \circ \langle \pi_1, \ldots, \pi_n, \pi_{n+2}, \pi_{n+1}\rangle \circ \langle id_{[\![\Gamma]\!]\times[\![\alpha_2]\!]}, f \circ \langle \pi_1, \ldots, \pi_n\rangle\rangle)\\
&= \lambda(g \circ \langle \pi_1, \ldots, \pi_n, f \circ \langle \pi_1, \ldots, \pi_n\rangle, \pi_{n+1}\rangle)
\end{aligned}$$

and (‡) maps to $\lambda(g) \circ \langle id_{[\![\Gamma]\!]}, f\rangle$.

By the uniqueness of the transpose, the equality of (†) and (‡) is proved by the following equation:

$$\begin{aligned}
&ev \circ ((\lambda(g) \circ \langle id_{[\![\Gamma]\!]}, f\rangle) \times id_{[\![\alpha_2]\!]})\\
&= ev \circ (\lambda(g) \times id_{[\![\alpha_2]\!]}) \circ (\langle id_{[\![\Gamma]\!]}, f\rangle \times id_{[\![\alpha_2]\!]})\\
&= g \circ (\langle id_{[\![\Gamma]\!]}, f\rangle \times id_{[\![\alpha_2]\!]})\\
&= g \circ \langle \pi_1, \ldots, \pi_n, f \circ \langle \pi_1, \ldots, \pi_n\rangle, \pi_{n+1}\rangle
\end{aligned}$$

- If $M = \lambda y.M$ ($x \not\equiv y \wedge x \in fv(M) \wedge y \in fv(L)$) and $\alpha' = \alpha_2 \to \alpha_1$, let m be a morphism $[\![\Delta \Vdash M : (\Gamma, x : \alpha, y : \alpha_2 \vdash \alpha_1)]\!]$. Then, the left side of the equation is

$$\begin{aligned}
&[\![\Delta \Vdash (\lambda y.M)[L/x] : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!]\\
&= [\![\Delta \Vdash \lambda z.M[z/y][L/x] : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!]\\
&= \lambda_{[\![\Gamma]\!]} \circ [\![\Delta \Vdash M[z/y][L/x] : (\Gamma, z : \alpha_2 \vdash \alpha_1)]\!]\\
&= \lambda_{[\![\Gamma]\!]} \circ \|\Delta \Vdash M[z/y] : (\Gamma, z : \alpha_2, x : \alpha \vdash \alpha_1)\|^{\Gamma, z:\alpha_2, x:\alpha}_{\mathcal{C}(\langle \pi_1,\ldots,\pi_n\rangle, [\![\alpha_1]\!])\circ l}\\
&= \lambda_{[\![\Gamma]\!]} \circ \mathcal{S}ub_{[\![\Gamma]\!]\times[\![\alpha_2]\!]} \circ\\
&\quad \langle [\![\Delta \Vdash M[z/y] : (\Gamma, z : \alpha_2, x : \alpha \vdash \alpha_1)]\!], \mathcal{C}(\langle \pi_1, \ldots, \pi_n\rangle, [\![\alpha_1]\!]) \circ l\rangle\\
&= \lambda_{[\![\Gamma]\!]} \circ \mathcal{S}ub_{[\![\Gamma]\!]\times[\![\alpha_2]\!]} \circ\\
&\quad \langle \|\Delta \Vdash M : (\Gamma, z : \alpha_2, x : \alpha, y : \alpha_2 \vdash \alpha_1)\|^{\Gamma, z:\alpha_2, x:\alpha, y:\alpha_2}_{tp_{[\![\alpha_2]\!]}(\pi_{n+1})\circ !_{[\![\Delta]\!]}},\\
&\quad \mathcal{C}(\langle \pi_1, \ldots, \pi_n\rangle, [\![\alpha_1]\!]) \circ l\rangle\\
&= \lambda_{[\![\Gamma]\!]} \circ \mathcal{S}ub_{[\![\Gamma]\!]\times[\![\alpha_2]\!]} \circ\\
&\quad \langle \mathcal{S}ub_{[\![\Gamma]\!]\times[\![\alpha_2]\!]\times[\![\alpha]\!]} \circ \langle [\![\Delta \Vdash M : (\Gamma, z : \alpha_2, x : \alpha, y : \alpha_2 \vdash \alpha_1)]\!], tp_{[\![\alpha_2]\!]}(\pi_{n+1})\circ !_{[\![\Delta]\!]}\rangle,\\
&\quad \mathcal{C}(\langle \pi_1, \ldots, \pi_n\rangle, [\![\alpha_1]\!]) \circ l\rangle\\
&= \lambda_{[\![\Gamma]\!]} \circ \mathcal{S}ub_{[\![\Gamma]\!]\times[\![\alpha_2]\!]} \circ\\
&\quad \langle \mathcal{S}ub_{[\![\Gamma]\!]\times[\![\alpha_2]\!]\times[\![\alpha]\!]} \circ \langle \mathcal{C}(\langle \pi_1, \ldots, \pi_n, \pi_{n+2}, \pi_{n+3}\rangle, [\![\alpha_1]\!]) \circ m, tp_{[\![\alpha_2]\!]}(\pi_{n+1})\circ !_{[\![\Delta]\!]}\rangle,\\
&\quad \mathcal{C}(\langle \pi_1, \ldots, \pi_n\rangle, [\![\alpha_1]\!]) \circ l\rangle \text{ — (†)}
\end{aligned}$$

The right side is the same as the one above: $\mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle \lambda_{[\![\Gamma]\!] \times [\![\alpha]\!]} \circ m, l \rangle$ — (‡)
Suppose that m maps $h_1, \ldots, h_d$ to $g : [\![\Gamma]\!] \times [\![\alpha]\!] \times [\![\alpha_2]\!] \longrightarrow [\![\alpha_1]\!]$. Then, (†) maps the d-tuple to

$$\begin{aligned}
&\lambda(g \circ \langle \pi_1, \ldots, \pi_n, \pi_{n+2}, \pi_{n+3} \rangle \circ \langle id_{[\![\Gamma]\!] \times [\![\alpha_2]\!] \times [\![\alpha]\!]}, \pi_{n+1} \rangle \\
&\quad \circ \langle id_{[\![\Gamma]\!] \times [\![\alpha_2]\!]}, f \circ \langle \pi_1, \ldots, \pi_n \rangle \rangle) \\
&= \lambda(g \circ \langle \pi_1, \ldots, \pi_n, \pi_{n+2}, \pi_{n+1} \rangle \circ \langle id_{[\![\Gamma]\!] \times [\![\alpha_2]\!]}, f \circ \langle \pi_1, \ldots, \pi_n \rangle \rangle) \\
&= \lambda(g \circ \langle \pi_1, \ldots, \pi_n, f \circ \langle \pi_1, \ldots, \pi_n \rangle, \pi_{n+1} \rangle)
\end{aligned}$$

and (‡) maps to $\lambda(g) \circ \langle id_{[\![\Gamma]\!]}, f \rangle$.
Therefore, following the argument above, (†) = (‡) holds.

- If $M = M_1 M_2$, let $\alpha' = \alpha_1$ and m_1, m_2 be the following morphisms:

$$\begin{aligned}
m_1 &: [\![\Delta \Vdash M_1 : (\Gamma, x : \alpha \vdash \alpha_2 \to \alpha_1)]\!] \\
m_2 &: [\![\Delta \Vdash M_2 : (\Gamma, x : \alpha \vdash \alpha_2)]\!]
\end{aligned}$$

Then, on one hand, the following equation holds.

$$\begin{aligned}
&[\![\Delta \Vdash (M_1 M_2)[L/x] : (\Gamma \vdash \alpha_1)]\!] \\
&= [\![\Delta \Vdash (M_1[L/x])(M_2[L/x]) : (\Gamma \vdash \alpha_1)]\!] \\
&= \mathcal{C}([\![\Gamma]\!], ev) \circ \ltimes_{[\![\Gamma]\!]} \circ \\
&\quad \langle [\![\Delta \Vdash M_1[L/x] : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!], [\![\Delta \Vdash M_2[L/x] : (\Gamma \vdash \alpha_2)]\!] \rangle \\
&= \mathcal{C}([\![\Gamma]\!], ev) \circ \ltimes_{[\![\Gamma]\!]} \circ \\
&\quad \langle \|\Delta \Vdash M_1 : (\Gamma, x : \alpha \vdash \alpha_2 \to \alpha_1)\|_l^{\Gamma, x:\alpha}, \|\Delta \Vdash M_2 : (\Gamma, x : \alpha \vdash \alpha_2)\|_l^{\Gamma, x:\alpha} \rangle \\
&= \mathcal{C}([\![\Gamma]\!], ev) \circ \ltimes_{[\![\Gamma]\!]} \circ \langle \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle m_1, l \rangle, \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle m_2, l \rangle \rangle \text{ — (†)}
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash M_1 M_2 : (\Gamma, x : \alpha \vdash \alpha_1)]\!], l \rangle \\
&= \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle \mathcal{C}([\![\Gamma]\!] \times [\![\alpha]\!], ev) \circ \ltimes_{[\![\Gamma]\!] \times [\![\alpha]\!]} \circ \langle m_1, m_2 \rangle, l \rangle \text{ — (‡)}
\end{aligned}$$

Suppose that m_1 and m_2 map $h_1, \ldots, h_d$ to $g_1 : [\![\Gamma]\!] \times [\![\alpha]\!] \longrightarrow [\![\alpha_1]\!]^{[\![\alpha_2]\!]}$ and $g_2 : [\![\Gamma]\!] \times [\![\alpha]\!] \longrightarrow [\![\alpha_2]\!]$, respectively.
Then (†) maps the d-tuple to $ev \circ \langle g_1 \circ \langle id_{[\![\Gamma]\!]}, f \rangle, g_2 \circ \langle id_{[\![\Gamma]\!]}, f \rangle \rangle$ and (‡) maps to $ev \circ \langle g_1, g_2 \rangle \circ \langle id_{[\![\Gamma]\!]}, f \rangle = ev \circ \langle g_1 \circ \langle id_{[\![\Gamma]\!]}, f \rangle, g_2 \circ \langle id_{[\![\Gamma]\!]}, f \rangle \rangle$. Therefore, (†) and (‡) are the same morphisms.

- If $M = p_i(M)$, the equation is proved as follows:

$$\begin{aligned}
&[\![\Delta \Vdash (p_i(M))[L/x] : (\Gamma \vdash \alpha')]\!] \\
&= [\![\Delta \Vdash p_i(M[L/x]) : (\Gamma \vdash \alpha')]\!] \\
&= \pi_i \circ [\![\Delta \Vdash M[L/x] : \tau_1 \otimes \cdots \otimes (\Gamma \vdash \alpha') \otimes \cdots \otimes \tau_k]\!] \\
&= \pi_i \circ \left\|\Delta \Vdash M : S_{\Gamma,\, x:\alpha}(\tau_1 \otimes \cdots \otimes (\Gamma \vdash \alpha') \otimes \cdots \otimes \tau_k)\right\|_l^{\Gamma, x:\alpha} \\
&= \pi_i \circ \left\|\Delta \Vdash M : S_{\Gamma,\, x:\alpha}(\tau_1) \otimes \cdots \otimes S_{\Gamma,\, x:\alpha}(\Gamma \vdash \alpha') \otimes \cdots \otimes S_{\Gamma,\, x:\alpha}(\tau_k)\right\|_l^{\Gamma, x:\alpha} \\
&= \pi_i \circ \langle \left\|\Delta \Vdash p_1(M) : S_{\Gamma,\, x:\alpha}(\tau_1)\right\|_l^{\Gamma, x:\alpha}, \ldots, \\
&\qquad \left\|\Delta \Vdash p_i(M) : (\Gamma, x : \alpha \vdash \alpha')\right\|_l^{\Gamma, x:\alpha}, \ldots, \left\|\Delta \Vdash p_k(M) : S_{\Gamma,\, x:\alpha}(\tau_k)\right\|_l^{\Gamma, x:\alpha} \rangle \\
&= \left\|\Delta \Vdash p_i(M) : (\Gamma, x : \alpha \vdash \alpha')\right\|_l^{\Gamma, x:\alpha}
\end{aligned}$$

- If $M = M_1 \between M_2$, the equation is proved as follows:

$$\begin{array}{ll} & [\![\Delta \Vdash (M_1 \between M_2)[L/x] : (\Gamma \vdash \alpha')]\!] \\ = & [\![\Delta \Vdash (M_1[L/x]) \between M_2 : (\Gamma \vdash \alpha')]\!] \\ = & ev \circ \langle [\![\Delta \Vdash M_1[L/x] : \tau \Rightarrow (\Gamma \vdash \alpha')]\!], [\![\Delta \Vdash M_2 : \tau]\!] \rangle \\ = & ev \circ \langle \|\Delta \Vdash M_1 : \tau \Rightarrow (\Gamma, x : \alpha \vdash \alpha')\|_l^{\Gamma,x:\alpha}, [\![\Delta \Vdash M_2 : \tau]\!] \rangle \\ = & ev \circ \langle \lambda(\|\Delta, X : \tau \Vdash M_1 \between X : (\Gamma, x : \alpha \vdash \alpha')\|_{l\circ\langle\pi_1,\ldots,\pi_d\rangle}^{\Gamma,x:\alpha}), [\![\Delta \Vdash M_2 : \tau]\!] \rangle \\ = & \|\Delta, X : \tau \Vdash M_1 \between X : (\Gamma, x : \alpha \vdash \alpha')\|_{l\circ\langle\pi_1,\ldots,\pi_d\rangle}^{\Gamma,x:\alpha} \circ \langle id_\Delta, [\![\Delta \Vdash M_2 : \tau]\!] \rangle \\ = & \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta, X : \tau \Vdash M_1 \between X : (\Gamma, x : \alpha \vdash \alpha')]\!], l \circ \langle \pi_1, \ldots, \pi_d \rangle \rangle \circ \\ & \quad \langle id_{[\![\Delta]\!]}, [\![\Delta \Vdash M_2 : \tau]\!] \rangle \\ = & \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta, X : \tau \Vdash M_1 \between X : (\Gamma, x : \alpha \vdash \alpha')]\!] \circ \langle id_{[\![\Delta]\!]}, [\![\Delta \Vdash M_2 : \tau]\!] \rangle, l \rangle \\ = & \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle ev \circ \langle \lambda([\![\Delta, X : \tau \Vdash M_1 \between X : (\Gamma, x : \alpha \vdash \alpha')]\!]), [\![\Delta \Vdash M_2 : \tau]\!] \rangle, l \rangle \\ = & \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle ev \circ \langle [\![\Delta \Vdash \zeta X.M_1 \between X : \tau \Rightarrow (\Gamma, x : \alpha \vdash \alpha')]\!], [\![\Delta \Vdash M_2 : \tau]\!] \rangle, l \rangle \\ =_{M\eta} & \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle ev \circ \langle [\![\Delta \Vdash M_1 : \tau \Rightarrow (\Gamma, x : \alpha \vdash \alpha')]\!], [\![\Delta \Vdash M_2 : \tau]\!] \rangle, l \rangle \\ = & \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash M_1 \between M_2 : (\Gamma, x : \alpha \vdash \alpha')]\!], l \rangle \end{array}$$

– If $\tau = \tau_2 \Rightarrow \tau_1$, the equation is proved as follows:

$$\begin{array}{ll} & [\![\Delta \Vdash M[L/x] : \tau_2 \Rightarrow \tau_1]\!] \\ =_{M\eta} & [\![\Delta \Vdash \zeta X.(M[L/x]) \between X : \tau_2 \Rightarrow \tau_1]\!] \\ = & \lambda([\![\Delta, X : \tau_2 \Vdash (M[L/x]) \between X : \tau_1]\!]) \\ = & \lambda([\![\Delta, X : \tau_2 \Vdash (M \between X)[L/x] : \tau_1]\!]) \\ = & \lambda(\|\Delta, X : \tau_2 \Vdash M \between X : S_{\Gamma,\ x:\alpha}(\tau_1)\|_{l\circ\langle\pi_1,\ldots,\pi_d\rangle}^{\Gamma,x:\alpha}) \\ = & \|\Delta \Vdash M : \tau_2 \Rightarrow S_{\Gamma,\ x:\alpha}(\tau_1)\|_l^{\Gamma,x:\alpha} \\ = & \|\Delta \Vdash M : S_{\Gamma,\ x:\alpha}(\tau_2 \Rightarrow \tau_1)\|_l^{\Gamma,x:\alpha} \end{array}$$

– If $\tau = munit$, the equation holds obviously since the morphism $!_{[\![\Delta]\!]}$ is unique.
– If $\tau = \tau_1 \otimes \cdots \otimes \tau_k$, the equation is proved as follows:

$$\begin{array}{l} [\![\Delta \Vdash M[L/x] : \tau_1 \otimes \cdots \otimes \tau_k]\!] \\ = [\![\Delta \Vdash \langle\!\langle p_1(M[L/x]), \ldots, p_k(M[L/x]) \rangle\!\rangle : \tau_1 \otimes \cdots \otimes \tau_k]\!] \\ = \langle [\![\Delta \Vdash p_1(M[L/x]) : \tau_1]\!], \ldots, [\![\Delta \Vdash p_k(M[L/x]) : \tau_k]\!] \rangle \\ = \langle [\![\Delta \Vdash (p_1(M))[L/x] : \tau_1]\!], \ldots, [\![\Delta \Vdash (p_k(M))[L/x] : \tau_k]\!] \rangle \\ = \langle \|\Delta \Vdash p_1(M) : S_{\Gamma,\ x:\alpha}(\tau_1)\|_l^{\Gamma,x:\alpha}, \ldots, \|\Delta \Vdash p_k(M) : S_{\Gamma,\ x:\alpha}(\tau_k)\|_l^{\Gamma,x:\alpha} \rangle \\ = \|\Delta \Vdash M : S_{\Gamma,\ x:\alpha}(\tau_1) \otimes \cdots \otimes S_{\Gamma,\ x:\alpha}(\tau_k)\|_l^{\Gamma,x:\alpha} \\ = \|\Delta \Vdash M : S_{\Gamma,\ x:\alpha}(\tau_1 \otimes \cdots \otimes \tau_k)\|_l^{\Gamma,x:\alpha} \end{array}$$

Thus, $[\![\Delta \Vdash M[L/x] : \tau]\!] = \|\Delta \Vdash M : S_{\Gamma,x:\alpha}(\tau)\|_l^{\Gamma,x:\alpha}$ holds.[10]

[10]Using meta-level η-conversion in the proof is not prohibited since the proof of the conversion is independent of this lemma.

Theorem 9 **(Soundness of Base-level α-Conversion)**

$$(\alpha)\,\frac{\Delta \Vdash M : (\Gamma, x : \alpha_2 \vdash \alpha_1) \qquad y \notin fv(M)}{[\![\Delta \Vdash \lambda x.M : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!] = [\![\Delta \Vdash \lambda y.M[y/x] : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!]}$$

Proof Suppose that $|\Delta| = d$, $|\Gamma| = n$ and let m be a morphism $[\![\Delta \Vdash M : (\Gamma, x : \alpha_2 \vdash \alpha_1)]\!]$.

$$[\![\Delta \Vdash \lambda x.M : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!] = \lambda_{[\![\Gamma]\!]} \circ m$$

$$\begin{aligned}
&[\![\Delta \Vdash \lambda y.M[y/x] : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!] \\
&= \lambda_{[\![\Gamma]\!]} \circ [\![\Delta \Vdash M[y/x] : (\Gamma, y : \alpha_2 \vdash \alpha_1)]\!] \\
&= \lambda_{[\![\Gamma]\!]} \circ \mathcal{S}ub_{[\![\Gamma]\!] \times [\![\alpha_2]\!]} \circ \langle \mathcal{C}(\langle \pi_1, \ldots, \pi_n, \pi_{n+2}\rangle, [\![\alpha_1]\!]) \circ m, tp_{[\![\alpha_2]\!]}(\pi_{n+1}) \circ !_{[\![\Delta]\!]}\rangle
\end{aligned}$$

Next, suppose that m maps a d-tuple $h_1, \ldots, h_d \in [\![\Delta]\!]$ to $f : [\![\Gamma]\!] \times [\![\alpha_2]\!] \longrightarrow [\![\alpha_1]\!]$. Then, the left side of the equation maps the d-tuple to $\lambda(f)$. On the other hand,

$$\begin{aligned}
&\lambda(f \circ \langle \pi_1, \ldots, \pi_n, \pi_{n+2}\rangle \circ \langle id_{[\![\Gamma]\!] \times [\![\alpha_2]\!]}, \pi_{n+1}\rangle) \\
=&\lambda(f \circ \langle \pi_1, \ldots, \pi_n, \pi_{n+1}\rangle) \\
=&\lambda(f)
\end{aligned}$$

Thus, $[\![\Delta \Vdash \lambda x.M : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!] = [\![\lambda y.M[y/x] : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!]$ holds.

Theorem 10 **(Soundness of Base-level β-Conversion)**

$$(\beta)\,\frac{\Delta \Vdash M_1 : (\Gamma, x : \alpha_2 \vdash \alpha_1) \qquad \Delta \Vdash M_2 : (\Gamma \vdash \alpha_2)}{[\![\Delta \Vdash (\lambda x.M_1)M_2 : (\Gamma \vdash \alpha_1)]\!] = [\![\Delta \Vdash M_1[M_2/x] : (\Gamma \vdash \alpha_1)]\!]}$$

Proof Suppose that $|\Delta| = d$ and let m_1, m_2 be the following morphisms:

$$\begin{aligned}
m_1 &: [\![\Delta \Vdash M_1 : (\Gamma, x : \alpha_2 \vdash \alpha_1)]\!] : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!] \times [\![\alpha_2]\!], [\![\alpha_1]\!]) \\
m_2 &: [\![\Delta \Vdash M_2 : (\Gamma \vdash \alpha_2)]\!] : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!], [\![\alpha_2]\!])
\end{aligned}$$

Suppose that m_1 and m_2 map a d-tuple $h_1, \ldots, h_d \in [\![\Delta]\!]$ to $f : [\![\Gamma]\!] \times [\![\alpha_2]\!] \longrightarrow [\![\alpha_1]\!]$ and $g : [\![\Gamma]\!] \longrightarrow [\![\alpha_2]\!]$, respectively. Then, the following equation holds.

$$\begin{aligned}
&[\![\Delta \Vdash (\lambda x.M_1)M_2 : (\Gamma \vdash \alpha_1)]\!] \\
&= \mathcal{C}([\![\Gamma]\!], ev) \circ \ltimes_{[\![\Gamma]\!]} \circ \langle [\![\Delta \Vdash \lambda x.M_1 : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!], m_2 \rangle \\
&= \mathcal{C}([\![\Gamma]\!], ev) \circ \ltimes_{[\![\Gamma]\!]} \circ \langle \lambda_{[\![\Gamma]\!]} \circ m_1, m_2 \rangle \text{ — } (\dagger)
\end{aligned}$$

Recall that $\lambda_{[\![\Gamma]\!]}$, $\ltimes_{[\![\Gamma]\!]}$ and $\mathcal{C}([\![\Gamma]\!], ev)$ map morphisms in the following way:

$$\begin{aligned}
\lambda_{[\![\Gamma]\!]} &: & f &\longmapsto \lambda(f) \\
\ltimes_{[\![\Gamma]\!]} &: & (\lambda(f), g) &\longmapsto \langle \lambda(f), g \rangle \\
\mathcal{C}([\![\Gamma]\!], ev) &: & \langle \lambda(f), g \rangle &\longmapsto ev \circ \langle \lambda(f), g \rangle
\end{aligned}$$

Since the following equation holds, the morphism (†) maps the d-tuple to a morphism $f \circ \langle id, g \rangle : [\![\Gamma]\!] \longrightarrow [\![\alpha_1]\!]$.

$$\begin{aligned}
ev \circ \langle \lambda(f), g \rangle &= ev \circ (\lambda(f) \times id_{[\![\alpha_2]\!]}) \circ \langle id_{[\![\Gamma]\!]}, g \rangle \\
&= f \circ \langle id_{[\![\Gamma]\!]}, g \rangle
\end{aligned}$$

On the other hand, the following equation holds.

$$\begin{aligned}
[\![\Delta \Vdash M_1[M_2/x] : (\Gamma \vdash \alpha_1)]\!] &= \| \Delta \Vdash M_1 : (\Gamma, x : \alpha_2 \vdash \alpha_1) \|^{\Gamma, x:\alpha_2}_{m_2} \\
&= \mathcal{S}ub_{[\![\Gamma]\!]} \circ \langle m_1, m_2 \rangle \text{ — } (\ddagger)
\end{aligned}$$

Let us also recall that $\mathcal{S}ub_{[\![\Gamma]\!]} : (f, g) \longmapsto f \circ \langle id_{[\![\Gamma]\!]}, g \rangle$. Thus, the morphism (‡) maps the d-tuple to a morphism $f \circ \langle id, g \rangle : [\![\Gamma]\!] \longrightarrow [\![\alpha_1]\!]$, which is equivalent to (†). Therefore, the following equation holds.

$$[\![\Delta \Vdash (\lambda x.M_1)M_2 : (\Gamma \vdash \alpha_1)]\!] = [\![\Delta \Vdash M_1[M_2/x] : (\Gamma \vdash \alpha_1)]\!]$$

Theorem 11 (Soundness of Base-level η-Conversion)

$$(\eta)\,\frac{\Delta \Vdash M : (\Gamma \vdash \alpha_2 \to \alpha_1) \qquad x \notin fv(M)}{[\![\Delta \Vdash \lambda x.Mx : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!] = [\![\Delta \Vdash M : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!]}$$

Proof Suppose that $|\Delta| = d$, $|\Gamma| = n$, and let m be a morphism $[\![\Delta \Vdash M : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!] : [\![\Delta]\!] \longrightarrow \mathcal{C}([\![\Gamma]\!].[\![\alpha_1]\!]^{[\![\alpha_2]\!]})$, which maps a d-tuple $h_1, \ldots, h_d \in [\![\Delta]\!]$ to $f : [\![\Gamma]\!] \longrightarrow [\![\alpha_1]\!]^{[\![\alpha_2]\!]}$

$$
\begin{aligned}
&[\![\Delta \Vdash \lambda x.Mx : (\Gamma \vdash \alpha_2 \to \alpha_1)]\!] \\
&= \lambda_{[\![\Gamma]\!]} \circ [\![\Delta \Vdash Mx : (\Gamma, x : \alpha_2 \vdash \alpha_1)]\!] \\
&= \lambda_{[\![\Gamma]\!]} \circ \mathcal{C}([\![\Gamma]\!] \times [\![\alpha_2]\!], ev) \circ \ltimes_{[\![\Gamma]\!] \times [\![\alpha_2]\!]} \circ \\
&\quad \langle [\![\Delta \Vdash M : (\Gamma, x : \alpha_2 \vdash \alpha_2 \to \alpha_1)]\!], [\![\Delta \Vdash x : (\Gamma, x : \alpha_2 \vdash \alpha_2)]\!] \rangle \\
&= \lambda_{[\![\Gamma]\!]} \circ \mathcal{C}([\![\Gamma]\!] \times [\![\alpha_2]\!], ev) \circ \ltimes_{[\![\Gamma]\!] \times [\![\alpha_2]\!]} \circ \\
&\quad \langle \mathcal{C}(\langle \pi_1, \ldots, \pi_n \rangle, [\![\alpha_1]\!]^{[\![\alpha_2]\!]}) \circ m, tp_{[\![\alpha_2]\!]}(\pi_{n+1}) \circ !_{[\![\Delta]\!]} \rangle \text{ — } (\dagger)
\end{aligned}
$$

$$
\begin{aligned}
tp_{[\![\alpha_2]\!]}(\pi_{n+1}) \circ !_{[\![\Delta]\!]} &: h_1, \ldots, h_d \longmapsto \pi_{n+1} \\
\mathcal{C}(\langle \pi_1, \ldots, \pi_n \rangle, [\![\alpha_1]\!]^{[\![\alpha_2]\!]}) &: f \longmapsto f \circ \langle \pi_1, \ldots, \pi_n \rangle \\
\ltimes_{[\![\Gamma]\!] \times [\![\alpha_2]\!]} &: (f \circ \langle \pi_1, \ldots, \pi_n \rangle, \pi_{n+1}) \longmapsto \langle f \circ \langle \pi_1, \ldots, \pi_n \rangle, \pi_{n+1} \rangle \\
\mathcal{C}([\![\Gamma]\!] \times [\![\alpha_2]\!], ev) &: \langle f \circ \langle \pi_1, \ldots, \pi_n \rangle, \pi_{n+1} \rangle \longmapsto ev \circ \langle f \circ \langle \pi_1, \ldots, \pi_n \rangle, \pi_{n+1} \rangle
\end{aligned}
$$

Thus, (†) maps the d-tuple to

$$
\begin{aligned}
&\lambda(ev \circ \langle f \circ \langle \pi_1, \ldots, \pi_n \rangle, \pi_{n+1} \rangle) \\
&= \lambda(ev \circ (f \times id_{[\![\alpha_2]\!]}) \circ \langle \pi_1, \ldots, \pi_n, \pi_{n+1} \rangle) \\
&= \lambda(ev \circ (f \times id_{[\![\alpha_2]\!]})) \\
&= f
\end{aligned}
$$

Therefore, (†) and m are identical morphisms.

7 Soundness of Meta-level Substitution and Conversions

Lemma 2 (Meta-level Substitution Lemma)

$$
\frac{\begin{array}{c}[\![\Delta, X : \tau' \Vdash M : \tau]\!] = m : [\![\Delta]\!] \times [\![\tau']\!] \longrightarrow [\![\tau]\!] \\ [\![\Delta \Vdash L : \tau']\!] = \ l \ : [\![\Delta]\!] \longrightarrow [\![\tau']\!]\end{array}}{[\![\Delta \Vdash M[L/X] : \tau]\!] = m \circ \langle id_{[\![\Delta]\!]}, l \rangle : [\![\Delta]\!] \longrightarrow [\![\tau]\!]}
$$

Proof By induction on the structure of the term M.

The substitution rule for meta-level variables is thus immune with respect to the binding of (base-level) variables, as the following equation implies.

(23) $$(\zeta X.\lambda x.X) \between x = (\lambda x.X)[x/X] = \lambda x.x \qquad (1)$$

This means that MLC has a term that corresponds to the following map at the level of the object language, which is not the case in STLC.

(24) $$\phi \longmapsto \lambda x.\phi \qquad (2)$$

Just as $\alpha/\beta/\eta$-conversions are sound in STLC, meta-level $\alpha/\beta/\eta$-conversions are sound in MLC.

Theorem 12 (**Soundness of Meta-level α -Conversion**)

$$(M\alpha)\frac{\Delta, X : \tau_2 \Vdash M : \tau_1 \qquad Y \notin fmv(M)}{[\![\Delta \Vdash \zeta X.M : \tau_2 \Rightarrow \tau_1]\!] = [\![\Delta \Vdash \zeta Y.M[Y/X] : \tau_2 \Rightarrow \tau_1]\!]}$$

Proof Suppose that $|\Delta| = d$.

$$\begin{aligned}
&[\![\Delta \Vdash \zeta X.M : \tau_2 \Rightarrow \tau_1]\!] \\
&= \lambda([\![\Delta, X : \tau_2 \Vdash M : \tau_1]\!]) \\
&= \lambda([\![\Delta, X : \tau_2 \Vdash M : \tau_1]\!] \circ \langle \pi_1, \ldots, \pi_{d+1} \rangle) \\
&= \lambda([\![\Delta, X : \tau_2 \Vdash M : \tau_1]\!] \circ \langle \pi_1, \ldots, \pi_d, \pi_{d+2} \rangle \circ \langle id_{[\![\Delta]\!] \times [\![\tau_2]\!]}, \pi_{d+1} \rangle) \\
&= \lambda([\![\Delta, Y : \tau_2, X : \tau_2 \Vdash M : \tau_1]\!] \circ \langle id_{[\![\Delta]\!] \times [\![\tau_2]\!]}, [\![\Delta, Y : \tau_2 \Vdash Y : \tau_2]\!] \rangle) \\
&= \lambda([\![\Delta, Y : \tau_2 \Vdash M[Y/X] : \tau_1]\!]) \\
&= [\![\Delta \Vdash \zeta Y.M[Y/X] : \tau_2 \Rightarrow \tau_1]\!]
\end{aligned}$$

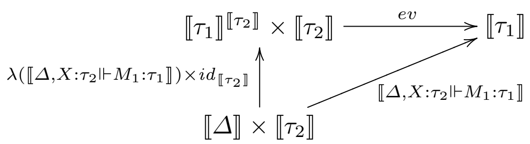
Theorem 13 (**Soundness of Meta-level β-Conversion**)

$$(M\beta)\frac{\Delta, X : \tau_2 \Vdash M_1 : \tau_1 \qquad \Delta \Vdash M_2 : \tau_2}{[\![\Delta \Vdash (\zeta X.M_1) \lightning M_2 : \tau_1]\!] = [\![\Delta \Vdash M_1[M_2/X] : \tau_1]\!]}$$

This is proved by a standard triangular identity in Cartesian closed categories.

Proof

$$\begin{aligned}
&[\![\Delta \Vdash (\zeta X.M_1) \lightning M_2 : \tau_1]\!] \\
&= ev \circ \langle [\![\Delta \Vdash \zeta X.M_1 : \tau_2 \Rightarrow \tau_1]\!], [\![\Delta \Vdash M_2 : \tau_2]\!] \rangle \\
&= ev \circ \langle \lambda([\![\Delta, X : \tau_2 \Vdash M_1 : \tau_1]\!]), [\![\Delta \Vdash M_2 : \tau_2]\!] \rangle \\
&= [\![\Delta, X : \tau_2 \Vdash M_1 : \tau_1]\!] \circ \langle id_{[\![\Delta]\!]}, [\![\Delta \Vdash M_2 : \tau_2]\!] \rangle \\
&= [\![\Delta \Vdash M_1[M_2/X] : \tau_1]\!]
\end{aligned}$$



Theorem 14 (Soundness of Meta-level η-Conversion)

$$(M\eta)\frac{\Delta \Vdash M : \tau_2 \Rightarrow \tau_1 \qquad X \notin fmv(M)}{[\![\Delta \Vdash \zeta X.M \lightning X : \tau_2 \Rightarrow \tau_1]\!] = [\![\Delta \Vdash M : \tau_2 \Rightarrow \tau_1]\!]}$$

Proof Suppose that $|\Delta| = d$.

$$\begin{aligned}
&[\![\Delta \Vdash \zeta X.M \lightning X : \tau_2 \Rightarrow \tau_1]\!] \\
&= \lambda([\![\Delta, X : \tau_2 \Vdash M \lightning X : \tau_1]\!]) \\
&= \lambda(ev \circ \langle [\![\Delta, X : \tau_2 \Vdash M : \tau_2 \Rightarrow \tau_1]\!], [\![\Delta, X : \tau_2 \Vdash X : \tau_2]\!] \rangle) \\
&= \lambda(ev \circ \langle [\![\Delta \Vdash M : \tau_2 \Rightarrow \tau_1]\!] \circ \langle \pi_1, \ldots, \pi_d \rangle, \pi_{d+1} \rangle) \\
&= \lambda(ev \circ ([\![\Delta \Vdash M : \tau_2 \Rightarrow \tau_1]\!] \times id_{[\![\tau_2]\!]}) \circ \langle \pi_1, \ldots, \pi_d, \pi_{d+1} \rangle) \\
&= \lambda(ev \circ ([\![\Delta \Vdash M : \tau_2 \Rightarrow \tau_1]\!] \times id_{[\![\tau_2]\!]})) \\
&= [\![\Delta \Vdash M : \tau_2 \Rightarrow \tau_1]\!]
\end{aligned}$$

8 Conclusion

We have presented a revised syntax, an equational theory, and a revised categorical semantics of MLC, by which we fixed several problems faced by previous formulations of MLC (Bekki 2009; Bekki and Asai 2010), and proved the soundness of the theory, including base-level/meta-level substitution and $\alpha/\beta/\eta$-conversions.

By virtue of this revision, we provide a more reliable foundation for the research based on monadic translations in terms of internal monads in Bekki (2009), Bekki and Asai (2010), Hayashishita and Bekki (2011), which enables us to maintain and further develop the result of those analyses.

MLC promises to find application in at least two areas: in linguistics, we are ready to investigate other linguistic phenomena to which monadic analyses may yield appropriate structures; and in theory of programming languages, we may be able to apply MLC to such tasks as code generation and partial evaluation, among others.

References

Bekki, D. (2009). Monads and meta-lambda calculus. In H. Hattori, T. Kawamura, T. Ide', M. Yokoo & Y. Murakami (Eds.), *New Frontiers in Artificial Intelligence Conference and Workshops, (JSAI 2008), Asahikawa, Japan, June 2008, Revised Selected Papers from LENLS5* (Vol. LNAI 5447, pp. 193–208), Springer.

Bekki, D., & Asai, K. (2010). Representing covert movements by delimited continuations. In K. Nakakoji, Y. Murakami & E. McCready (Eds.), *New Frontiers in Artificial Intelligence JSAI-isAI 2009 Workshops, Tokyo, Japan, November 2009, Selected Papers from LENLS7* (Vol. LNAI 6284, pp. 161–180). Heidelberg: Springer.

Crole, R. L. (1993) *Categories for types*. Cambridge: Cambridge University Press.

Danvy, O., & Filinski, A. (1990) Abstracting control. *LFP90, the 1990 ACM Conference on Lisp and Functional Programming* (pp. 151–160).

Hayashishita, J. R., & Bekki, D. (2011). Conjoined nominal expressions in Japanese: Interpretation through monad, In *The Eighth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS8)* (pp. 139–152). JSAI International Symposia on AI 2011, Takamatsu, Kagawa, Japan.

Kock, A. (1970). Strong functors and monoidal monads. Various Publications Series 11, Aarhus Universitet, August 1970.

Lambek, J. (1980). From λ-calculus to cartesian closed categories. In: J. P. Seldin & J. R. Hindley (Eds.), *To H. B. Curry: Essays on combinatory logic, lambda calculus and formalism*. New York: Academic Press.

Lambek, J., & Scott, P. J. (1986). *Introduction to higher order categorical logic*. Cambridge: Cambridge University Press.

MacLane, S. (1997). *Categories for the working mathematician. Graduate texts in mathematics* (2nd ed.). New Yok: Springer.

Masuko, M., & Bekki, D. (2011). Categorical semantics of meta-lambda calculus. *The 13th JSSST Workshop on Programming and Programming Languages (PPL2011) (in Japanese)* (pp. 60–74), Joozankei, Japan.

Moggi, E. (1989). Computational lambda-calculus and monads. *Fourth Annual IEEE Symposium on Logic in Computer, Science* (pp. 14–23).

Ogata, N. (2008). Towards computational non-associative lambek lambda-calculi for formal pragmatics. *The Fifth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS2008) in Conjunction with the 22nd Annual Conference of the Japanese Society for Artificial Intelligence* (pp. 79–102), Asahikawa, Japan.

Shan, C. C. (2001). Monads for natural language semantics. In K. Striegnitz (Ed.), The ESSLLI-2001 student session (13th European summer school in logic, language and information, 2001). pp. 285–298.

Unger, C. (2011). Dynamic semantics as monadic computation. *The Eighth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS8)* (pp. 153–164). JSAI International Symposia on AI 2011, Takamatsu, Kagawa, Japan.

Wadler, P. (1992). Comprehending monads. *Mathematical structure in computer science* (Vol. 2, pp. 461–493). Cambridge: Cambridge University Press. (an ealier version appears in Conference on Lisp and Functional Programming, Nice, France, ACM, June 1990).

Coordinating and Subordinating Binding Dependencies

Alastair Butler

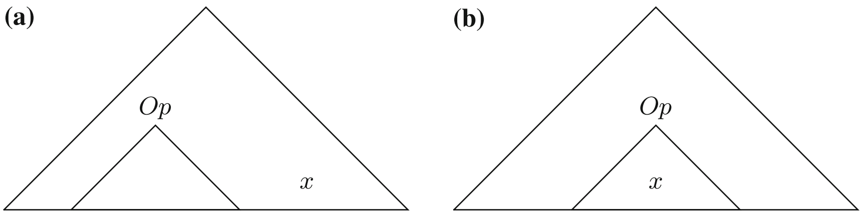
Abstract This chapter focuses on similarities between coordinating and (distant) subordinating binding dependencies. We start from natural language data suggesting such dependencies are established by the same underlying mechanism, e.g., "*A collector didn't buy because she was influenced.*" is structurally ambiguous without consequences for the pronominal binding. We compare and contrast four related systems that capture coordinating and subordinating binding dependencies, the first with distinct mechanisms, the others with single mechanisms.

Keywords Coordination · Subordination · Anaphoric binding · Covaluation · Clause internal relations · Pronouns

1 Introduction

This chapter focuses on the differences and similarities between the dependency types pictured in (1), where Op is some form of binding operator and x is a bindee dependent on Op. In (1a) x is outside and positioned to the right of the syntactic scope of Op. Let us call this a binding dependency of *coordination*. In (1b) x is within the syntactic scope of Op. Let us call this a binding dependency of *subordination*.

(1)
(a) **(b)**



A. Butler (✉)
PRESTO, Japan Science and Technology Agency, Center for the Advancement of Higher Education, Tohoku University, Miyagi, Japan
e-mail: ajb129@hotmail.com

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*, Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_4,

While (1) suggests no reason to expect the two dependency types to be of the same kind, English appears to employ the same pronominal mechanism to establish both dependency types. For example, consider (2) which allows the distinct readings of (3). The ambiguity hinges on the scope placement of negation with respect to *because*, with *she* anaphorically dependent on *a collector* for both readings.

(2) A collector didn't buy because she was influenced.

(3) a. For the reason stated, a collector didn't make the purchase.
(*because* > *neg*)
b. A collector made the purchase, for a reason not yet stated.
(*neg* > *because*)

Reading (3a) is captured with the bracketing of (4a) that conforms to the picture of (1a). In contrast reading (3b) is only possible with the bracketing of (4b) that has the attached adjunct clause falling under the scope of negation and so under the scope of the main clause subject *a collector* to conform to the picture of (1b).

(4) a. [A collector didn't buy] because [she was influenced].
b. A collector didn't [buy because [she was influenced]].

Other examples to stress the similarity of coordinating and subordinating binding dependencies come from observing covaluation effects. Effects of covaluation readily arise across sentences in discourse and so as coordinating binding dependencies. Consider (5), a binding dependency established with a mandatory reflexive. (Accounting for reflexive binding is addressed in Sects. 4 and 5.) Typically, *him* should not be able to occur in the same environment as *himself* and take the same referent; yet it does in (6). This is possible because the antecedent of *him* is not the subject *John himself*, but rather the previous *John* in A's question. This gives an instance of covaluation: that *him* is coreferential with the subject is not the consequence of a binding localised to the current clause, and so does not violate the restriction that bars a pronoun occurring where a reflexive is able to occur.

(5) **John** voted for **himself**.

(6) A: Who voted for **John**?
B: Well, John himself voted for **him**, but I doubt others did.

What is interesting for our current concern is the observation made in Heim (1993) (see also Reinhart 2000; Büring 2005) that covaluation effects can arise under the scope of a quantifier and so with a subordinating binding dependency. What is required for the covaluation effect is for the dependency to be sufficiently deeply embedded for the bindee to take the from of a non-reflexive pronoun. For example, (7) allows for a covaluation reading where every candidate is surprised to find 'No one else voted for me!'.

(7) Every candidate is surprised because only he voted for him.

That the phenomena of covaluation is observable with both subordinating and coordinating binding dependencies, and furthermore arises with the employment of the same lexical items in English, is very suggestive that a single unified mechanism of pronominal binding is responsible for both dependency types, in English at least. Whatever explains the coordinating binding dependencies in (4a) and (6) should bear on what explains the subordinating binding dependencies in (4b) and (7), and vice versa.

The remainder of this chapter is organised as follows. Section 2 gives our starting point in formulating an account by presenting *Predicate Logic with Anaphora* (PLA) from Dekker (2002). This implements the two dependency types with distinct binding mechanisms. Section 3 looks at a system built from the parts of PLA that are not predicate logic, offering a way to derive subordinating and coordinating binding dependencies with a single mechanism. Section 4 looks at *Dynamic Predicate Logic with Exit operators* (DPLE) from Vermeulen (1993). This offers an alternative unified way to capture the two dependency types, with primitive coordinating relations and derived subordinating effects. Section 5 offers yet another system, a minimal version of the *Scope Control Theory* of Butler (2010), which brings together insights gained from the other systems.

Throughout this chapter we make use of the following notation for sequences:

- $[x_0, \ldots, x_{n-1}]$ a sequence with n elements, x_0 being frontmost
- $\hat{x}$ abbreviation for a sequence
- $\hat{x}_i$ returns the i-th element of a sequence: $[x_0, \ldots, x_{n-1}]_i = x_i$, where $0 \leq i < n$.
- $|\hat{x}|$ returns the length of a sequence: $|[x_0, \ldots, x_{n-1}]| = n$.
- $\hat{x} \mathbin{@} \hat{y}$ returns the concatenation of sequences $\hat{x}$ and $\hat{y}$.

2 Distinct Coordinating and Subordinating Mechanisms

Suppose we wish to encode the discourse of (8) with predicate logic.

(8) Someone^1 enters. He$_1$ smiles. Someone^2 laughs. She$_2$ likes him$_1$.

One possibility is given by (9).

(9) $\exists x(\text{enters}(x) \wedge \text{smiles}(x) \wedge \exists y(\text{laughs}(y) \wedge \text{likes}(y, x)))$

While (9) captures the truth-conditional semantics of (8), it does so with existential quantifiers corresponding to *someone* taking syntactic scope over the remaining discourse. This has the shortcoming of leaving no separable encoding for the sentences of (8), a situation that appears forced because predicate logic supports only subordinating binding dependencies.

This section considers *Predicate Logic with Anaphora* (PLA) from Dekker (2002). A central design goal of PLA is to allow a direct coding of coordinating binding dependencies with minimal deviation from a predicate logic core (that captures subordinating binding dependencies). The language of PLA is that of predicate logic with atomic formulas taking, in addition to variables, "pronouns" $\{p_0, p_1, p_2, \ldots\}$ as terms.

Definition 1 (*PLA satisfaction and truth*).

Suppose a first-order model M with domain of individuals D. Suppose $\hat{\sigma}$ is a sequence of individuals from D. Suppose g is an assignment from variables to individuals of D. $g[x/d]$ is an assignment that is like g in all respects, except (possibly) differing with d assigned to x. $\texttt{drop}(\hat{\sigma}, n)$ returns what is left after dropping the first n elements of $\hat{\sigma}$ for $0 \leq n < |\hat{\sigma}|$. The PLA semantics can be given as follows:

- $M, \hat{\sigma}, g \models \exists x\phi$ iff $M, \texttt{drop}(\hat{\sigma}, 1), g[x/\hat{\sigma}_0] \models \phi$
- $M, \hat{\sigma}, g \models \phi \wedge \psi$ iff $M, \texttt{drop}(\hat{\sigma}, n(\psi)), g \models \phi$ and $M, \hat{\sigma}, g \models \psi$
- $M, \hat{\sigma}, g \models \neg\phi$ iff there is no $\hat{\sigma}' \in D^{n(\phi)}$ such that $M, \hat{\sigma}'\hat{\sigma}, g \models \phi$
- $M, \hat{\sigma}, g \models \mathrm{P}(t_1, \ldots, t_n)$ iff $([t_1]_{\hat{\sigma},g}, \ldots, [t_n]_{\hat{\sigma},g}) \in M(\mathrm{P})$

and

- $[x]_{\hat{\sigma},g} = g(x)$
- $[p_i]_{\hat{\sigma},g} = \hat{\sigma}_i$

where $n(\phi)$ is a count of existentials in ϕ that are outside the scope of negation: $n(\exists x\phi) = n(\phi) + 1$, $n(\phi \wedge \psi) = n(\phi) + n(\psi)$, $n(\neg\phi) = 0$, $n(\mathrm{P}(t_1, \ldots, t_n)) = 0$.

A formula ϕ is *true* with respect to $\hat{\sigma}$ and g in M iff there is a $\hat{\sigma}' \in D^{n(\phi)}$ such that $M, \hat{\sigma}'\hat{\sigma}, g \models \phi$.

As with the classical interpretation of predicate logic the clause for existential quantification brings about the creation of a new x binding by resetting the value assigned to x. However the reset value is not entirely *new* but is rather the dropped frontmost sequence value of $\hat{\sigma}$.

The clause for conjunction brings about a division between 'fresh' and 'old' sequence positions. Fresh positions are at the front of $\hat{\sigma}$ and occur so as to be dropped by an existential. Old positions occur towards the rear of $\hat{\sigma}$ and were dropped by existentials of prior conjuncts. This division falls out from $\phi \wedge \psi$ since for the evaluation of ϕ the fresh positions for ψ are dropped on the basis of an $n(\psi)$ count to reveal the fresh positions for ϕ, as well as the old positions. For the evaluation of ψ there are no drops with the consequence that the fresh and old positions for ϕ are old positions for ψ.

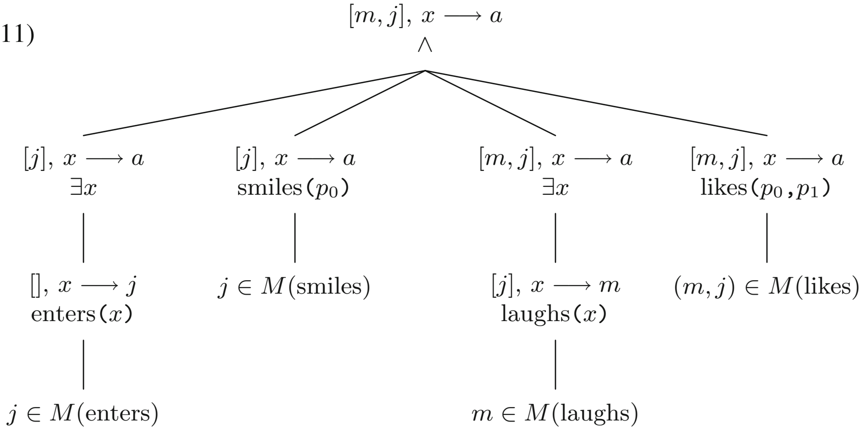
The negation of a formula ϕ acts as a 'test': it tells us that values for the existentials in ϕ cannot be found, or rather that there is no way to feed the existentials of ϕ values for their bindings from a sequence $\hat{\sigma}'$ with length $n(\phi)$. Consequently for a negated formula evaluated with respect to $\hat{\sigma}$, all positions of $\hat{\sigma}$ are old. Similarly for atomic formula evaluation all fresh positions will have dropped and so only old positions are left as potential values for pronouns as terms within the atomic formula.

As an example of PLA in action, (8), can be rendered as (10).

(10) $\exists x \text{enters}(x) \wedge \text{smiles}(p_0) \wedge \exists x \text{laughs}(x) \wedge \text{likes}(p_0, p_1)$

An evaluation of (10) is illustrated in (11), happening against the sequence of individuals $[m, j]$ and an assignment of a to variable x. Conjunction in PLA is associative, so $(\phi \wedge \psi) \wedge \chi \equiv \phi \wedge (\psi \wedge \chi)$, and to reflect this the illustration is left underspecified as to the order in which the instances of $\wedge$ apply.

(11)



We can see that during the course of evaluation occurrences of $\wedge$ manage when frontmost positions of the sequence are kept in reserve for subsequent discourse. If positions are not reserved, they are either: (i) destined to be dropped to have their values entered into the assignment as existential binding values before a 'test' is encountered (either a predicate or negated formula), as demonstrated by the instances of $\exists x$; or (ii) serve as accessible old positions for pronouns, as demonstrated by the first p_0 instance and p_1 taking as antecedent the sequence value that serves as the binding value adopted by the first instance of $\exists x$, and the second instance of p_0 taking as antecedent the sequence value that serves as the binding value adopted by the second $\exists x$.

The binding value for an existential quantification comes from the sequence $\hat{\sigma}$. Pronouns likewise take their antecedent from $\hat{\sigma}$. Since $\hat{\sigma}$ is an open environment parameter of evaluation, this gives (10) the same interpretation as the predicate logic formula (12).

(12) $\text{enters}(x_1) \wedge \text{smiles}(x_1) \wedge \text{laughs}(x_2) \wedge \text{likes}(x_2, x_1)$

This accords with the use of predicate logic in Cresswell (2002) to encode discourse while maintaining separable encodings for constituent sentences, e.g., (8) can be captured along the lines of (13).

(13) $\exists x(x = x_1 \wedge \text{enters}(x)) \wedge \text{smiles}(x_1) \wedge \exists x(x = x_2 \wedge \text{laughs}(x)) \wedge \text{likes}(x_2, x_1)$

Furthermore, Cresswell suggests that what we are seeing as an essentially variable use of indefinites is reflected in English by the word "namely", e.g., it is possible to say (14).

(14) Someone, namely John, enters. He smiles. Someone, namely Mary, laughs. She likes him.

With the illustration of (11) we also see that what is the antecedent of a pronoun in PLA is not necessarily determined by the same index. Rather the antecedent of a pronoun must be calculated by taking into account the number of occurrences of existentials in intervening conjuncts. This seems appropriate for characterising pronouns. Pronouns do not have grammatically fixed antecedents and must instead have their antecedent resolved, that is, chosen from potentially a number of accessible antecedents. The index provides notation to indicate a specific choice of antecedent. Moreover, under the assumption that a lower integer is in some sense easier to allocate than a higher integer, pronouns are seen to favour closer (more salient) antecedents.

We are now in a position to consider PLA renderings of the examples in Sect. 1. Reading (3a) of (2) can be captured with the PLA formula (15a) that has bracketing conforming to (4a), while reading (3b) is captured by (15b) with bracketing that conforms to (4b).

(15) a. $\exists x(\text{collector}(x) \wedge \neg\text{buy}(x)) \wedge \text{influenced}(p_0)$
b. $\exists x(\text{collector}(x) \wedge \neg(\text{buy}(x) \wedge \text{influenced}(x)))$

Due to the bracketing that makes the dependency with the pronoun a coordinating binding dependency in (15a) and a subordinating binding dependency in (15b) the encoding of the pronoun cannot be captured in a consistent manner: a dependency with the configuration of (1a) requires a PLA pronoun, while configuration (1b) necessitates a PLA variable.

Considering (5), the binding of the reflexive pronoun is captured with a PLA variable as in (16), while the covaluation data (6) necessitates a PLA pronoun as in (17).

(16) $\exists x(\text{John}(x) \wedge \text{voted\_for}(x, x))$

(17) $\exists x(\text{John}(x) \wedge \text{voted\_for}(x, p_1))$

Encodings (16) and (17) are appropriate for distinguishing covaluation from binding. PLA pronoun binding is essentially an ad-hoc dependency, with the actual dependency that obtains fixed with the allocation of an index for the pronoun—a change of index would allow a different antecedent. In contrast, the binding of a variable is determined solely by the scope under which the variable falls. This is appropriate for capturing dependencies that could not be otherwise, which holds for reflexive binding, but also for other clause internal dependencies, e.g., connecting subject and object arguments to the verb.

However the distinction between binding and covaluation achieved with (16) and (17) breaks down when we consider providing a PLA encoding for the covaluation reading of (7). Arising as the consequence of a subordinating configuration, the covaluation reading of (7) requires the pronoun to be coded with a PLA variable, as in (18).

(18) $\neg\exists x(\text{candidate}(x) \wedge \neg(\text{surprised}(x) \wedge \neg\exists y(\text{voted\_for}(y, x) \wedge \neg y = x)))$

3 Only a Pronoun Mechanism

Context is structured with PLA as a sequence that is not changed but which can be extended with insertions from instances of negation that are subsequently lost on exit from the scope of the negation, resulting in DRT-like accessibility effects (see e.g., Kamp and Reyle 1993). Dynamism emerges from an advancing through the given context as managed by instances of $\wedge$, or to put this another way, the unveiling of previously inaccessible portions of context. The current position reached comes from a division between sequence values in active use and sequence values not in active use. Values in active use are either values that will be used as the values for quantifications in the current conjunct, in which case they are dropped from the sequence machinery and imported into the variable binding machinery, or else are values accessible to pronouns, in which case they will have been used as the values for quantifications in prior conjuncts. Values not in active use are held back to serve as values for the quantifications of subsequent conjuncts.

We might imagine other ways of keeping track on where we are in a structured context. In this section we consider a system with a "pointer" to mark the location reached in a linearly structured context which we can take to be an infinite sequence. Having infinite sequences removes the possibility of undefinedness, but the system would operate the same over finite sequences with sufficient sequence values.

Definition 2 (*Pointer Semantics satisfaction and truth*).

Suppose a first-order model M with domain of individuals D. We will use $\hat{\sigma}[k..m]$ for finite sequence $(\hat{\sigma}_k, \ldots, \hat{\sigma}_m)$, $\hat{\sigma}[k..\omega]$ for the infinite sequence $(\hat{\sigma}_k, \ldots)$, and $\hat{\sigma}[-\omega..k]$ for the infinite sequence $(\ldots, \hat{\sigma}_k)$. We will write $\hat{\sigma}$ for $\hat{\sigma}[-\omega..\omega]$ and suppose individuals from D are assigned to the positions of $\hat{\sigma}$. Satisfaction is given as follows:

- $M, \hat{\sigma} \models_k \exists\phi$ iff $M, \hat{\sigma} \models_{k+1} \phi$
- $M, \hat{\sigma} \models_k \phi \wedge \psi$ iff $M, \hat{\sigma} \models_k \phi$ and $M, \hat{\sigma} \models_{k+n(\phi)} \psi$
- $M, \hat{\sigma} \models_k \neg\phi$ iff $M, \hat{\sigma}[-\omega..k]\hat{\sigma}'[1..n(\phi)]\hat{\sigma}[k+1..\omega] \not\models_k \phi$ for all $\hat{\sigma}'[1..n(\phi)]$
- $M, \hat{\sigma} \models_k P(t_1, \ldots, t_n)$ iff $([\![t_1]\!]_{\hat{\sigma},k}, \ldots, [\![t_n]\!]_{\hat{\sigma},k}) \in M(P)$
- $[\![p_i]\!]_{\hat{\sigma},k} = \hat{\sigma}_{k-i}$

where $n(\phi)$ is a count of the binding operators $\exists$ in ϕ that are outside the scope of negation: $n(\exists\phi) = n(\phi)+1, n(\phi\wedge\psi) = n(\phi)+n(\psi), n(\neg\phi) = 0, n(P(t_1, \ldots, t_n)) = 0.$

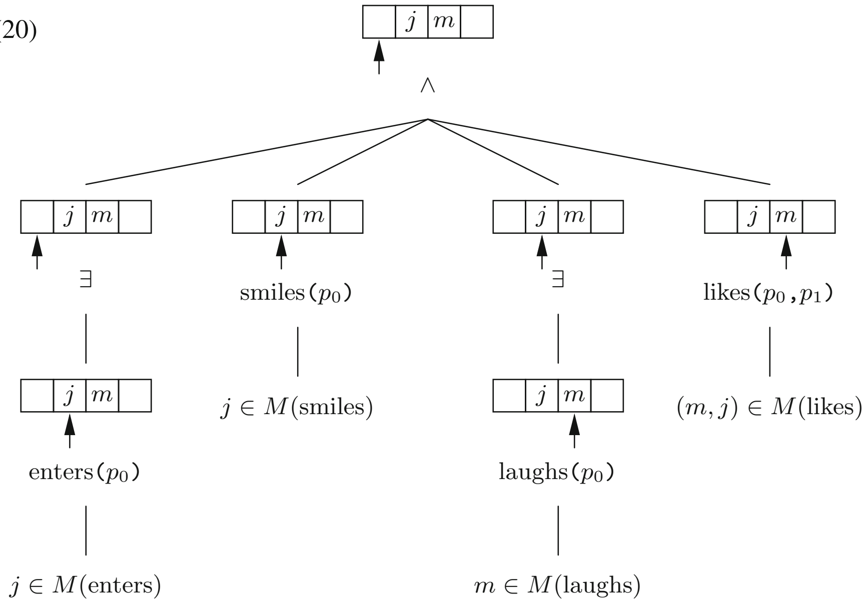
A formula ϕ is *true* with respect to k and $\hat{\sigma}$ in M iff there is a $\hat{\sigma}'[1..n(\phi)]$ such that $M, \hat{\sigma}[-\omega..k]\hat{\sigma}'[1..n(\phi)]\hat{\sigma}[k+1..\omega] \models_k \phi$.

Binding operator $\exists\phi$ opens a fresh binding by advancing the pointer by one sequence value, which is thereafter accessible for pronouns inside ϕ. With a conjunct $\phi \wedge \psi$, crossing to ψ from ϕ advances the pointer by $n(\phi)$ sequence values, that is, by the number of instances of $\exists$ visible in ϕ. This is demonstrated with a rendering of (8) as (19).

(19) $\exists$enters (p_0) $\wedge$ smiles (p_0) $\wedge$ $\exists$laughs (p_0) $\wedge$ likes (p_0, p_1)

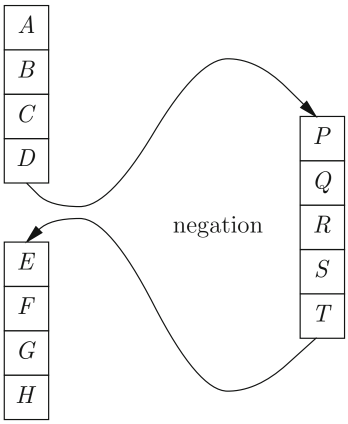
Evaluation of (19) is illustrated in (20), with the order in which the three instances of $\wedge$ apply left underspecified, since conjunction is associative, as was the case with PLA. Also note how the evaluation of terminals ends exactly as with (11).

(20)



Negation works as illustrated in (21) by inserting into the main sequence a sequence of values that serve as the positions to which the pointer is moved by existentials while inside the scope of negation, returning back to the main sequence outside the scope of negation.

(21)



We are now in a position to apply the system of Pointer Semantics to the examples of Sect. 1. Reading (3a) of (2) follows from (22a) with bracketing conforming to (4a), while reading (3b) is captured by (22b) with bracketing that conforms to (4b).

(22) a. $\exists(\text{collector}(p_0) \wedge \neg\text{buy}(p_0)) \wedge \text{influenced}(p_0)$
b. $\exists(\text{collector}(p_0) \wedge \neg(\text{buy}(p_0) \wedge \text{influenced}(p_0)))$

Reflexive binding in (5) is captured with (23), while covaluation readings for (6) and (7) are obtained with (24) and (25), respectively.

(23) $\exists(\text{John}(p_0) \wedge \text{voted\_for}(p_0, p_0))$

(24) $\exists(\text{John}(p_0) \wedge \text{voted\_for}(p_0, p_1))$

(25) $\neg\exists(\text{candidate}(p_0) \wedge\neg(\text{surprised}(p_0) \wedge\neg\exists(\text{voted\_for}(p_0, p_1) \wedge\neg p_0 = p_1)))$

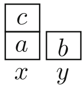
The encodings of (22)–(25) successfully capture the uniform role of linking played by pronouns in both subordinating and coordinating configurations. However this uniformity is achieved at the expense of turning all operator-variable dependencies into PLA-like pronominal bindings, including internal clause relations, such as subject arguments binding verbal predicates. As was noted when looking at PLA, PLA pronouns capture the ad-hoc nature of pronominal linking, with the actual dependency that holds relying on a choice of index with respect to the presence of intervening (formula) material. This is inappropriate for capturing core clause internal links, which need to be grammatically enforced and essentially unalterable dependencies, but see van Eijck (2001) where a constrained system of pronoun management is developed in a dynamic setting. What we appear to need is a system that has both the option of pronominal linking and variable name linking with subordinating configurations.

4 Only Coordinating Binding Dependencies

In this section we look at *Dynamic Predicate Logic with Exit operators* (DPLE) as introduced by Vermeulen (1993, 2000), Hollenberg and Vermeulen (1996), with the exception that we adopt predicate logic notation for the purpose of comparison. The system itself extends the system of Dynamic Predicate Logic from Groenendijk and Stokhof (1991) with the distinctive feature of treating the scoping of variables by means of sequences. Each variable is assigned its own sequence by a sequence assignment. This allows the introduction of an *Exit* operator for terminating otherwise persistent dynamic scopes.

We will illustrate sequence assignments at work. The sequence value of an assignment to a variable is depicted directly above the variable as boxes on top of each other forming a stack. The uppermost box is the frontmost element of the sequence. Only sequences for variable of interest are represented. For example, (26) illustrates an assignment g with $g(x) = [c, a]$ and $g(y) = [b]$.

(26)

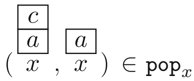


The DPLE system builds on a primitive relation pop on sequence assignments:

- $(g, h) \in \texttt{pop}_x$ iff h is just like g, except that $g(x) = [(g(x))_0] @ h(x)$.

There are actually two ways to interpret $(g, h) \in \texttt{pop}_x$. Read from g to h the relation describes the *popping* of the frontmost sequence element that is assigned to x by g. Read from h to g the relation describes the *pushing* of a new sequence value onto the front of the sequence that is assigned to x by h. An example is illustrated in (27).

(27)



Formulas are interpreted as relations on sequence assignments that change the before-state of a sequence assignment to an after-state in accordance with definition 3.

Definition 3 (*DPLE satisfaction and truth*).

Suppose a first-order model M with domain of individuals D. Suppose SA is the set of assignments from variables to sequences of individuals from D. Each formula ϕ is assigned to a relation $[\![\phi]\!]_M \subseteq SA \times SA$ as follows:

- $(g, h) \in [\![\phi \wedge \psi]\!]_M$ iff $\exists j : (g, j) \in [\![\phi]\!]_M$ and $(j, h) \in [\![\psi]\!]_M$
- $(g, h) \in [\![\exists x]\!]_M$ iff $(h, g) \in \texttt{pop}_x$
- $(g, h) \in [\![Exit_x]\!]_M$ iff $(g, h) \in \texttt{pop}_x$
- $(g, h) \in [\![x = y]\!]_M$ iff $g = h$ and $(g(x))_0 = (g(y))_0$
- $(g, h) \in [\![\mathrm{P}(x_1, \ldots, x_n)]\!]_M$ iff $g = h$ and $((g(x_1))_0, \ldots, (g(x_n))_0) \in M(\mathrm{P})$
- $(g, h) \in [\![\neg\phi]\!]_M$ iff $g = h$ and $\neg\exists j : (g, j) \in [\![\phi]\!]_M$

A formula ϕ is *true* with respect to g in M iff there is a h such that $(g, h) \in [\![\phi]\!]_M$.

From definition 3, we see how the after-state of a formula is the before-state of the next conjoined formula. Existential quantification is captured as the pushing of a random value onto the front of the x-sequence of the input assignment, while the *Exit* operator is realised as the popping of the frontmost value of the x-sequence. Predicate formulas correspond to tests on the frontmost values of variable sequences. The implementation of negation follows Dynamic Predicate Logic in possibly changing the assignment only internally to the scope of negation to act overall as a static test.

The DPLE system can be employed to provide a rendering of (8) as in (28).

(28) $\exists x \wedge \mathrm{enters}(x) \wedge \mathrm{smiles}(x) \wedge \exists y \wedge \mathrm{laughs}(y) \wedge \mathrm{likes}(x, y)$

Reading (3a) of (2) obtains with the formula (29a) with bracketing that conforms to (4a), while reading (3b) is captured by (29b) with bracketing that conforms to (4b).

(29) a. $\exists x \wedge \mathrm{collector}(x) \wedge \neg\mathrm{buy}(x) \wedge \mathrm{influenced}(x)$
b. $\exists x \wedge \mathrm{collector}(x) \wedge \neg(\mathrm{buy}(x) \wedge \mathrm{influenced}(x))$

Reflexive binding in (5) is captured with (30), while covaluation readings for (6) and (7) are obtained with (31) and (32), respectively.

(30) $\exists x \wedge \mathrm{John}(x) \wedge \mathrm{voted\_for}(x, x)$

(31) $\exists x \wedge \mathrm{John}(x) \wedge \mathrm{voted\_for}(x, y)$

(32) $\neg(\exists x \wedge \mathrm{candidate}(x) \wedge \neg(\mathrm{surprised}(x) \wedge \neg(\exists y \wedge \mathrm{voted\_for}(y, x) \wedge \neg y = x)))$

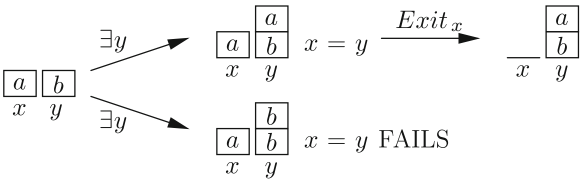
However the encodings of (28)–(32) are essentially what we might have expected to give had we been employing the system of Dynamic Predicate Logic, because there is no use made of the one feature that is unique to DPLE, namely the *Exit* operator.

As a more interesting application of DPLE consider (33). This acts to move the frontmost element of the x-sequence to the front of the y-sequence, where $x \neq y$.

(33) $\mathtt{move}_{x,y} \equiv \exists y \wedge x = y \wedge Exit_x$

We can give (34) as an illustration of how (33) works.

(34)



From (34) we see how $\exists y$ changes an input assignment to a new assignment where what was the original content of the y-sequence is stored ‘under’ a new frontmost element. For the next instruction to succeed the new frontmost element of the y-sequence must be a, else $x = y$ fails. Consequently, $\exists y \wedge x = y$ has a unique output. Then $Exit_x$ applies to remove the frontmost element of the x-sequence to give the final output.

To illustrate an application for `move` of (33) let us define operations $\bullet$ and $\Box$, where utilisation of the specific variables sbj, obj and p is rigidly fixed by the definitions.

(35) $\bullet \equiv \mathtt{move}_{sbj,p}$
$\Box \equiv \mathtt{move}_{obj,p} \wedge \bullet$

Let us also employ the convention in (36) to bring about formula iteration.

(36) $R^0 \equiv \top$
$R^n \equiv R^{n-1} \wedge R$

We will now aim to mimic the contribution of some English words and phrases with DPLE expressions. In (37) we define the contribution of several subject noun phrases that create *sbj* bindings, noting that *sbj* variable names are rigidly specified as parts of the encodings.

(37) `Someone` $\equiv \exists sbj$
`A_collector` $\equiv \exists sbj \wedge \mathrm{collector}(sbj)$
`John` $\equiv \exists sbj \wedge \mathrm{John}(sbj)$
`every_candidate` $\phi \equiv \neg(\exists sbj \wedge \mathrm{candidate}(sbj) \wedge \neg\phi)$

In (38) we aim to capture the contribution of both subject and object pronouns that have an open *i* parameter to take an index value, as well as a subject pronoun that is accompanied by *only* and also a reflexive pronoun. The reflexive pronoun has no index but is rather rigidly fixed to create an *obj* binding that is linked to the open *sbj* binding.

(38) $\texttt{she}_i, \texttt{he}_i \equiv \exists sbj \wedge (\texttt{move}_{p,r})^i \wedge sbj = p \wedge (\texttt{move}_{r,p})^i$
$\texttt{her}_i, \texttt{him}_i \equiv \exists obj \wedge obj = p \wedge (\texttt{move}_{p,r})^i \wedge (\texttt{move}_{r,p})^i$
$\texttt{only\_he}_i\ \phi \equiv \neg(\exists sbj \wedge \phi \wedge \neg((\texttt{move}_{p,r})^i \wedge sbj = p))$
$\texttt{herself}, \texttt{himself} \equiv \exists obj \wedge obj = sbj$

Finally we provide DPLE encodings for verbs in (39), again noting that the variable instances are rigidly specified.

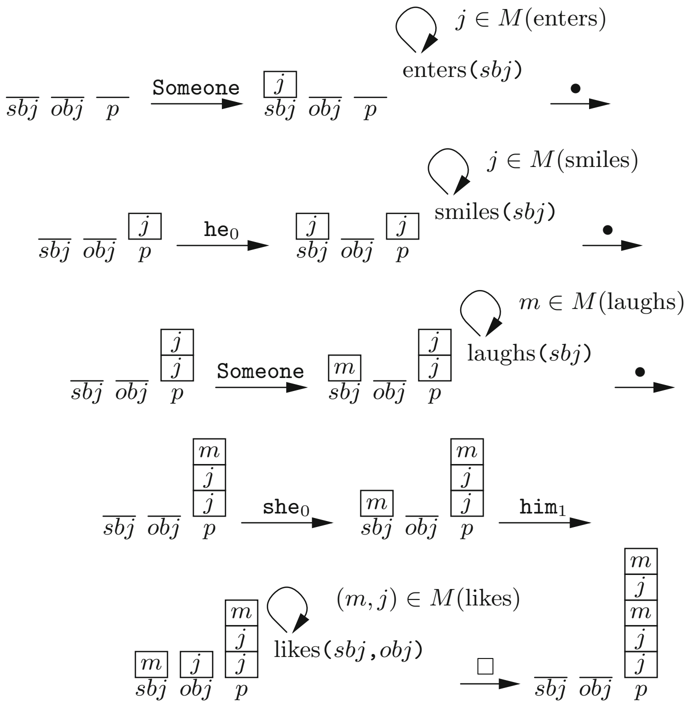
(39) $\texttt{enters} \equiv \text{enters}(sbj)$, $\texttt{smiles} \equiv \text{smiles}(sbj)$, $\texttt{laughs} \equiv \text{laughs}(sbj)$, $\texttt{likes} \equiv \text{likes}(sbj, obj)$, $\texttt{buy} \equiv \text{buy}(sbj)$, $\texttt{was\_influenced} \equiv \text{influenced}(sbj)$, $\texttt{voted\_for} \equiv \text{buy}(sbj)$, $\texttt{is\_surprised} \equiv \text{surprised}(sbj)$

We are now in a position to consider (40) as a rendering of (8).

(40) $\texttt{Someone} \wedge \texttt{enters} \wedge \bullet \wedge \texttt{he}_0 \wedge \texttt{smiles} \wedge \bullet \wedge \texttt{Someone} \wedge \texttt{laughs} \wedge \bullet \wedge \texttt{she}_0 \wedge \texttt{him}_1 \wedge \texttt{likes} \wedge \square$

In (41) an illustration is given of the changes that arise to a sequence assignment when (40) is evaluated, resulting in terminals that end exactly as with (11) and (20).

(41)



We can also capture reading (3a) of (2) with (42a), while reading (3b) follows from (42b).

(42) (a.) $\texttt{A\_collector} \wedge \neg\texttt{buy} \wedge \bullet \wedge \texttt{she}_0 \wedge \texttt{was\_influenced} \wedge \bullet$

(b.) $\texttt{A\_collector} \wedge \neg(\texttt{buy} \wedge \bullet \wedge \texttt{she}_0 \wedge \texttt{was\_influenced}) \wedge \bullet$

Reflexive binding in (5) is captured with (43), while covaluation readings for (6) and (7) are obtained with (44) and (45), respectively.

(43) $\texttt{John} \wedge \texttt{himself} \wedge \texttt{voted\_for} \wedge \Box$

(44) $\texttt{John} \wedge \texttt{him}_0 \wedge \texttt{voted\_for} \wedge \Box$

(45) $\texttt{every\_candidate}\,(\texttt{is\_surprised} \wedge \bullet \wedge \texttt{only\_he}_0\,(\texttt{him}_0 \wedge \texttt{voted\_for}))$

Examples (40) and (42)–(45) illustrate how it is possible to make bindings with DPLE uniform and principled with the dependency management that the *Exit* operator makes possible. However this comes at the cost of having to stipulate management instructions. Worse still, relevant management instructions are dependent on what has gone on before, with there being different ways to terminate a sentence, notably, □, • or no operation, depending on whether the sentence contains an object, only a subject, or falls under the scope of negation. To sum up, DPLE facilitates achieving uniform bindings. What we now require is a way to automate the management of open dependencies. This we aim to achieve with the system of the next section.

5 Pronominal and Variable Binding Again

In this section we illustrate a system with options for pronominal and variable name binding in subordinating contexts, and the option of pronominal binding for coordinating contexts.

One way to achieve such a system would be to place the machinery of predicate logic alongside Pointer Semantics of Sect. 3. While feasible, the resulting system would have two binding options—pronominal and variable binding—simultaneously available for subordinating dependencies. In a natural language like English the distinct binding options of reflexive binding and pronominal binding have a complementary distribution.

What we aim for instead in this section is to present a system that has quantification to open a variable binding that will serve as the mechanism for establishing grammatically determined dependencies such as linking subject and object arguments to the main predicate as well as reflexive binding, and for this to be subsequently handed over to a binding that is available to the pronominal machinery, either as part of a coordinating dependency or as part of a sufficiently embedded subordinating dependency. This will be accomplished with some of the dynamic control DPLE allowed over sequence assignments.

First we introduce the operations of `cons` and `snoc`:

- `cons` adds an element to the front of a sequence: `cons` $y\ \hat{x} = [y]$ `@` $\hat{x}$.
- `snoc` adds an element to the end of a sequence: `snoc` $y\ \hat{x} = \hat{x}$ `@` $[y]$.

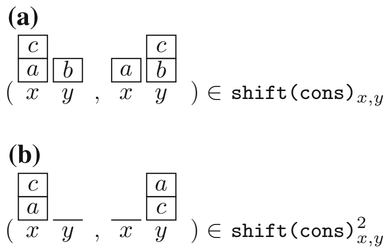
We now define `shift(`*op*`)` on pairs of assignments (g, h) to move from g to h or vice versa. For `shift(`*op*`)` the operation *op* needs to be specified, with suitable candidates being `cons` and `snoc` to give `shift(cons)` and `shift(snoc)`.

- $(g, h) \in \mathtt{shift}(op)_{x,y}$ iff $\exists k : (h, k) \in \mathtt{pop}_y$ and k is just like g, except that $g(x) = op((h(y))_0, k(x))$.

We will also use an iteration convention for relations on sequence assignments, e.g., implemented as in (36).

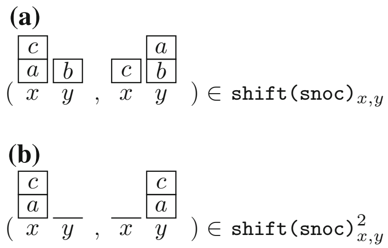
Reading from g to h, $(g, h) \in \mathtt{shift(cons)}_{x,y}$ moves the frontmost scope of the x sequence to the front of the y sequence. Examples are given in (46): (46a) illustrates a single action of `shift(cons)`; (46b) illustrates how iterated `shift(cons)` can move a sequence of scopes to the front of another sequence with the original ordering of scopes reversed.

(46)



Reading from g to h, $(g, h) \in \mathtt{shift(snoc)}_{x,y}$ moves the endmost scope of the x sequence to the front of the y sequence. Examples are given in (47): (47a) illustrates a single action of `shift(snoc)`; (47b) illustrates how iterated `shift(snoc)` can move a sequence of scopes to the front of another sequence with the original ordering of scopes intact.

(47)



Having iterable relations `shift(cons)` and `shift(snoc)` in addition to `pop` from Sect. 4, we introduce a minimal version of Scope Control Theory (SCT) from Butler (2010).

Definition 4 (*Minimal Scope Control Theory satisfaction and truth*).

Formulas are evaluated with respect to a first-order model $M = (D, I)$ and sequence assignment $g \in SA$, where SA is the set of assignments from variables to sequences of individuals from D. First, we define term evaluation:

- $g(v_n) = d$ if $\exists h : (h, g) \in \mathtt{pop}_v^i$ and $d = \uparrow(h(v))$, else undefined.

Next, we define formula evaluation:

- $M, g \models \exists x\phi$ iff $\exists h : (g, h) \in \mathtt{shift(cons)}_{e,x}$ and $M, h \models \phi$
- $M, g \models \rceil x\phi$ iff $\exists h : (g, h) \in \mathtt{shift(cons)}_{x,p}$ and $M, h \models \phi$
- $M, g \models \phi \wedge \psi$ iff $\exists h : ((g, h) \in \mathtt{pop}_e^{n(\psi)}$ and $M, h \models \phi)$ and $\exists h((g, h) \in \mathtt{shift(snoc)}_{e,p}^{n(\phi)}$ and $M, h \models \psi)$
- $M, g \models \neg\phi$ iff $\neg\exists h : (h, g) \in \mathtt{pop}_e^{n(\phi)}$ and $M, h \models \phi$
- $M, g \models P(t_1, \ldots, t_n)$ iff $(g(t_1), \ldots, g(t_n)) \in I(P)$

where $n(\phi)$ is a count of existentials in ϕ that are outside the scope of negation: $n(\exists x\phi) = n(\phi) + 1$, $n(\rceil x\phi) = n(\phi)$, $n(\phi \wedge \psi) = n(\phi) + n(\psi)$, $n(\neg\phi) = 0$, $n(\mathrm{P}(t_1, \ldots, t_n)) = 0$.
A formula ϕ is *true* with respect to g in M iff there is a $h \in SA$ such that $(h, g) \in \mathtt{pop}_e^{n(\psi)}$ and $M, h \models \phi$.
A term has form v_n in a predicate formula $\mathrm{P}(\ldots v_n \ldots)$, and denotes the $n + 1$ element of the v-sequence. In practice, indices different from 0 appear rarely. Nevertheless, they can appear. We will assume the convention that the index 0 can be omitted to keep the language to a conservative extension of traditional predicate logic notation.

The existential quantifier is defined to trigger an instance of `shift(cons)` that relocates the frontmost sequence element of the e-sequence to the x-sequence, where e is a privileged name and x is determined by the quantification instance. Similarly $\rceil x$ serves to relocate with `shift(cons)` an assigned sequence value, only relocating the frontmost value of the x-sequence, where x is determined by the given instance of $\rceil x$, to become the frontmost sequence value that is assigned to the privileged p variable.

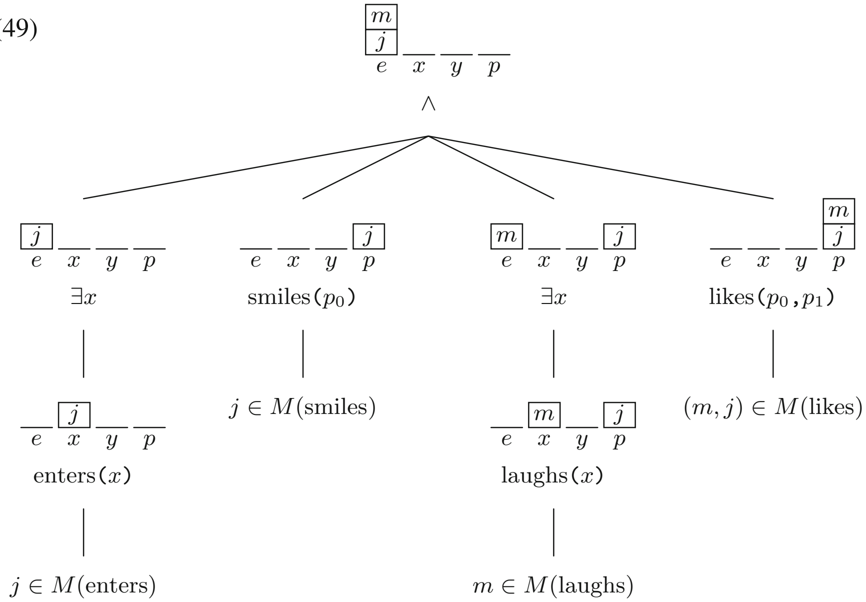
Conjunction also acts to modify the content of what is assigned to the privileged variables of e and p. Before evaluation of the first conjunct can proceed, the e-sequence has $n(\psi)$ values popped. Before evaluation of the second conjunct can proceed, there is the relocation with `shift(snoc)` of $n(\phi)$ values from the end of the sequence assigned to e to the front of the sequence that is assigned to p. Evaluation of $\neg\phi$ involves showing there is no way to extend the sequence assigned to the privileged e variable by $n(\phi)$ sequence values and have ϕ hold true.

Having now introduced Minimal SCT we can demonstrate the system with a rendering of (8) as in (48), which is syntactically identical to the PLA rendering of (10).

(48) $\exists x\mathrm{enters}(x) \wedge \mathrm{smiles}(p_0) \wedge \exists x\mathrm{laughs}(x) \wedge \mathrm{likes}(p_0, p_1)$

An evaluation of (48) is illustrated in (49), with the ordering of the three instances of $\wedge$ left underspecified since conjunction is once again associative, and happening against a sequence assignment with an initial state in which e is assigned the sequence $[m, j]$ while other variables are assigned the empty sequence. As the evaluation proceeds this assignment is modified, so that the terminals end exactly as with (11), (20) and (41).

(49)



Reading (3a) of (2) is captured with the formula (50a) with bracketing that conforms to (4a), while reading (3b) is captured by (50b) with bracketing that conforms to (4b). It is with the rendering of (50b) that we see the first deviation from what was possible with PLA. Notably (50b) employs $\rceil x$ with the consequence that the pronoun can be rendered as p_0 and so uniformity is achieved with the treatment of the pronoun in (50a).

(50) a. $\exists x(\text{collector}(x) \wedge \neg\text{buy}(x)) \wedge \text{influenced}(p_0)$
b. $\exists x(\text{collector}(x) \wedge \neg(\text{buy}(x) \wedge \rceil x \text{influenced}(p_0)))$

Reflexive binding in (5) is captured with (51), while covaluation readings for (6) and (7) are obtained with (52) and (53), respectively. Again this illustrates the consistent use of variables named p (akin to PLA-like pronouns) to capture English pronouns, while other variables capture clause internal linking and reflexive binding.

(51) $\exists x(\text{John}(x) \wedge \text{voted\_for}(x, x))$

(52) $\exists x(\text{John}(x) \wedge \text{voted\_for}(x, p_0))$

(53) $\neg\exists x(\text{candidate}(x) \wedge \neg(\text{surprised}(x) \wedge \rceil x \neg\exists y(\text{voted\_for}(y, p_0) \wedge \neg y = p_0)))$

With the management of dependencies that occurs with conjunction and the $\rceil$-operator it is possible to establish a more direct correspondence with a natural language like English. For example, following what was possible in (37) with DPLE, we define the contribution of several subject noun phrases in (54) with rigidly prescribed bindings.

(54) $\texttt{Someone}\ \phi \equiv \exists sbj\phi$
$\texttt{A\_collector}\ \phi \equiv \exists sbj(\text{collector}(sbj) \wedge \phi)$
$\texttt{John}\ \phi \equiv \exists sbj(\text{John}(sbj) \wedge \phi)$
$\texttt{every\_candidate}\ \phi \equiv \neg\exists sbj(\text{candidate}(sbj) \wedge \neg\phi)$

In (55) we capture subject and object pronouns, as well as a subject pronoun accompanied by *only* and also a reflexive pronoun.

(55) $\texttt{she}_i, \texttt{he}_i\ \phi \equiv \exists sbj(sbj = p_i \wedge \phi)$
$\texttt{her}ensuremath_i, \texttt{him}_i\ \phi \equiv \exists obj(obj = p_i \wedge \phi)$
$\texttt{only\_he}_i\ \phi \equiv \neg\exists sbj(\phi \wedge \neg sbj = p_i)$
$\texttt{herself}, \texttt{himself}\ \phi \equiv \exists obj(obj = sbj \wedge \phi)$

Having (54)–(55), and adopting the codings of verbs in (39) without change, we are in a position to render (8) as (56), The notable advantage over the DPLE version of (40) is that there is no longer a need for the explicit management instructions of • and □.

(56) Someone enters $\wedge$ $\texttt{he}_0$ smiles $\wedge$ Someone laughs $\wedge$ $\texttt{she}_0$ ($\texttt{him}_1$ likes)

Reading (3a) of (2) is captured with (57a), while reading (3b) follows from (57b).

(57) a. A_collector $\neg$buy $\wedge$ $\texttt{she}_0$ was_influenced
b. A_collector $\neg$(buy$\wedge\rceil sbj$($\texttt{she}_0$ was_influenced))

Reflexive binding in (5) is captured with (58), while covaluation readings for (6) and (7) are obtained with (59) and (60), respectively.

(58) John (himself voted_for)

(59) John ($\texttt{him}_0$ voted_for)

(60) every_candidate (is_surprised$\wedge\rceil sbj$(only_$\texttt{he}_0$ ($\texttt{him}_0$ voted_for)))

An apparent weakness that remains with the encodings of (57b) and (60) is that the management of $\rceil sbj$ must be stipulated. What we require is a trigger to motivate the hand-over from a non-*p*(ronominal) binding to a *p* binding. We might suppose such hand overs come about whenever the non-*p* binding is set to be reused. So in the examples of (57b) and (60) the fact that a new clause is entered which itself contains a subject argument opening a sbj binding is reason enough for the occurrence of $\rceil sbj$.

Extending this rational we should expect that when there is no subject in an embedded clause there will be no corresponding $\rceil sbj$, e.g., with the consequence that a sbj dependency can be maintained across a clause boundary, as holds for the control dependency in (61).

(61) John needed to go.

Interestingly a control predicate may take as complement an embedded clause that contains a main predicate that is unable to receive a subject binding, but only provided there is the presence of expletive *it*, as the contrast between (62a) and (62b) demonstrates.

(62) a. John needed it to rain.
b. *John needed to rain.

Supposing expletive *it* to contribute $\rceil sbj$ removes the subject binding inherited from the control verb before *rain* is encountered.

Finally we should note that for the current approach to scale, e.g., to handle passivisation, a more sophisticated interaction with syntax is required to implement how binding names are obtained, which is left for future work.

6 Summary

This chapter began with the observation that in English properties for coordinating dependencies also held for sufficiently deeply embedded subordinating dependencies. This was taken to suggest that the two dependency types arise with the same mechanism, and the purpose of this chapter has been to explore formal ways to implement such a mechanism. This chapter has ended with a system that preserves classical variable name binding, which captures well the more restricted role of clause internal argument linking and reflexive binding in natural language, and have a subsequent hand-over to a pronominal machinery in subordinating contexts with sufficient embedding, which captures well the ad-hoc nature of pronominal binding. The set up of the system was such that the same pronominal machinery carried over as the means to establish dependencies across coordinating contexts. With its management of dependencies the system was demonstrated to reduce the distance between natural language form and its realisation as a formula for interpretation.

Acknowledgments This research has been supported by the Japan Science and Technology Agency PRESTO program (Synthesis of Knowledge for Information Oriented Society). I would like to thank the three anonymous reviewers for their extremely helpful comments that prompted considerable improvements to the chapter.

References

Büring, D. (2005). Bound to bind. *Linguistic Inquiry*, *36*(2), 259–274.
Butler, A. (2010). *The semantics of grammatical dependencies, vol. 23 of current research in the semantics/pragmatics interface*. Bingley: Emerald.
Cresswell, M. J. (2002). Static semantics for dynamic discourse. *Linguistics and Philosophy*, *25*, 545–571.
Dekker, Paul. (2002). Meaning and use of indefinite expressions. *Journal of Logic, Language and Information*, *11*, 141–194.

Groenendijk, Jeroen, & Stokhof, Martin. (1991). Dynamic predicate logic. *Linguistics and Philosophy*, *14*(1), 39–100.

Heim, I. (1993). Anaphora and semantic interpretation: A reinterpretation of Reinhart's approach. Technical report SfS-Report-07-93, University of Tübingen.

Hollenberg, M., & Vermeulen, C. F. M. (1996). Counting variables in a dynamic setting. *Journal of Logic and Computation*, *6*, 725–744.

Kamp, H., & Reyle, U. (1993). *From discourse to logic: Introduction to model-theoretic semantics of natural language, formal logic and discourse representation theory*. Dordrecht: Kluwer.

Reinhart, T. (2000). Strategies of anaphora resolution. In H. Bennis, M. Everaert, & E. Reuland (Eds.), *Interface strategies* (pp. 295–325). Amsterdam: Royal Academy of Arts and Sciences.

van Eijck, J. (2001). Incremental dynamics. *Journal of Logic, Language and Information*, *10*, 319–351.

Vermeulen, C. F. M. (1993). Sequence semantics for dynamic predicate logic. *Journal of Logic, Language and Information*, *2*, 217–254.

Vermeulen, C. F. M. (2000). Variables as stacks: A case study in dynamic model theory. *Journal of Logic, Language and Information*, *9*, 143–167.

What is a Universal? On the Explanatory Potential of Evolutionary Game Theory in Linguistics

Gerhard Jäger

Abstract Natural languages are shaped by evolutionary processes, both in the sense of biological evolution of our species, and, on a much shorter time scale, by a form of cultural evolution. There are long research traditions in theoretical biology and economics (a) to model communication by means of game theory, and (b) to use game theory to study biological and cultural evolution. Drawing mostly on work by Huttegger (2007) and Pawlowitsch (2008), this chapter argues that results and methods from game theory are apt to formalize the intuitive notion of linguistic universals as emergent properties of communication.

Keywords Evolutionary game theory · Linguistic universals · Language evolution · Signaling games

1 Introduction: Language and Evolutionary Game Theory

One of the central issues in modern linguistics is the search for **universals**, that is properties that are shared by all languages. Empirical research over the past few decades has unearthed a solid amount of universals or quasi-universals (properties that are shared by almost all natural languages), some of them quite contingent.

There is an ongoing intense and at times ideological discussion within the linguistic community what a satisfactory explanation for universals should look like. One school of thought (most prominently defended by Noam Chomsky; see for instance Chomsky 1957, 1995) takes it that universals are reflexes of our innate linguistic competence that is shared basically by all humans and ultimately genetically determined. We will call this the *nativist* position. The other creed (see for instance Bybee 2001; Barlow and Kemmer 2000 or Haspelmath 1999) holds the view that languages

G. Jäger (✉)
Institute of Linguistics, University of Tübingen, Wilhelmstr. 19, 72072 Tübingen, Germany
e-mail: gerhard.jaeger@uni-tuebingen.de

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*, Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_5,

are adapted to its functions in communication and cognition. Therefore linguistic universals should be explained in terms of language function. This school of thought is usually dubbed *functionalist*.

Both approaches face severe epistemological problems. It is not much of an explanation to claim that a certain universal is based on innate, genetically determined, properties of the human brain. Without a independent justification for that assumed innate feature of the brain, the innateness hypothesis amounts to little more than a restatement of the facts. Functionalism, by itself, is no more explanatory either. To show that a certain property of human languages facilitates their usage has to be complemented by a plausible reconstruction of a causality that leads to this kind of adaptation. Additionally, functionalist explanations are frequently *post hoc*, i.e. functionalist research frequently starts out with some empirically established universal and tries to find a function that the universal is an adaptation for, rather than truly predicting universals from function.

However, both approaches seem to be fundamentally true, if properly conceived. Nobody would deny that there are innate constraints on what kind of linguistic items an infant is able to acquire and to put to use. To take a simple example: Many languages employ the *relative pitch* of speech sounds to encode grammatical distinctions. For instance, in English a rising intonation at the end of a sentence can be used to mark the sentence as a question. In other languages, as for instance in Chinese, the same sequence of sounds can acquire a completely different meaning, depending on whether it is pronounced with a rising or a falling intonation. There are no languages, however, that would employ *absolute pitch* to mark linguistic distinctions. A low pitch of a female speaker might still be higher, in terms of absolute frequency, than the high pitch of a male speaker. What matters for natural languages though is the position of a certain pitch within the range of pitches that a particular speaker is able to produce. For all we know, the reliance on relative pitch is a good candidate for a linguistic universal that is based on innate properties of the human brain.

On the other hand, some linguistic universals cannot be reduced to the cognitive state of a single language user. They are irreducibly social in nature. A well-known candidate is Zipf's law (Zipf 1935, 1949). It states that in a sufficiently large corpus of naturally occurring speech, the frequency of occurrence of the words of the language are distributed according to a power law. The most frequent word is used about twice as many times as the second most frequent, about ten times as often as the tenth most frequent word etc. More generally, if you list the words of a language according to their frequency of usage, the rank of a word in this list is approximately inversely proportional to its frequency of occurrence. This law has been validated many times for many languages. It is a good candidate for a universal property of language usage. However, this regularity is certainly not cognitively represented in the minds of language users. Rather, it is an emergent property of the usage of language in social interaction. The two approaches should thus be considered as complementary rather than as mutually exclusive. Evolutionary game theory, I would like to argue, provides a formal framework that model questions of language evolution that allows us to integrate the two approaches, innateness and the social function of language.

2 Modeling Language Evolution as a Game

Game theory is a formal language that allows us to study the *interaction of individual agents*, and the joint outcome of this interaction. The evolutionary branch of game theory explicitly takes into account the *path dependence* of this interaction under some limited form of rationality.

An evolutionary game consists of two parts. First, the so–called *stage game*, that is, the basic situation of interaction that is encountered repeatedly among a group of individuals, where it is defined what are possible action choices of individuals, so–called *strategies*, and how individuals' payoffs are calculated as a function of the strategy profile in the whole population. Second, the *game dynamics*, which described according to which rule individuals' strategies are updated from one period to the next. This rule can be either exogenously given, representing something like a "law of nature", that is, a feedback mechanism of the environment; or it can be derived from some optimizing behavior of individual agents. This is usually seen as the part of the model setup where different degrees of rationality enter the description of the problem. In an economic model it is often plausible to assume that agents are perfectly rational, in the sense that they can calculate their optimal choice of action given their state of knowledge and their beliefs about other agents actions. On the other hand, in most ecological or biological models it rarely makes sense to assume that agents consciously interact. Rather they perform some predefined program whenever they get the stimulus to do so. Taking these as the two extremes of one scale, I would argue that most linguistic applications lie somewhere in between, depending on the particular problem at hand. If we are interested in an aspect of the origins of language in an anthropological sense we are most probably closer to a biological set up, whereas if we are interested in some aspects of the pragmatics of language use, it definitely makes sense to assume that agents are conscious about their interaction, that they have beliefs about other players actions, that they employ some kind of optimizing behavior given all these considerations etc.

Any optimizing behavior, or even imitation, in the last event, implicitly assumes some innate abilities or properties that are not further explained in the model. Assumptions on innate abilities of individual agents, however, do not only enter the construction of the game dynamics, they also may be part of the description of possible actions choices of individual agents. For example, if a strategy of an agent represents a program to perform a particular utterance whenever he or she observes a particular event of nature, this definitely involves the assumption that this agent has an abstract notion of this event and that he or she has the ability to link this to an arbitrary sign.

What is specific about applications of evolutionary game theory to linguistic questions is that strategies in the stage game are very often not just verbal descriptions like "cooperate" or "do not cooperate", but take the form of a *mathematically complex object*, similar to a quantitative trait in biology. This is a direct consequence of the fact that linguistic interaction is sequential in nature, and thus most naturally modeled as an extensive game. Normalizing an extensive game leads to strategies that are functions of a particular kind. As such these strategies can display properties that we can describe in some formal language. For the kind of game I am going to discuss below, strategies can be described by stochastic matrices, using the language of linear algebra. Eventually, we aim at learning something about the *regularity patterns* of these strategies that are used in an equilibrium outcome of the model.

Thus, innateness assumptions typically enter the description of a game, whereas the functionalist point of view is reflected in the solution concept applied. This is not to say that innateness enters the model *only* at the level of assumptions. An evolutionary game can also serve the purpose of explaining some property of language that is considered to be innate—but what sustains this property are the equilibrium conditions of individuals interaction. It is in this sense I consider the explanatory value of a game theoretic model as *functionalist*—even though it may rely on assumptions that concern innate abilities of individual agents.

I will discuss this and related issues in more detail for a specific class of games that are widely used in game theoretic approaches to language.

3 Signaling Games

Signaling games in the style of Lewis (1969) or Nowak and Krakauer (1999) have received particular attention in the newly arising literature that uses evolutionary game theory to study questions about the evolution of language.

In these games there is a finite number of events that potentially become the object of communication and a finite number of arbitrary signs. In each round of the game, nature presents the sender with one of the events. The sender in turn emits a signal that is visible to the receiver. Finally, the receiver guesses an event, possibly using the signal received as a clue. If the guess of the receiver is correct, both players score a point, otherwise neither receives a payoff.

A strategy in the role of the sender is thus a map from the set of events to the set of signals, and a strategy in the role of the receiver is a map from received signals to events.

As long as domain and range of these functions are finite, it is convenient to use the language of linear algebra and to represent functions as matrices.

A strategy in the role of the sender can be represented by an $n \times m$ matrix P (n being the number of events, and m the number of signals), such that each row contains exactly one cell with the entry 1, while all other cells have the entry 0. The intended interpretation is that $p_{ij} = 1$ if event i is mapped to signal j. Likewise a strategy in the role of the receiver can be represented by an $m \times n$ matrix Q, where here the interpretation is that $q_{ji} = 1$ if signal j is associated event i and 0 otherwise.

For the sake of simplicity, in the context of this chapter I confine attention to cases where $n = m$, i.e. there are exactly as many signals as events. Also, I assume that all events occur with the same probability, and that sending or receiving signals does not incur any costs. With these assumptions, the payoff function is identical for both players, and it can be defined by

$$\pi(P, Q) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} q_{ji},$$

(The utility that was informally described in the text can be obtained if $\pi(P, Q)$ is divided by n, because each event occurs with probability $\frac{1}{n}$. However, multiplying all payoffs by a constant factor does not alter the structure of a game, and we can thus as well drop this factor.)

In the language of linear algebra, this can conveniently be written as

$$\pi(P, Q) = \mathrm{tr}(PQ),$$

where $\mathrm{tr}(PQ)$ denotes the trace of (PQ).

It is assumed that all individuals of a linguistic community find themselves in the roles of sender and receiver with equal probabilities. A strategy for the *symmetrized game*, then, is a pair of two matrices, (P, Q), and the payoff function is given by

$$F[(P, Q), (P', Q')] = \frac{1}{2}\mathrm{tr}(PQ') + \frac{1}{2}\mathrm{tr}(P'Q).$$

Note that this payoff function is symmetric,

$$F[(P, Q), (P', Q')] = F[(P', Q'), (P, Q)],$$

giving rise to a so–called *doubly symmetric* or *partnership game*, that is, a symmetric game with a symmetric payoff function.

3.1 Mixed Strategies and Population Games

Rephrasing the model at hand in a population based framework, every pair (P, Q) can be identified with a particular *type* of agent. A *state of the population*, then, is a vector of type frequencies,

$$x = (x_1, x_2, \ldots, x_L) \text{ s.t.} \sum_{l=1}^{L} x_l = 1,$$

where x_l is the fraction of agents using pure strategy l. Formally such a vector of type frequencies is equivalent to a *mixed strategy*.

The average payoff of a type is interpreted as its *fitness*,

$$f_l(x) = \sum_{l'=1}^{L} x_{l'} F\left[(P_l, Q_l), (P_{l'}, Q_{l'})\right];$$

the *average fitness in the population* is denoted by

$$\bar{f}(x) = \sum_{l=1}^{L} x_l f_l(x).$$

To every vector of type frequencies $x = (x_1, x_2, \ldots, x_L)$ we can assign the *population's average strategy profile* in terms of the P and Q matrices,

$$(\bar{P}(x), \bar{Q}(x)) = \left(\sum_{1}^{L} x_l P_l, \sum_{1}^{L} x_l Q_l\right),$$

which can be written as

$$(\bar{P}(x), \bar{Q}(x)) = \left[\begin{pmatrix} \bar{p}_{11} & \ldots & \bar{p}_{1j} & \ldots & \bar{p}_{1m} \\ \vdots & & \vdots & & \\ \bar{p}_{i1} & \ldots & \bar{p}_{ij} & \ldots & \bar{p}_{im} \\ \vdots & & \vdots & & \\ \bar{p}_{n1} & \ldots & \bar{p}_{nj} & \ldots & \bar{p}_{nm} \end{pmatrix}, \begin{pmatrix} \bar{q}_{11} & \ldots & \bar{q}_{1i} & \ldots & \bar{q}_{1n} \\ \vdots & & \vdots & & \\ \bar{q}_{j1} & \ldots & \bar{q}_{ji} & \ldots & \bar{q}_{jn} \\ \vdots & & \vdots & & \\ \bar{q}_{m1} & \ldots & \bar{q}_{mj} & \ldots & \bar{q}_{mn} \end{pmatrix}\right],$$

where $\bar{p}_{ij}$ is the sum of all type frequencies whose i, j–th entry in P is equal to 1, and $\bar{q}_{ji}$ is the sum of all type frequencies whose j, i–th entry in Q is equal to 1, that is,

$$\bar{p}_{ij} = \sum_{l:p^l_{ij}=1} x_l \quad \text{and} \quad \bar{q}_{ji} = \sum_{l:q^l_{ji}=1} x_l.$$

In general the average strategy profile is a pair of two stochastic matrices, which are denoted by $(\bar{P}, \bar{Q})$. The average payoff or *fitness* of a type $f_l(x)$ then can be written as the payoff of this strategy from play against the population's average strategy,

$$f_l(x) = F\left[(P_l, Q_l), \left(\bar{P}(x), \bar{Q}(x)\right)\right] = \frac{1}{2}\mathrm{tr}\left(P_l \bar{Q}(x)\right) + \frac{1}{2}\mathrm{tr}\left(\bar{P}(x) Q_l\right),$$

and the *average fitness in the population* is the payoff of the population's average strategy from play against itself,

$$\bar{f}(x) = F\left[\left(\bar{P}(x), \bar{Q}(x)\right), \left(\bar{P}(x), \bar{Q}(x)\right)\right] = \mathrm{tr}\left(\bar{P}(x)\bar{Q}(x)\right).$$

Hurford (1989) introduces essentially the same model, though he does not use the language of game theory. He uses this model to study the evolutionary emergence of a particular linguistic universal known as *bidirectionality*, that is, the property that whenever an individual links a particular sign to a particular concept (object of communication), then this individual will also link this concept to that sign.[1] In a nutshell, this means that each adult speaker of a language uses, in the role of the speaker, a code that she is able understand in the role of the hearer. This feature is so deeply entrenched in our conception of "language" that it seems to be almost too obvious to mention. However, there are plenty of signaling systems that do not have this property. An obvious example are intermediate stages in the acquisition of a natural language (both in first language acquisition by infants and in second language acquisition by adults). New linguistic items are much faster acquired passively than actively. This means that a language learner can *interpret* some linguistic items correctly without being able to *produce* them.

In some signaling games there are strategies that display perfect bidirectionality. For 2 events and 2 signals, for instance, these are

$$(P_1, Q_1) = \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right], \text{ and } (P_2, Q_1) = \left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right].$$

But there are also strategies that are not bidirectional at all; rather they display what one would call "perfectly inconsistent behavior" in the role of the sender and the receiver, namely

$$(P_1, Q_2) = \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right], \text{ and } (P_2, Q_1) = \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right].$$

Most linguistic theories tacitly assume that bidirectionality is an innate property, which is ultimately genetically determined. If we look at isolated sender–receiver interaction, bidirectionality is not required for successful communication. But obviously bidirectionality seems to be fixed in humans. Shifting the question to an evolutionary framework, Hurford asks whether bidirectionality could get fixed due to some evolutionary advantage in a population based setting. In order to test this hypothesis, Hurford ran a series of computer experiments, where he let "Saussurean strategists", that is, individuals who update their Q according to their received P in a best–response fashion according to the asymmetric game, compete against types with other behavioral rules. Individuals who communicate better leave relatively more offspring and parents transmit their type to their kids. This simulation is run

[1] Sometimes this is also referred to as the notion of the Saussurean sign.

for different initial conditions. There seems to be good evidence that for most initial conditions Saussurean strategists indeed do better than other behavioral types. However, this is not true for all initial conditions.

3.2 Equilibrium Selection

The kind of evolutionary dynamics that Hurford, Nowak and their coworkers assume in their simulations assigns fitness to communicating agents which is proportional to the average communicative success of that agent within the given population. This success rate, in turn, depends on the relative frequency of communicative strategies within this population. We are thus dealing with frequency dependent selection, which can be modeled by means of evolutionary game theory.

In the tradition of evolutionary game theory, a Nash equilibrium in mixed strategies is interpreted as an equilibrium composition of the population. It generally holds that in a Nash equilibrium in mixed strategies, all strategies that are played with some positive probability must yield the same payoff given the specific probability mix of all the other players. In the population interpretation of mixed strategies in mind, this translates into the condition that in a Nash–equilibrium state of the population every type l that is present with some positive frequency must yield the same average payoff as all the other resident types—given the actual composition of the population.

For a Nash–equilibrium strategy of a symmetrized game, the strategies choices in the two roles have to be best responses to each other, that is, P is a best response to Q and Q is a best response to P. The condition that Q is a best response to P is the criterion that Hurford (1989) uses to characterize bidirectionality. Adopting this concept of bidirectionality, we may say that in a Nash–equilibrium composition of the population, bidirectionality is satisfied on the level of the population's average sender and receiver matrices. Interestingly, this property does not extend from the population's average to the individuals' level.

Let us consider an example. Suppose $n = m = 2$, i.e. we have two signals and two event. Then the set of all sender matrices is given by

$$\mathcal{P}_{2\times 2} = \left\{ P_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ P_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\right.$$
$$\left. P_3 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix},\ P_4 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\},$$

and the set of all receiver matrices is given by

$$\mathcal{Q}_{2\times 2} = \left\{ Q_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ Q_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\right.$$
$$\left. Q_3 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix},\ Q_4 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\}.$$

Now suppose one half of the population uses pure strategy (P_1, Q_2), and the other half uses (P_2, Q_1). Then population's average strategy is

$$(\bar{P}, \bar{Q}) = \left[\begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}\right].$$

In this case $\bar{P}$ is a best response to $\bar{Q}$ and vice versa, but still every individual agent is perfectly inconsistent between his or her sender and receiver strategy.

As for most models used in language evolution, the signaling game considered has an abundance of equilibria. In the case of 2 events and 2 signals, there are already 6 symmetric Nash equilibria in pure strategies, (P_1, Q_1), (P_2, Q_2), (P_3, Q_3), (P_3, Q_4), (P_4, Q_3), (P_4, Q_4), and there are whole continua of equilibria in mixed strategies. For these equilibria the average strategy profile is of the form

$$\left[\begin{pmatrix} 1-\alpha & \alpha \\ 1-\alpha & \alpha \end{pmatrix}, \begin{pmatrix} 1-\beta & \beta \\ 1-\beta & \beta \end{pmatrix}\right],$$

where α and β are between 0 and 1—for example, if one half of the population uses pure strategy (P_1, Q_1), and the other half uses (P_2, Q_2). The population's average strategy profile then is

$$\left[\begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}\right].$$

$$\left[\begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}\right].$$

The problem of equilibrium selection is therefore of major importance.

An important part of the program in evolutionary game theory has focused on establishing links between the static analysis of Nash equilibria, and its refinement concepts, and the stability properties of game dynamics. A great deal of effort has been devoted to the so–called *replicator dynamics*.

In continuous time, the replicator dynamics is given by a system of L differential equations,

$$\frac{\dot{x}_l}{x_l} = f_l(x) - \bar{f}(x), \quad l = 1, \ldots, L,$$

where $\dot{x}_l$ denotes the derivative of x_l with respect to time. So the growth rate of type l is given by the difference of its average payoff minus the average payoff in the population.

This dynamics can be interpreted in the sense of both biological as well as cultural transmission of strategies from one generation to the next (roughly corresponding to the nativist and the functionalist perspective in linguistics that were mentioned in the beginning). Agents who communicate more successfully are more successful in finding mates, acquiring sufficient sources of food, escaping dangers, and so on, which yields them either a direct or an indirect advantage in reproduction, or increases the chances of their offspring to reach reproduction age. Or, in the context of cultural evolution, agents who communicate more successfully are more likely to be imitated by other agents and therefore the strategies that they use will reproduce with a higher rate.

A Nash–equilibrium strategy is necessarily a rest point of the replicator dynamics. The reverse is not generally true. For example, every monomorphic state, that is, a state where the whole population consists of only one type, trivially is a rest point of the replicator dynamics, but it is not necessarily a Nash–equilibrium strategy.

In language evolution we are not so much interested in any one particular equilibrium; rather we want to understand where a particular dynamics typically will lead us to and what this implies for the qualitative regularity patterns of the communicative strategies that are used by agents.

The most commonly applied refinement concept in an evolutionary context is that of an *evolutionarily stable strategy*. A strategy played in a symmetric Nash equilibrium is evolutionarily stable if either (i) it has no alternative best replies (that is, if it is a strict equilibrium), or if (ii) in case that there is an alternative best reply, this alternative best reply yields a strictly lower payoff against itself than the original Nash strategy yields against this alternative best reply. As a lighter version of this, if in case (ii) the alternative best reply yields only a lower or equal payoff against the original Nash strategy, then the original Nash strategy is called a *neutrally stable* or *weakly evolutionarily stable strategy*.

Though their name seems to hint at some dynamic story, both evolutionary and neutral stability are as such static refinement criteria for symmetric Nash equilibria. However, as it has been shown by Taylor and Jonker (1978) for the continuous case, and by Hofbauer et al. (1979) for the discrete case, every *evolutionarily stable* strategy is a *locally asymptotically stable* rest point of the replicator dynamics. This means that a system that has reached such a state will return there if it is disturbed by a small perturbation. In analogy to this, Thomas (1985), and in a more general context Bomze and Weibull (1995), show that every *neutrally stable* strategy is *Lyapunov stable* in the replicator dynamics. If a system is in a Lyapunov stable state, it will remain within the local environment of this state if a small perturbation occurs.

Unfortunately, none of the converses of these results is true in general. This means that whenever we have found an evolutionarily stable strategy, we know that once a population has attained such a strategy, it is locally asymptotically stable, but this does not rule out there to be other asymptotically stable rest points that do not correspond to an evolutionarily stable strategy; analogously for neutral stability and Lyapunov stability. However, in the case of signaling games, help comes from their symmetry properties.

Akin and Hofbauer (1982) show that for doubly symmetric games each orbit of the replicator dynamics, indeed, *converges to some rest point*. This is related to the property that for such games, the replicator dynamics induces a strictly monotonic increase in the average payoff along every non–stationary solution path, which in biology is known as *Fisher's fundamental theorem of natural selection*. Hofbauer and Sigmund (1988) show that in this case a locally asymptotically stable rest point is evolutionarily stable, so that for doubly symmetric games, evolutionary stability indeed *coincides* with asymptotic stability. Bomze (2002) shows that an analogous result holds true between neutral stability and Lyapunov stability so that for doubly symmetric games with pairwise interaction a rest point of the replicator dynamics is Lyapunov stable if and only if it corresponds to a neutrally stable Nash equilibrium. So we can rule out that there are other rest points that are Lyapunov stable, and the dynamics typically will lead to a Nash equilibrium that satisfies neutral stability. Once we are able to identify general patterns in the strategies used in population states that satisfy the essentially static criterion of neutral stability, this will tell us something about the regularity patterns of the communicative strategies that can be expected to arise in the long run.

For the signaling game considered here evolutionarily stable strategies are all pairs of permutation matrices (P, Q) such that one matrix is the transpose of the other (Wärneryd 1993; Trapa and Nowak 2000). For 2 events and 2 signals, these are the two pairs

$$(P_1, Q_1) = \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right], \text{ and } (P_2, Q_2) = \left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right].$$

For 2 events and 2 signals evolutionarily stable strategies are also the only neutrally stable strategies. However, if the number of events or signals is greater or equal to 3, this is no longer true.

For example, if $n = m = 3$ a possible neutrally stable strategy looks like

$$(P, Q) = \left[\begin{pmatrix} 1-\alpha & \alpha & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1-\beta & \beta \end{pmatrix}\right], \ \alpha, \beta \in (0, 1).$$

There is no strategy $(P', Q') \in \mathcal{P}^{\Delta} \times \mathcal{Q}^{\Delta}$ such that $F[(P, Q), (P', Q')] = F[(P, Q), (P, Q)]$ and such that $F[(P, Q), (P', Q')] > F[(P', Q'), (P', Q')]$. The same is true if in the example above *either* α *or* β is 0 or 1, which means that one of the two matrices contains a column that consists entirely of zeros.

This example shows that signaling games may possess equilibrium states that are neutrally but not evolutionarily stable. In fact, Huttegger (2007) and Pawlowitsch (2008) show that games with more than two events and signals have infinitely many such states. These are generally states that involve a certain amount of synonymy and homonomy on the population level; they represent sub-optimal communication strategies.

The mentioned authors furthermore show that this set of neutrally but not evolutionarily stable states always has a basin of attraction that has a Lebesgue measure larger than zero. So if an initial state is chosen at random and each initial state has a positive probability density, there is a positive probability that the population will converge to such a suboptimal state. I performed a numerical approximation which revealed that the joint basin of attraction of the set of neutrally but not evolutionarily stable strategies in a 3×3 game comprises about 1.2 % of the state space. (See the Appendix for details.)

4 Evolution of Signaling Games and Linguistic Universals

Let us return to the main methodological point of this chapter, the potential of evolutionary game theory for illuminating the concept of a linguistic universal. Intuitively, a universal of a certain game under a certain dynamics is a set of population states U such that with probability one, a population will eventually enter U and never leave it afterwards. To make this precise, one has to assume a probability distribution over initial states. Since I assume that the stage game by definition excludes all "impossible" states, I take it that each strategy, pure or mixed, has a positive probability density. Since under the replicator dynamics, each trajectory in a doubly symmetric game converges to some rest point (as shown in Akin and Hofbauer 1982), a set U is a universal if and only if *almost all initial states converge to some point* $x^* \in U$.

Given the considerations in the previous section, we can conclude that any set containing the set of Nash equilibria is a universal. This follows from the facts that (a) all orbits converge to a rest point, and (b) if an interior point converges to a single point, this point is a Nash equilibrium (as shown for instance in Hofbauer and Sigmund 1998:69/70). Since the boundary of a simplex is a null set, almost all points converge to some Nash equilibrium.

The significance of Huttegger's and Pawlowitsch's result is that each universal has to include all neutrally stable states, including those that are not evolutionarily stable. In the model chosen, perfect bidirectionality (in the sense of a one-one map between forms and meanings) is thus not a universal.

Is it possible to give a more precise characterization of the set of universals of this game? Huttegger (2007) proves that almost all points of the state space converge to a Nash equilibrium at the boundary of the state space. Translated into our terminology, this means that it is a universal that not all possible grammars are represented in the population.

It is important to stress that these universals are social in nature, rather than cognitive. Being in a Nash equilibrium state at the boundary of the state space is a property of a population, not of an individual member of such a population.

To draw conclusions about universal properties of the state of individual agents, it would be necessary to narrow down the class of population level universals further. This problem proves to be surprisingly difficult though, and I have to close this section with some conjectures and suggestions for further research.

It might perhaps seem plausible to assume that almost all orbits converge to a Lyapunov stable—and thus neutrally stable—point. This would amount to the claim that any set containing all neutrally stable states is a universal. However, it is possible to come up with games where this is not the case. In the Appendix I discuss a doubly symmetric game which has a non-neutrally stable Nash equilibrium that attracts a set with a positive measure.

Of course, this game is not a signaling game. It is thus possible—and in fact, it seems highly likely—that in signaling games, almost all points converge to a neutrally stable strategy. At the present point, I have to leave this open as an issue for further research.

The notion of a universal depends on the underlying dynamics, and modifying this dynamics may thus lead to different universals. An empirically well-motivated choice would be the kind of stochastic dynamics of finite populations that is studied in Kandori et al. (1993) and Young (1993). Strictly speaking, these models predict that there are no universals at all except for the trivial one comprising the entire state space. However, linguists frequently operate with the notion of a *statistical universal*, which is a property that is shared by almost all languages. This concept could be formalized as a set with the property that an orbit with a randomly chosen initial state, observed at a randomly chosen time, will be within this set with a sufficiently high probability. Under stochastic evolution, this would correspond to a set containing all stochastically stable states and their environments. In fact, van Rooij (2004) and Jäger (2007) employ the notion of stochastic stability to derive certain empirically attested statistical universals from a game theoretic model.

5 An Example: The Evolution of Case Marking

Let me finally illustrate the theoretical notions developed above by an example that is linguistically somewhat more informed than the highly schematic signaling games considered so far. In Jäger (2007), I give a formalization of the typology of case marking systems in terms of a game that is a slight generalization of a signaling game. The presentation of this model here is necessarily rather dense; the interested reader is referred to the mentioned article.

Let us assume that there are two options to morphologically mark the agent argument of a transitive verb, ergative case or nominative case, i.e. the morphological form of subjects of intransitive verbs. Likewise, we assume two options for marking the patient argument, accusative or nominative (sometimes called absolutive case in the context of ergative languages). We furthermore assume that nominative case is unmarked, i.e. less costly than both ergative and accusative marking. So both syntactic core roles in a transitive clause may be marked in a unambiguous but costly or in an ambiguous but cheap way.

We are interested in split systems, i.e. systems where the assignment of overt ergative/accusative marking is conditioned by semantic properties of the NP in question. For simplicity's sake, only those strategies are considered where pronouns may follow a different case marking paradigm than full NPs. Of course non-split strategies are also taken into account. A sender strategy has to specify

1. whether pronominal agents are realized in ergative or nominative,
2. whether non-pronominal agents are realized in ergative or nominative,
3. whether pronominal patients are realized in accusative or nominative, and
4. whether non-pronominal patients are realized in accusative or nominative.

Hence a sender strategy can be represented as a binary 4-tuple, where 1 means "unambiguous marking" (ergative or accusative, depending on the role of the NP), and 0 means "nominative marking". There are 16 such strategies.

When interpreting a transitive clause, the receiver observes whether the two NPs in question are pronominal or not, and their case marking. Based on this information, he has to decide which of the two NPs is agent and which is patient. As a further abstraction, we assume that the receiver has to make this decision solely on the basis of this information, i.e. factors such as word order, semantic plausibility etc. are disregarded. If at least one of the two NPs is in a non-nominative case, the case morphology completely disambiguates the role assignment. If both NPs are in nominative and both have the same status with regard to pronominality, the receiver has no choice but to make a random guess. The only scenario where he can make a strategic choice arises if one NP is a pronoun and the other one a full NP, and both are in nominative case. So essentially there are only two receiver strategies. In the scenario just described, the possible choice are:

1. the pronoun is agent and the full NP is patient (abbreviated as *pA*), or
2. the pronoun is patient and the full NP is agent (abbreviated as *pP*).

The utilities of sender and receiver as assumed to be identical. If the receiver opts for the correct role assignment, both players score a point. Additionally, each occurrence of a non-nominative case marking in a clause incurs a cost for both players.[2]

In Jäger (2007), the probabilities of the four different combinations of syntactic roles and pronominality status are estimated using a corpus study. Furthermore, a range of various differential costs for case marking are considered. In the present context, I will only discuss one configuration, where each case exponent incurs a cost of 0.1. The normalized asymmetric utility matrix then comes out as in Table 1.

Of the 16 sender strategies, only 3 are *strictly undominated* (shown in bold face). These are

- 1100: all agents are marked in ergative and all patients in nominative, i.e. an unconditional ergative system,
- 0110: ergative marking only occurs with full NPs and accusative marking only with pronouns, i.e. a typical double split system, and
- 0011: all agents are in nominative and all patients in accusative, i.e. an unconditional accusative system.

[2] Arguably these costs apply to the sender but not to the receiver. However, it can be shown that the stability properties of the resulting game remain unchanged if signaling costs are assigned to both players.

Table 1 Asymmetric normalized utilities for the case marking game

Sender strategies	Receiver strategies	
	pA	*pP*
1111	0.80	0.80
1110	0.88	0.88
1101	0.82	0.82
1100	**0.90**	**0.90**
1011	0.81	0.81
1010	0.85	0.85
1001	0.81	0.83
1000	0.86	0.87
0111	0.89	0.89
0110	**0.97**	**0.26**
0101	0.81	0.81
0110	0.89	0.18
0011	**0.90**	**0.90**
0010	0.94	0.23
0001	0.81	0.82
0000	0.85	0.15

Table 2 Reduced game

Sender strategies	Receiver strategies	
	pA	*pP*
1100	0.90	**0.90**
0110	**0.97**	0.26

It follows directly from the definition of the replicator dynamics that all trajectories converge to a state where only strictly undominated strategies have a positive probability.[3] Therefore we can restrict attention to the sub-game that only comprises undominated strategies. Also, since 1100 and 0011 have the same utility profile, we can collapse them into a single strategy for the purpose of analysis. I will continue to use the name 0011, with the intended meaning that this covers any mixture of 0011 and 1100. The utility matrix is given in Table 2. This game has two pure Nash equilibria (shown in bold). Additionally, it has infinitely many mixed strategy equilibria. There is a threshold $\theta \approx 0.90$ such that each mixed strategy with $p_s(0110) = 0$ and $p_r(pP) > \theta$ is also a Nash equilibrium.

This structure carries over the symmetrized version of this game, shown in Table 3: This game has two pure symmetric Nash equilibria as well (shown in bold). Additionally, each mixed strategy with $p(1100/pA) + p(1100/pP) = 1$ and $p(1100/pP) \geq \theta$ is a symmetric Nash equilibrium.

As was pointed out in Jäger (2007), the equilibrium $0110/pA$ is the only evolutionarily stable strategy in this game (as well as in the larger original game). This

[3] This carries to the symmetrized version of the game, i.e. in the symmetrized game only those agents survive that play an undominated strategy in the sender role.

Table 3 Symmetrized reduced game

	1100/*pA*	0110/*pA*	1100/*pP*	0110/*pP*
1100/*pA*	0.90	0.93	0.90	0.93
0110/*pA*	0.93	**0.97**	0.58	0.61
1100/*pP*	0.90	0.58	**0.90**	0.58
0110/*pP*	0.93	0.61	0.58	0.26

does not entail though that a population will evolve towards this strategy from each initial point. The set of strategies with $p(1100/pA) + p(0011/pP) = 1$ and $p(1100/pP) > \theta$ (note the strict inequality!) consists exclusively of neutrally stable strategies. With an argument along the lines of Pawlowitsch (2008) it can straightforwardly be shown that this continuum of neutrally stable strategies attracts a positive measure of the state space.

As every orbit in the interior of the state space converges to some Nash equilibrium, we can conclude that the set of Nash equilibria of this game (and therefore also of the original game comprising all 16 sender strategies) is the smallest universal here. The fact that the notion of a universal is more inclusive than the set of evolutionarily stable strategies is empirically well-justified here. The neutrally stable strategies of this game represent case marking systems without splits. These are empirically well-attested, for instance with a pure accusative system in Hungarian and a pure ergative system in the past tense paradigm of Burushaski.

6 Conclusion

To wrap up, I tried to argue that the notion of a linguistic universal can be illuminated by considering a (cultural) evolutionary dynamics, because this approach promises to give a causal explanation for seemingly functionally motivated features of natural languages. Evolutionary game theory is a suitable mathematical framework to carry out this programme because the dynamics of language use arguably involves frequency dependent selection, and because linguistic interaction can naturally be formalized by game theoretic means. I have to close with a note of caution though: to actually derive universals analytically from a game theoretic model, it is not sufficient to study static equilibrium properties. Rather, a careful exploration of the dynamic properties of the model is inevitable.

Acknowledgments Thanks to Simon Huttegger, Christina Pawlowitsch and two anonymous reviewers for important comments on an earlier version of the article.

Appendix

Numerical estimation of the size of the basin of attraction As mentioned in the text, the joint basin of attraction of the set of neutrally but not evolutionarily stable strategies is about 1.2 % of the entire strategy space. The result was obtained by using a

Monte-Carlo method. A random initial state s was picked out by a random generator according to the uniform distribution over the 729-dimensional simplex, the replicator dynamics was solved numerically for the initial condition s, and the asymptotic behavior was analyzed at $t = 1000$ (by which time all time series have converged towards a rest point, within the limits of precision imposed by the numerical algorithm used). This procedure was repeated 5,000 times. It turned out that in 4,941 cases the solution converged towards an evolutionarily stable state, and in 59 cases to a neutrally but not evolutionarily stable state. This means that with a confidence of 95 %, the true value lies between 0.9 and 1.5 %, with a maximum likelihood estimation of 1.2 %.

Basins of attraction The utility matrix of the game in question is:

$$A = \begin{pmatrix} 2 & 2 & 2 \\ 2 & 3 & -1 \\ 2 & -1 & 0 \end{pmatrix}$$

This game is doubly symmetric, and it has two Nash equilibria: e_1 and e_2. Consider the linear manifold $E = \{x \in \text{int}S_3 : x_3 = 4x_2\}$. For points $x \in E$, we have

$$\begin{aligned} \dot{x}_2 &= x_2(2x_1 + 3x_2 - x_3 - x \cdot Ax) \\ &= x_2(2x_1 - x_2 - x \cdot Ax) \\ \dot{x}_3 &= x_3(2x_1 - x_2 - x \cdot Ax) \\ &= 4x_2(2x_1 - x_2 - x \cdot Ax), \end{aligned}$$

and thus $\dot{x}_3 = 4\dot{x}_2$. So $\dot{x}$ is always tangential to E, and E is thus an invariant set. As e_1 is the only Nash equilibrium within the closure of this set, all points in E converge towards e_1. Now consider the set $F = \{x \in \text{int}S_3 : x_3 > 4x_2\}$. Since e_2 is not in the closure of this set, a trajectory starting in F and converging towards e_2 would have to leave F. This is impossible though because all interior points remain within the interior under the replicator dynamics—so no orbit can touch a boundary face—and a trajectory cannot cross E due to uniqueness of solutions of autonomous differential equations. We thus conclude that all points in F converge towards e_1. Since F has a positive Lebesgue measure, e_1 has a basin of attraction that is not a null set. Nevertheless e_1 is not neutrally stable, because

$$\begin{aligned} e_2 \cdot Ae_1 &= e_1 \cdot Ae_1 \\ e_2 \cdot Ae_2 &> e_1 \cdot Ae_2 \end{aligned}$$

In Fig. A.1 the phase portrait of the game is sketched. The bold line indicates the boundary between the basins of attraction of e_1 and e_2. It is easy to see that e_1 is

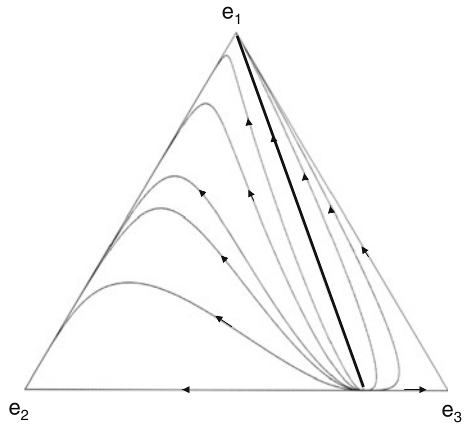


Fig. A.1 Phase portrait

not Lyapunov stable (because every open environment has a non-empty intersection with the basin of attraction of e_2), but nevertheless attracts a non-null set.

References

Akin, E., & Hofbauer, J. (1982). Recurrence of the unfit. *Mathematical Biosciences*, *61*, 51–62.

Barlow, M., & Kemmer, S. (Eds.). (2000). *Usage-based models of language*. Stanford: CSLI Publications.

Bomze, I. M., & Weibull, J. W. (1995). Does neutral stability imply Lyapunov stability? *Games and Economic Behavior*, *11*(2), 173–192.

Bomze, I. (2002). Regularity versus degeneracy in dynamics, games, and optimization: A unified approach to different aspects. *SIAM Review*, *44*(3), 394–441.

Bybee, J. (2001). *Phonology and language use*. Cambridge, UK: Cambridge University Press.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Haspelmath, M. (1999). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft*, *18*(2), 180–205.

Hofbauer, J., & Sigmund, K. (1988). *The theory of evolution and dynamical systems*. Cambridge: Cambridge University Press.

Hofbauer, J., Schuster, P., & Sigmund, K. (1979). A note on evolutionarily stable strategies and game dynamics. *Journal of Theoretical Biology*, *81*(3), 609–612.

Hofbauer, J., & Sigmund, K. (1998). *Evolutionary games and population dynamics*. Cambridge, UK: Cambridge University Press.

Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, *77*, 187–222.

Huttegger, S. H. (2007). Evolution and the explanation of meaning. *Philosophy of Science*, *74*, 1–27.

Jäger, G. (2007). Evolutionary game theory and typology: A case study. *Language*, *83*(1), 74–109.

Kandori, M., Mailath, G., & Rob, R. (1993). Learning, mutation, and long-run equilibria in games. *Econometrica*, *61*, 29–56.

Lewis, D. (1969). *Convention*. Cambridge, Mass: Harvard University Press.
Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, *96*(14), 8028–8033.
Pawlowitsch, C. (2008). Why evolution does not always lead to an optimal signaling system. *Games and Economic Behavior*, *63*, 203–226.
Taylor, P., & Jonker, L. (1978). Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, *40*, 145–156.
Thomas, B. (1985). On evolutionarily stable sets. *Journal of Mathematical Biology*, *22*, 105–115.
Trapa, P., & Nowak, M. (2000). Nash equilibria for an evolutionary language game. *Journal of Mathematical Biology*, *41*, 172–188.
van Rooij, R. (2004). Signalling games select Horn strategies. *Linguistics and Philosophy*, *27*, 493–527.
Wärneryd, K. (1993). Cheap talk, coordination and evolutionary stability. *Games and Economic Behavior*, *5*, 532–546.
Young, H. P. (1993). The evolution of conventions. *Econometrica*, *61*, 57–84.
Zipf, G. (1935). *The psycho-biology of language*. Cambridge, Massachusetts: MIT Press.
Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge: Addison Wesley.

Continuation Hierarchy and Quantifier Scope

Oleg Kiselyov and Chung-chieh Shan

Abstract We present a directly compositional and type-directed analysis of quantifier ambiguity, scope islands, wide-scope indefinites and inverse linking. It is based on Danvy and Filinski's continuation hierarchy, with deterministic semantic composition rules that are *uniquely* determined by the formation rules of the overt syntax. We thus obtain a compositional, uniform and parsimonious treatment of quantifiers in subject, object, embedded-NP and embedded-clause positions without resorting to Logical Forms, Cooper storage, type-shifting and other ad hoc mechanisms. To safely combine the continuation hierarchy with quantification, we give a precise logical meaning to often used informal devices such as picking a variable and binding it off. Type inference determines variable names, banishing "unbound traces". Quantifier ambiguity arises in our analysis solely because quantifier words are polysemous, or come in several strengths. The continuation hierarchy lets us assign strengths to quantifiers, which determines their scope. Indefinites and universals differ in their scoping behavior because their lexical entries are assigned different strengths. PPs and embedded clauses, like the main clause, delimit the scope of embedded quantifiers. Unlike the main clause, their limit extends only up to a certain hierarchy level, letting higher-level quantifiers escape and take wider scope. This interplay of strength and islands accounts for the complex quantifier scope phenomena. We present an economical "direct style", or continuation hierarchy on-demand, in which quantifier-free lexical entries and phrases keep their simple, unlifted types.

Keywords Semantics · Continuation semantics · Quantifier scope · Quantifier ambiguity · Continuation hierarchy · CPS · Delimited continuation · Direct compositionality

O. Kiselyov (✉)
University of Tsukuba, Tsukuba, Japan
e-mail: oleg@okmij.org

C.-c. Shan
School of Informatics and Computing, Indiana University, Bloomington, IN, USA
e-mail: ccshan@indiana.edu

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*, Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_6,

1 Introduction

The proper treatment of quantification has become a large research area ever since Montague called attention to "the puzzling cases of quantification and reference" back in 1974 (Montague 1974). The impressive breadth of the area is evident from two recent surveys (Szabolcsi 2000, 2009), which concentrate only on interactions of quantifier phrases among themselves (leaving out, for example, binding of pronouns by quantifiers). The two surveys collect a great amount of empirical data—more and more puzzles. There is also a great number of proposals for a theory to explain the puzzles. And yet even the basic features of the theory remain undecided. In the conclusion of her survey (Szabolcsi 2000) poses the following three challenges that call for significant new research:

1. "develop the tools, logical as well as syntactic, that are necessary to account for the whole range of existing readings;"
2. "draw the proper empirical distinction between readings that are actually available and those that are not;"
3. determine "whether 'spell-out syntax' is sufficient for the above two purposes" [in other words, if quantifier scope can be determined without resorting to Logical Form]

This chapter takes on the challenges and develops a logical tool that is expressive to capture empirical data—available and unavailable readings—for a range of quantifier phenomena, from quantifier ambiguity to scope islands, wide-scope indefinites and inverse linking. The "spell-out syntax" proved sufficient: we directly compose meanings that are model-theoretic, not trees. There is quite more work yet to do. Future work dealing with numeric and downward-entailing quantifiers, plural indefinites, and quantificational binding will hopefully clarify presently ad hoc parameters such as the number of hierarchy levels.

1.1 What is Quantifier Scope

"The scope of an operator is the domain within which it has the ability to affect the interpretation of other expressions" (Szabolcsi 2000, Sect. 1.1). In this chapter, we concentrate on how a quantifier affects the interpretation of another quantified phrase. For example,

(1) I showed every boy a planet.

has the reading that I showed each boy a possibly different planet. The quantifier 'every' affected the interpretation of 'a planet', which refers to a possibly different planet for a different boy. That reading is called *linear* scope. The sentence has another—*inverse*—reading, whereupon each boy was shown the same planet. The example thus exhibits quantifier ambiguity. Although the inverse-scope reading of (1) entails the linear reading (which lead to doubts if inverse readings have to be accounted for at all (Reinhart 1979)), this is not always the case. For example, the linear and inverse readings of

(2) Two of the students attended three of the seminars.

(3) Neither student attended a seminar on rectangular circles.

do not entail each other. Szabolcsi (2000) demonstrates solid inverse-scope readings on many more examples. A theory of scope must also explain why no quantifier ambiguity arises in examples like

(4) That every boy left upset a teacher.

(5) Someone reported that John saw everyone.

(6) Some professor admires every student and hates the Dean.

and yet other examples with a quantifier within an embedded clause, such as

(7) Everyone reported that [Max and some lady] disappeared.

are ambiguous. Szabolcsi argues (Szabolcsi 2000, Sect. 3.2) that "different quantifier types have different scope-taking abilities". The theory should therefore support lexical entries for quantifiers that take scope differently and compositionally in relation to each other. The present chapter describes such a theory.

1.2 Why Continuations

Our theory of quantifier scope is based on *continuation semantics*, which emerged (Barker 2002; de Groote 2001) as a compelling alternative to traditional approaches to quantification—Montague's proper treatment, Quantifier Raising (QR), type-shifting (surveyed by Barker (2002))—as well as to the Minimalism views (surveyed by Szabolcsi (2000); she also extensively discusses QR and its empirical inadequacy). Continuation semantics is compelling because it can interpret quantificational NPs (QNPs) compositionally in situ, without type-shifting, Cooper storage, or building any structures like Logical Forms beyond overt syntax. Accordingly, QNPs in subject and other positions are treated the same, QNPs and NPs are treated the same, and scope taking is semantic. Central to the approach is the hypothesis that "some linguistic expressions (in particular, QNPs) have denotations that manipulate their own continuations" (Barker 2002, Sect. 1). Although continuation semantics is only a decade old, its origin can be traced to Montague's proper treatment: "saying that NPs denote generalized quantifiers amounts to saying that NPs denote functions on their own continuations" (Barker 2002, Sect. 2.2; see also de Groote (2001)) . Several continuation approaches have been developed since Barker (2002), using so-called control operators (de Groote 2001; Shan 2007a; Bekki and Asai 2009) or Lambek-Grishin calculus (Bernardi and Moortgat 2010).

1.3 Contributions

Like all continuation approaches, our theory features a compositional, uniform and in-situ analysis of QNPs in object, subject and other positions. Moreover, we address the following open issues.

Inverse scope, scope islands and wide-scope indefinites One way to account for these phenomena is to combine control operators with metalinguistic quotation (Shan 2007b). More common—see for example Shan (2004)—is using a continuation hierarchy, such as Danvy and Filinski's (D&F) hierarchy (Danvy and Filinski 1990), which has been thoroughly investigated in the Computer Science theory. The common problem, which has not been addressed in the metalinguistic quotation and the previous D&F hierarchy approaches, is avoiding "unbound traces"—preventing denotations with unbound variables. Barker and Shan's essentially 'variable-free' semantics (Barker and Shan 2008) side-steps the unbound traces problem altogether. However it relies on a different and little investigated hierarchy. The corresponding direct-style (see the next point) is unknown.
Our approach is the first to give a rigorous account of inverse scope, scope islands and wide-scope indefinites using the D&F hierarchy. We rely on types to prevent unbound traces. We formalize the pervasive intuition that a QNP is represented by a trace (QR), pronoun (Montague) or variable (Cooper storage) that gets bound somehow. We make this intuition precise and give it logical meaning, banishing unbound traces once and for all.

Direct style In Barker's continuation approach (Barker 2002), every constituent's denotation explicitly receives its continuation, even though few constituents need to manipulate these continuations. Combining such *continuation-passing-style* (CPS) denotations is quite cumbersome, as we see in Sect. 2.2. Thus, we would like to avoid CPS denotations for quantifier-free constituents, in particular, for lexical entries other than quantifiers. *Direct-style* continuation semantics lets us combine continuation-manipulating denotations directly with ordinary denotations, simplifying analyses and keeping most lexical entries 'uncomplicated', which we illustrate in Sect. 2.3.
We present a version of direct-style for the D&F hierarchy. Unlike other direct-style approaches (Shan 2007a, b), ours uses the ordinary λ-calculus and denotational semantics rather than operational semantics and a calculus with control operators. Our treatment of inverse scope relies on the properties of the D&F hierarchy extensively, as detailed in Sect. 4.

Source of quantifier ambiguity It is common to explain quantifier ambiguity by the nondeterminism of semantic composition rules (Barker 2002; de Groote 2001). One syntactic formation operation may correspond to several semantic composition functions, or the analysis may include operators like 'lift' or 'wrap' that may be freely applied to any denotation.
In contrast, our semantic composition rules are all deterministic. Although we extensively rely on schematic rules to ease notation and emphasize commonality, how these schemas are instantiated is determined unambiguously by types.

Furthermore, our analysis has no optional or freely applicable rules or semantic combinators. Each syntactic formation operation maps to a unique semantic composition operation, and *vice versa*: each operation on denotations has a syntactic counterpart. This one-to-one correspondence between surface syntax and semantic composition underlies our entire approach—which is thus *directly compositional*. (See Sect. 6 of Barker (2002) for discussion of compositionality and how nondeterminism in semantic composition rules constitutes a threat.)
The source of quantifier ambiguity in our approach is solely in the lexical entries for the quantifier words rather than in the rules of syntactic formation or semantic composition. Different lexical entries for the same quantifier word have denotations corresponding to different levels of the continuation hierarchy, thus having different *strength*, or ability to scope over wider contexts.[1]
One advantage of our approach is better control over overgeneration: when only lexical entries are ambiguous, it is easier to see all available denotations and hence assure against overgeneration.

To summarize: our contribution is a directly compositional analysis of quantifier ambiguity, scope islands, inverse linking and wide-scope indefinites in the D&F continuation hierarchy, in direct style, without risking unbound traces, and using deterministic semantic composition rules. We analyze QNP in situ and compositionally, relying on no structure beyond the overt syntax. All non-determinism is in the choice of lexical entries for quantifier words. The presentation uses the familiar denotational semantics.

1.4 The Structure of the Chapter

The warm-up Sect. 2 gradually introduces continuation semantics on a small fragment and explains our notation and terminology. Section 2.3 presents the direct-style continuation semantics as an economical CPS-on-demand. We treat bound variables rigorously in Sect. 3, with type annotations to infer variable names and to prevent unbound variables in final denotations. Section 4 presents the continuation hierarchy and uses it to analyze quantifier ambiguity. The corresponding direct-style, or CPS hierarchy on-demand, is described in Sect. 4.2. Scope islands, wide-scope indefinites and briefly inverse linking are the subject of Sect. 5.

For illustrations we use a small fragment of English with context-free syntax and extensional semantics, extending and refining the fragment throughout the chapter. Figure 1 shows the relationship between the fragments, illustrating parallel development in CPS and direct style.

[1] The different lexical entries for the same quantifier have a regular structure. In fact, all higher-strength quantifier entries are mechanically derived from the entry for the lowest-strength quantifier, as shown in Figs. 12, 13. The number of lexical entries, that is, the assignment of the levels of strength to a quantifier is determined from empirical data.

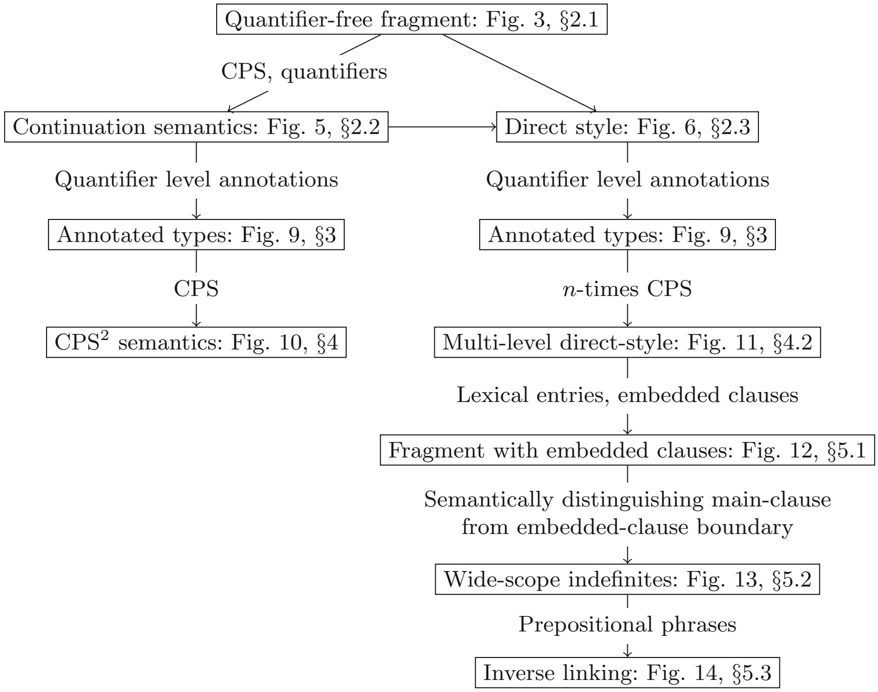


Fig. 1 Relationship between the fragments used in the chapter

The continuation hierarchy of quantifier scope described in the chapter has been implemented. The complete Haskell code is available online at http://okmij.org/ftp/gengo/QuanCPS.hs. The file implements the fragment of the chapter in the spirit of the Penn Lambda Calculator (Champollion et al. 2007), letting the user write parse trees and determine their denotations. We have used our semantic calculator for all the examples in the chapter.

2 Warm-Up: The Proper Continuation Treatment of Quantifiers

In this warm-up section, we recall Barker's continuation semantics (Barker 2002) and summarize it in our notation. Alongside, we also introduce Barker and Shan's continuation semantics (Barker and Shan 2004; Shan 2007a) in direct style, which avoids pervasive type lifting of lexical entries. We use the simplicity of the examples to introduce notation and calculi to be used in further sections.

$$
\begin{array}{llcl}
\text{Base types} & \upsilon & ::= & e \mid t \\
\text{Types} & \sigma & ::= & \upsilon \mid (\sigma\sigma) \\
\\
\text{Constants} & c & ::= & \wedge \mid \vee \mid \Rightarrow \mid \neg \mid \mathbf{john} \mid \mathbf{mary} \mid \mathbf{see} \mid \ldots \\
\text{Expressions} & d & ::= & c \mid d \centerdot d
\end{array}
$$

Fig. 2 The language $\mathcal{D}$ of denotations

Syntax	Semantic type	Denotation $[\![\cdot]\!]$
M → S .	t	$(\!|[\![\mathsf{S}]\!]|\!)$
S → NP VP	t	$[\![\mathsf{NP}]\!] \,^{<}\, [\![\mathsf{VP}]\!]$
VP → Vt NP	et	$[\![\mathsf{Vt}]\!] \,^{>}\, [\![\mathsf{NP}]\!]$
NP → John	e	**john**
NP → Mary	e	**mary**
VP → left	et	**leave**
Vt → saw	$e(et)$	**see**

Fig. 3 Syntax and direct semantics for a small quantifier-free fragment

2.1 Direct Semantics

Like Barker (2002), we start with a simple, quantifier-free fragment, with context-free syntax and extensional semantics. The language of denotations is a plain higher-order language, Fig. 2 with the obvious model-theoretical interpretation. The language has base types e and t and function types, for example $(e(et))$. We will often omit outer parentheses. Expressions (denoted by 'non-terminal' d) comprise constants (denoted by c) and applications $d_1 \cdot d_2$, which are left associative: $d_1 \cdot d_2 \cdot d_3$ stands for $(d_1 \cdot d_2) \cdot d_3$. Constants are logical constants (negation, etc) and domain constants (such as **john**). Logical connectives $\wedge$ (conjunction), $\vee$ (disjunction), $\Longrightarrow$ (implication) are constants of the type $t(tt)$, whose applications are written in infix, for example, $d_1 \wedge d_2$.

The syntax and semantics for our fragment is given in Fig. 3. The syntax formation operation Merge corresponds to forward application $^{>}$ or backward application $^{<}$ in semantics (see (8)).[2] The notation $d_1 {}^{>} d_2$ says nothing at all whether d_1 takes scope over d_2. The category M stands for the complete (matrix) sentence, terminated by the period. The corresponding semantic operation is $(\!|\cdot|\!)$. For now, these semantic composition operations are defined as follows:

[2] In our simple context-free syntax, the choice of forward or backward application is determined by the semantic types. If we used combinatorial categorial grammar (CCG), the choice of the application is evident from the categories of the nodes being combined.

(8)
$$d_1 {}^{>} d_2 \stackrel{\text{def}}{=} d_1 \cdot d_2$$
$$d_1 {}^{<} d_2 \stackrel{\text{def}}{=} d_2 \cdot d_1$$
$$\llparenthesis e \rrparenthesis \stackrel{\text{def}}{=} e$$

We extend these definitions in Sect. 2.2 when we add quantifiers, and we extend the definition of $\llparenthesis \cdot \rrparenthesis$ a few more times. It will become clear then that the latter semantic operation is not vacuous at all. Finally, Sect. 5.1 will make it clear that $\llparenthesis \cdot \rrparenthesis$ plays the role of the delimiter of the quantifier scope.

Figure 3 and the similar figures in the following sections demonstrate that each syntactic formation operation maps to a semantic composition operation and *vice versa*: each operation on denotations is reflected in syntax. This one-to-one syntax-semantic composition correspondence underlies our entire approach. We easily determine the denotation of a sample sentence

(9) [$_{\mathsf{M}}$ [$_{\mathsf{S}}$ [$_{\mathsf{NP}}$ John] [$_{\mathsf{VP}}$[$_{\mathsf{Vt}}$ saw] [$_{\mathsf{NP}}$ Mary]]].]
to be **see** · **mary** · **john**.

2.2 CPS Semantics

We now review continuation semantics, which lets us add quantifiers to our fragment. Barker (2002) has argued that the denotations of quantified phrases need access to their context. Here is a simple illustration. Suppose we had a magic domain constant **everyone** as the denotation of everyone. We could write the meaning of [$_{\mathsf{M}}$ [$_{\mathsf{S}}$ John [$_{\mathsf{VP}}$ saw [$_{\mathsf{NP}}$ everyone]]].] as $\llparenthesis$**see** · **everyone** · **john**$\rrparenthesis$, whose model-theoretical interpretation must be the same as that of the logical formula $\forall x.$ **see** · x · **john**. Removing **everyone** from $\llparenthesis$**see** · **everyone** · **john**$\rrparenthesis$ leaves the "term with a hole" $\llparenthesis$**see** · [] · **john**$\rrparenthesis$—the *context* of **everyone** in the original term. We intuit that **everyone** manages to grab its context, up to the enclosing $\llparenthesis \cdot \rrparenthesis$, and quantify over it.

To give each term the ability to grab its context, we write the terms in a *continuation-passing style* (CPS), whereupon each expression receives as an argument its context represented as a function, or *continuation*. Before we can write any CPS term, we have to resolve a small problem. To represent contexts we have to be able to build functions—an operation our language of denotations $\mathcal{D}$ (Fig. 2) does not support. Therefore, we "inject" $\mathcal{D}$ into the full λ-calculus, with λ-abstractions. This calculus, or language $\mathcal{L}$, is presented in Fig. 4.

The expressions of the language $\mathcal{D}$ (Fig. 2) are all constants of the λ-calculus $\mathcal{L}$; the types of $\mathcal{D}$ are all base types of $\mathcal{L}$. In this sense, $\mathcal{D}$ is embedded in $\mathcal{L}$. The language $\mathcal{L}$ has its own function types, written with an arrow $\rightarrow$. Distinguishing two kinds of function types makes the continuation argument stand out in CPS terms as well as types. We exploit this distinction in Sect. 2.3.

We take $\rightarrow$ to be right associative and hence we write $t \rightarrow (t \rightarrow t)$ as $t \rightarrow t \rightarrow t$. Besides the constants, $\mathcal{L}$ has variables, abstractions and applications. The application is again left associative, with $m_1 m_2 m_3$ standing for $(m_1 m_2) m_3$.

Types	$\tau ::= \sigma \mid \tau \to \tau$		
Variables	x, y, z, v, f, k		
Expressions	$m ::= d \mid x \mid \lambda x.\, m \mid m\ m$		
Reductions $m \rightsquigarrow m'$	$(\lambda x.\, m)m' \rightsquigarrow m\{x \mapsto m'\}$	(β)	

Fig. 4 Simply-typed λ-calculus, the language $\mathcal{L}$. (Base types σ and constants d are introduced in Fig. 2)

Syntax			Semantic type	Denotation $[\![\cdot]\!]$
M	→	S .	t	$(\![\![\mathsf{S}]\!]\!)$
S	→	NP VP	$(t \to t) \to t$	$[\![\mathsf{NP}]\!]\ {}^{<}\ [\![\mathsf{VP}]\!]$
VP	→	Vt NP	$((et) \to t) \to t$	$[\![\mathsf{Vt}]\!]\ {}^{>}\ [\![\mathsf{NP}]\!]$
NP	→	John	$(e \to t) \to t$	$\lambda k.\, k\ \mathbf{john}$
NP	→	Mary	$(e \to t) \to t$	$\lambda k.\, k\ \mathbf{mary}$
VP	→	left	$((et) \to t) \to t$	$\lambda k.\, k\ \mathbf{leave}$
Vt	→	saw	$((e(et)) \to t) \to t$	$\lambda k.\, k\ \mathbf{see}$
NP	→	everyone	$(e \to t) \to t$	$\lambda k.\, \forall x.\, k\ x$
NP	→	someone	$(e \to t) \to t$	$\lambda k.\, \exists x.\, k\ x$

Fig. 5 Syntax and continuation semantics for the small fragment

$\mathcal{L}$ is the full λ-calculus and has reductions, $m \rightsquigarrow m'$. An expression is in normal form if no reduction applies to it or any of its sub-expressions. The notation $m\{x \mapsto m'\}$ in the β-reduction rule stands for the capture-avoiding substitution of m' for x in m. A unique normal form always exists and can be reached by any sequence of reductions; in other words, $\mathcal{L}$ is strongly normalizing.

We are set to write CPS denotations for our fragment. Constants like **john** have little to do but to "plug themselves" into their context: $\lambda k.\, k\ \mathbf{john}$.[3] Here k represents the context of **john** within the whole sentence denotation. The whole denotation must be of the type t; hence k has the type $e \to t$ and the type of the CPS form of **john** is $(e \to t) \to t$. With the CPS denotations, our fragment now reads as in Fig. 5. The semantic composition operators are now defined as follows.

(10)
$$m_1{}^{>}m_2 \stackrel{\text{def}}{=} \lambda k.\, m_1(\lambda f.\, m_2(\lambda x.\, k(f \bullet x)))$$
$$m_1{}^{<}m_2 \stackrel{\text{def}}{=} \lambda k.\, m_1(\lambda x.\, m_2(\lambda f.\, k(f \bullet x)))$$
$$(\!|m|\!) \stackrel{\text{def}}{=} m(\lambda v.\, v)$$

The CPS form of $m_1{}^{>}m_2$ is $\lambda k.\, m_1(\lambda f.\, m_2(\lambda x.\, k\ (f \bullet x)))$: it fills its context k with $f \bullet x$, where f is what m_1 fills its context with, and x is what m_2 fills its context with.

Using Fig. 5 to compute the denotation of the sample sentence (9) gives us:

[3] When a context is represented by a continuation function k, filling the hole in the context with a term e—or, plugging e into the context—is represented by the application $k\ e$.

(11) $[\![[_{\mathsf{M}}\ [_{\mathsf{S}}\ [_{\mathsf{NP}}\ \mathsf{John}]\ [_{\mathsf{VP}}\ [_{\mathsf{Vt}}\ \mathsf{saw}]\ [_{\mathsf{NP}}\ \mathsf{Mary}]]].]]\!]$

$$= (\lambda k_0.\, (\lambda k.\, k\ \mathbf{john})(\lambda x.\, (\lambda k_1.\, (\lambda k.\, k\ \mathbf{see})(\lambda f'.\, (\lambda k.\, k\ \mathbf{mary})(\lambda x'.\, k_1\ (f' \cdot x')))) (\lambda f.\, k_0\ (f \cdot x)))) (\lambda v.\, v)$$

$$\rightsquigarrow (\lambda k_0.\, (\lambda k.\, k\ \mathbf{john})(\lambda x.\, (\lambda k_1.\, (\lambda k.\, k\ \mathbf{mary})(\lambda x'.\, k_1\ (\mathbf{see} \cdot x'))) (\lambda f.\, k_0\ (f \cdot x)))) (\lambda v.\, v)$$

$$\rightsquigarrow (\lambda k_0.\, (\lambda k.\, k\ \mathbf{john})(\lambda x.\, (\lambda k_1.\, k_1\ (\mathbf{see} \cdot \mathbf{mary})) (\lambda f.\, k_0\ (f \cdot x)))) (\lambda v.\, v)$$

$$\rightsquigarrow (\lambda k_0.\, (\lambda k.\, k\ \mathbf{john})(\lambda x.\, (k_0\ ((\mathbf{see} \cdot \mathbf{mary}) \cdot x)))) (\lambda v.\, v)$$

$$\rightsquigarrow (\lambda k_0.\, (k_0\ ((\mathbf{see} \cdot \mathbf{mary}) \cdot \mathbf{john}))) (\lambda v.\, v)$$

$$\rightsquigarrow ((\mathbf{see} \cdot \mathbf{mary}) \cdot \mathbf{john})$$

The β-reductions lead to the same expression $((\mathbf{see} \cdot \mathbf{mary}) \cdot \mathbf{john})$ as in Sect. 2.1. The argument k_1 was the continuation of $[\![\mathsf{saw\ Mary}]\!]$. The term $(\lambda k_0. \ldots)$ was the denotation of the main clause $[_{\mathsf{S}}\ \mathsf{John}\ [_{\mathsf{VP}}\ \mathsf{saw\ Mary}]]$, whose context is empty, represented by $\lambda v.\, v$. (If the clause were an embedded one, its context would not have been empty. We discuss embedded clauses in Sect. 5.1.)

Figure 5 contains two extra rows, not present in Fig. 3: The CPS semantics lets us express QNPs. The denotation of $\mathsf{everyone}$, $\lambda k.\, \forall x.\, k\ x$, is what we have informally argued at the beginning of Sect. 2.2 the denotation of $\mathsf{everyone}$ should be: the quantifier grabs its continuation k and quantifies over it. The denotation is a bit sloppy since we have not yet introduced quantifiers in any of our languages, $\mathcal{D}$ or $\mathcal{L}$. Such an informal style, appealing to predicate logic, is very common. For now, we go along; we come back to this point in Sect. 3, arguing that it pays to be formal. Let us see how quantification works:

(12) $[\![[_{\mathsf{M}}\ [_{\mathsf{S}}\ [_{\mathsf{NP}}\ \mathsf{John}]\ [_{\mathsf{VP}}\ [_{\mathsf{Vt}}\ \mathsf{saw}]\ [_{\mathsf{NP}}\ \mathsf{everyone}]]].]]\!]$

$$= (\lambda k_0.\, (\lambda k.\, k\ \mathbf{john})(\lambda x.\, (\lambda k_1.\, (\lambda k.\, k\ \mathbf{see})(\lambda f'.\, (\lambda k.\, \forall x''.\, k\ x'')(\lambda x'.\, k_1\ (f' \cdot x')))) (\lambda f.\, k_0\ (f \cdot x)))) (\lambda v.\, v)$$

$$\rightsquigarrow (\lambda k_0.\, (\lambda k_1.\, (\lambda k.\, \forall x''.\, k\ x'')(\lambda x'.\, k_1\ (\mathbf{see} \cdot x'))) (\lambda f.\, k_0\ (f \cdot \mathbf{john}))) (\lambda v.\, v)$$

$$\begin{aligned}
&\rightsquigarrow (\lambda k_0.\, (\lambda k_1.\, \forall x''.\, k_1\ (\mathbf{see} \cdot x''))\\
&\qquad (\lambda f.\, k_0\ (f \cdot \mathbf{john})))\\
&\quad (\lambda v.\, v)\\
&\rightsquigarrow (\lambda k_0.\, \forall x''.\, k_0\ (\mathbf{see} \cdot x'') \cdot \mathbf{john})\\
&\quad (\lambda v.\, v)\\
&\rightsquigarrow \forall x''.\, (\mathbf{see} \cdot x'') \cdot \mathbf{john}
\end{aligned}$$

The sample sentence “John saw everyone” had the quantifier in the object position, and yet we, unlike Montague, did not have to do anything special to accommodate it. In fact, comparing (11) against (12) shows that everyone is treated just like Mary. The β-reductions accumulate the context captured by the quantifier until it eventually becomes the full sentence context.

A quantifier in the subject position, unlike with QR, is treated just like a quantifier in the object position:

(13) $[\![[_{\mathsf{M}}\ [_{\mathsf{S}}\ [_{\mathsf{NP}}\ \mathsf{Someone}]\ [_{\mathsf{VP}}\ [_{\mathsf{Vt}}\ \mathsf{saw}]\ [_{\mathsf{NP}}\ \mathsf{everyone}]]].]]\!]$

$$\begin{aligned}
&= (\lambda k_0.\, (\lambda k.\, \exists y.\, k\ y)(\lambda x.\\
&\qquad (\lambda k_1.\, (\lambda k.\, k\ \mathbf{see})(\lambda f'.\, (\lambda k.\, \forall x''.\, k\ x'')(\lambda x'.\, k_1\ (f' \cdot x'))))\\
&\qquad (\lambda f.\, k_0\ (f \cdot x))))\\
&\quad (\lambda v.\, v)\\
&\rightsquigarrow (\lambda k_0.\, (\lambda k.\, \exists y.\, k\ y)(\lambda x.\\
&\qquad (\lambda k_1.\, \forall x''.\, k_1\ (\mathbf{see} \cdot x''))\\
&\qquad (\lambda f.\, k_0\ (f \cdot x))))\\
&\quad (\lambda v.\, v)\\
&\rightsquigarrow (\lambda k_0.\, \exists y.\, (\lambda k_1.\, \forall x''.\, k_1\ (\mathbf{see} \cdot x''))(\lambda f.\, k_0\ (f \cdot y)))\\
&\quad (\lambda v.\, v)\\
&\rightsquigarrow (\lambda k_0.\, \exists y.\, \forall x''.\, k_0((\mathbf{see} \cdot x'') \cdot y))\\
&\quad (\lambda v.\, v)\\
&\rightsquigarrow \exists y.\, \forall x''.\, (\mathbf{see} \cdot x'') \cdot y
\end{aligned}$$

Thus, continuation semantics can treat QNPs in any syntactic position with no type-shifting and no surgery on the syntactic derivation. The resulting denotation for “Someone saw everyone” is the linear-scope reading. Deriving the inverse-scope reading is the subject of Sect. 4.

2.3 Direct-Style Continuation Semantics

This section describes a “direct style” advocated by Barker and Shan (2004) and Shan (2004, 2007a). Its great appeal is in simple, non-CPS denotations for quantifier-free phrases. In particular, lexical entries other than quantifiers keep their straightforward

Syntax			Semantic type	Denotation $[\![\cdot]\!]$
M	$\rightarrow$	S .	t	$(\![[\![\mathsf{S}]\!]]\!)$
S	$\rightarrow$	NP VP	t or $(t \to t) \to t$	$[\![\mathsf{NP}]\!] {}^{<} [\![\mathsf{VP}]\!]$
VP	$\rightarrow$	Vt NP	et or $((et) \to t) \to t$	$[\![\mathsf{Vt}]\!] {}^{>} [\![\mathsf{NP}]\!]$
NP	$\rightarrow$	John	e	**john**
NP	$\rightarrow$	Mary	e	**mary**
VP	$\rightarrow$	left	et	**leave**
Vt	$\rightarrow$	saw	$e(et)$	**see**
NP	$\rightarrow$	everyone	$(e \to t) \to t$	$\lambda k.\, \forall x. k\ x$
NP	$\rightarrow$	someone	$(e \to t) \to t$	$\lambda k.\, \exists x. k\ x$

Fig. 6 Syntax and direct-style continuation semantics for the small fragment: the merger of Figs. 3 and 5. Lexical entries other than the quantifiers keep the simple denotations from Fig. 3

mapping to domain constants, like the mapping in Fig. 3. Our presentation of direct style is different from that of Shan (2007a): we use the ordinary λ-calculus and the denotational semantics, without introducing operational semantics and so-called control operators (although the informed reader will readily recognize these operators in our presentation). We introduce direct style as 'CPS on-demand'.

We start with an observation about CPS denotations:

$$[\![\mathsf{John}]\!] = \lambda k.\, k\ \mathbf{john}$$
$$[\![\mathsf{saw\ Mary}]\!] = \lambda k.\, k\ (\mathbf{see} \cdot \mathbf{mary})$$

In general, the CPS denotation of a quantifier-free term can be built by first determining the denotation according to the non-CPS rules (8), then wrapping $\lambda k.\, k\ (\cdot)$ around the result.

This observation gives us the idea to merge quantifier-free and CPS semantics; see Fig. 6. If denotations are quantifier-free—that is, if their types have no arrows—we use the non-CPS composition rules (8), which constitute the first case in (14) and (15) below. For CPS denotations, we use the CPS composition rules (10), written as the last case in (14) and (15). When composing CPS and non-CPS denotations, we implicitly promote the latter into CPS by wrapping them in $\lambda k.\, k\ (\cdot)$. The two middle cases of (14) and (15) show the result of that promotion after simplification (β-reductions). Thus the composition rules $^{>}$ and $^{<}$ become schematic with four cases. Likewise, $(\![\cdot]\!)$ becomes schematic with two cases, shown in (16). We stress the absence of any nondeterminism: which of the four composition rules to apply is uniquely determined by the types of the denotations being combined.

$$(1) \quad m_1{}^{>}m_2 \stackrel{\text{def}}{=} \begin{cases} m_1 \cdot m_2 & \text{if } m_1 : (\sigma\sigma'), m_2 : \sigma \\ \lambda k.\, m_2(\lambda x.\, k(m_1 \cdot x)) & \text{if } m_1 : (\sigma\sigma'), m_2 : (\sigma \to t) \to t \\ \lambda k.\, m_1(\lambda f.\, k(f \cdot m_2)) & \text{if } m_1 : ((\sigma\sigma') \to t) \to t, m_2 : \sigma \\ \lambda k.\, m_1(\lambda f.\, m_2(\lambda x.\, k(f \cdot x))) & \text{if } m_1 : ((\sigma\sigma') \to t) \to t, \\ & \quad m_2 : (\sigma \to t) \to t \end{cases}$$

$$(2) \quad m_1{}^{<}m_2 \stackrel{\text{def}}{=} \begin{cases} m_2 \cdot m_1 & \text{if } m_1 : \sigma, m_2 : (\sigma\sigma') \\ \lambda k.\, m_2(\lambda f.\, k(f \cdot m_1)) & \text{if } m_1 : \sigma, m_2 : ((\sigma\sigma') \to t) \to t \\ \lambda k.\, m_1(\lambda x.\, k(m_2 \cdot x)) & \text{if } m_1 : (\sigma \to t) \to t, m_2 : (\sigma\sigma') \\ \lambda k.\, m_1(\lambda x.\, m_2(\lambda f.\, k(f \cdot x))) & \text{if } m_1 : (\sigma \to t) \to t, \\ & \quad m_2 : ((\sigma\sigma') \to t) \to t \end{cases}$$

$$(3) \quad (\!| m |\!) \stackrel{\text{def}}{=} \begin{cases} m & \text{if } m : t \\ m(\lambda v.\, v) & \text{if } m : (t \to t) \to t \end{cases}$$

Since the sentence [$_{\mathsf{M}}$ John [$_{\mathsf{VP}}$ saw Mary].] is quantifier-free, its denotation is trivially determined as in Sect. 2.1, with no β-reductions—in marked contrast with Sect. 2.2. For [$_{\mathsf{M}}$ Someone [$_{\mathsf{VP}}$ saw Mary].], we compute $[\![$[$_{\mathsf{VP}}$ saw Mary]$]\!]$ as **see** · **mary** of the type (et) by the simple rules of (8). The denotation of someone has the type $(e \to t) \to t$, which is a CPS type: it has arrows. The types tell us to use the third case of (15) to combine $[\![$someone$]\!]$ with $[\![$[$_{\mathsf{VP}}$ saw Mary]$]\!]$. We obtain the final result $\exists y.\, \mathbf{see} \cdot \mathbf{mary} \cdot y$ after applying the second case of (16).

Direct style thus keeps quantifier-free lexical entries 'unlifted' and removes the tedium of the CPS semantics. Such CPS-on-demand, or *selective* CPS, has been used to implement delimited control in Scala (Rompf et al. 2009).

3 The Nature of Quantification

Before we advance to the main topic, scope and ambiguity, we take a hard look at logical quantification. So far, we have used quantified logical formulas like $\forall x.\, \mathbf{see} \cdot x \cdot \mathbf{john}$ without formally introducing quantifiers. The informality, however attractive, makes it hard to specify how to correctly use a logical quantifier to obtain a well-formed closed formula. For example, QR approaches may produce a denotation with an unbound trace, which must then be somehow fixed or avoided. A proper theory should not let sentence denotations with unbound variables arise in the first place.

We go back to the language $\mathcal{D}$, Fig. 2, and extend it with standard first-order quantifiers. The result is the language $\mathcal{D}_Q$ in Fig. 7.

We added variables, which are natural numbers, and two expression forms $\forall_n d$ and $\exists_n d$ to quantify over the variable n. Their model-theoretical semantics is standard, relying on the variable assignment ϕ, which maps variables to entities. Then $\forall_n d$ is true for the assignment ϕ iff d is true for every assignment that differs from ϕ only in the mapping of the variable n.

Levels	$n, l \in \mathbb{N}$
Base types	$\upsilon ::= e \mid t$
Types	$\sigma ::= \upsilon \mid (\sigma\sigma)$
Annotated types	$\rho ::= \sigma^n$
Constants	$c ::= \wedge \mid \vee \mid \Rightarrow \mid \neg \mid \textbf{john} \mid \textbf{mary} \mid \textbf{see} \mid \ldots$
Variables	n, l
Expressions	$d ::= c \mid d \mathbin{.} d \mid n \mid \forall_n d \mid \exists_n d$

Type system for judgments $d : \rho$

$$\frac{}{n : e^{n+1}} \qquad \frac{d_1 : (\sigma_2\sigma_1)^{n_1} \quad d_2 : \sigma_2^{n_2}}{d_1 \mathbin{.} d_2 : \sigma_1^{\max(n_1,n_2)}} \qquad \frac{d : t^{n+1}}{\forall_n d : t^n} \qquad \frac{d : t^{n+1}}{\exists_n d : t^n}$$

Fig. 7 The language $\mathcal{D}_Q$ of denotations

Figure 7 also extends the type system, with annotated types ρ and judgments $d : \rho$ of d having the annotated type ρ. Expression types σ are annotated with the upper bound on the variable names that may occur in the expression. For example, $d : \sigma^1$ means that d *may* have (several) occurrences of the variable 0; $d : \sigma^2$ means d may contain the variables 0 and 1. Our variables are de Bruijn *levels*. An expression d of the type σ^0 is a closed expression. We will often omit the type annotation (superscript) 0—hence $\mathcal{D}$ can be regarded as the variable-free fragment of $\mathcal{D}_Q$.

The language $\mathcal{L}$ will now use the expressions of $\mathcal{D}_Q$ as constants, and annotated types ρ as base types. Although the semantic composition functions in (14), (15) and (16) remain the same, their typing becomes more precise, as shown in Fig. 8. (Recall $(\!|\cdot|\!)$ is the semantic composition function that corresponds to the clause boundary, which we will discuss in detail in Sect. 5.1.) As usual, the typing rules are schematic: m_1 and m_2 stand for arbitrary expressions of $\mathcal{L}$, σ_1 and σ_2 stand for arbitrary $\mathcal{D}_Q$ types, and n_1, n_2, l_1, l_2, etc. are arbitrary levels. The choice n or l for the name of level metavariables has no significance beyond notational convenience. The English fragments in Figs. 5, 6 remain practically the same; the quantifier words now receive precisely defined rather than informal denotations, and precise semantic types; see Fig. 9.

Figure 9 assigns denotations and types to everyone and someone that are schematic in n. That is, there is an instance of the denotation for each natural number n. One may worry about choosing the right n and possible ambiguities. The worries are unfounded. As we demonstrate below, the requirement that the whole sentence denotation be closed (that is, have the type t^0) uniquely determines the choice of n in the denotation schemas for the quantifier words. The choice of variable names n is hence type-directed and deterministic. As an example, we show the typing derivation for "Someone saw everyone", which we explain below.

$$\frac{m_1 : (\sigma_2\sigma_1)^{n_1} \quad m_2 : \sigma_2^{n_2}}{m_1{}^{>}m_2 : \sigma_1^{\max(n_1,n_2)}} \qquad \frac{m_1 : (\sigma_2\sigma_1)^{n_1} \quad m_2 : (\sigma_2^{n_2} \to t^{l_1}) \to t^{l_2}}{m_1{}^{>}m_2 : (\sigma_1^{\max(n_1,n_2)} \to t^{l_1}) \to t^{l_2}}$$

$$\frac{m_1 : ((\sigma_2\sigma_1)^{n_1} \to t^{l_1}) \to t^{l_2} \quad m_2 : \sigma_2^{n_2}}{m_1{}^{>}m_2 : (\sigma_1^{\max(n_1,n_2)} \to t^{l_1}) \to t^{l_2}}$$

$$\frac{m_1 : ((\sigma_2\sigma_1)^{n_1} \to t^{l_1}) \to t^{l_2} \quad m_2 : (\sigma_2^{n_2} \to t^{l_3}) \to t^{l_1}}{m_1{}^{>}m_2 : (\sigma_1^{\max(n_1,n_2)} \to t^{l_3}) \to t^{l_2}} \qquad \frac{m : t^0}{(\!|m|\!) : t^0} \qquad \frac{m : (t^n \to t^n) \to t^0}{(\!|m|\!) : t^0}$$

Fig. 8 Typing rules for $^{>}$ in (14) ($^{<}$ is analogous) and for $(\!|\cdot|\!)$ in (16)

Syntax	Semantic type	Denotation $[\![\cdot]\!]$
...		
NP → everyone	$(e^{n+1} \to t^{n+1}) \to t^n$	$\lambda k.\, \forall_n (k\ n)$
NP → someone	$(e^{n+1} \to t^{n+1}) \to t^n$	$\lambda k.\, \exists_n (k\ n)$

Fig. 9 Precise denotations of quantifiers and their annotated types. The rest of the fragment remains the same; see Figs. 5 or 6

$$\frac{[\![\mathsf{someone}]\!] : (e^1 \to t^1) \to t^0 \qquad \dfrac{[\![\mathsf{see}]\!] : e(et)^0 \quad [\![\mathsf{everyone}]\!] : (e^2 \to t^2) \to t^1}{\mathbf{see}^{>}(\lambda k.\, \forall_1 (k 1)) : ((et)^2 \to t^2) \to t^1}}{\dfrac{(\lambda k.\, \exists_0 (k 0))^{<}(\mathbf{see}^{>} \lambda k.\, \forall_1 (k 1)) : (t^2 \to t^2) \to t^0}{(|(\lambda k.\, \exists_0 (k 0))^{<}(\mathbf{see}^{>} \lambda k.\, \forall_1 (k 1))|) : t^0}}$$

The resulting denotation β-reduces to $\exists_0 \forall_1 \mathbf{see} \cdot 1 \cdot 0$, as in Sect. 2.2. The other derivations in Sects. 2.2 and 2.3 are made rigorous similarly.

In the derivation above, the schematic denotation $[\![\mathsf{someone}]\!]$ was instantiated with $n = 0$, and the schema $[\![\mathsf{everyone}]\!]$ was instantiated with $n = 1$. It may be unclear how we have made this choice. It is a simple exercise to see that no other choice fits. Relying on the simplicity of the example, we now demonstrate the general method of choosing the variable names n appearing in schematic denotations. We repeat the derivation, this time assuming that $[\![\mathsf{someone}]\!]$ is instantiated with some variable name n and $[\![\mathsf{everyone}]\!]$ is instantiated with some name l. These so-called *schematic* or logical meta-variables n and l stand for some natural numbers that we do not know yet. As we build the derivation and fit the denotations, we discover constraints on n and l, which in the end let us determine these numbers.

$$\frac{[\![\mathsf{someone}]\!] : (e^{n+1} \to t^{n+1}) \to t^n \qquad \dfrac{[\![\mathsf{see}]\!]{:}e(et)^0 \quad [\![\mathsf{everyone}]\!]{:}(e^{l+1} \to t^{l+1}) \to t^l}{\mathbf{see}^{>}(\lambda k.\, \forall_l (k\ l)){:}((et)^{l+1} \to t^{l+1}) \to t^l}}{\dfrac{(\lambda k.\, \exists_n (k\ n))^{<}(\mathbf{see}^{>} \lambda k.\, \forall_l (k\ l)){:}(t^{\max(n+1,l+1)} \to t^{l+1}) \to t^n \quad \text{where } n{+}1{=}l}{\left(\!\left|(\lambda k.\, \exists_n (k\ n))^{<}(\mathbf{see}^{>} \lambda k.\, \forall_l (k\ l))\right|\!\right){:}t^0 \quad \text{where } n{=}0,\ \max(n{+}1,l{+}1){=}l{+}1}}$$

In the last-but-one step of the derivation, we attempt to type $(\lambda k.\, \exists_n (k\ n))^{<}$ $(\mathbf{see}^{>} \lambda k.\, \forall_l (k\ l))$ using the rule

$$\frac{m_1 : (\sigma_2^{n_1} \to t^{l_1}) \to t^{l_2} \quad m_2 : ((\sigma_2\sigma_1)^{n_2} \to t^{l_3}) \to t^{l_1}}{m_1{}^{<}m_2 : (\sigma_1^{\max(n_1,n_2)} \to t^{l_3}) \to t^{l_2}}.$$

This attempt only works if $n + 1 = l$, because according to the rule, the types of m_1 and m_2 must share the same name l_1. In the last step of the derivation, applying the typing rule for $(\!|\cdot|\!)$ from Fig. 8 gives two other constraints: $n = 0$ and $\max(n + 1, l + 1) = l + 1$. The three constraints have a unique solution: $n = 0$, $l = 1$.

More complex sentences with more quantifiers require us to deal with more variable names n_1, n_2, n_3, etc., and more constraints on them. The overall principle remains straightforward: since typing is syntax-directed there is never a puzzle as to which typing rule to use at any stage of the derivation. At most one typing rule applies. An application of a typing rule generally imposes constraints on the levels. We collect all constraints and solve them at the end (some constraints can be solved as we go).

Accumulating and solving such constraints is a logic programming problem. Luckily, in modern functional and logic programming languages like Haskell, Twelf or Agda, type checking propagates and solves constraints in a very similar way. If we write our denotations in, say, Haskell, the Haskell type checker automatically determines the names of schematic meta-variables and resolves schematic denotations and rules. We have indeed used the Haskell interpreter GHCi as such a 'semantic calculator', which infers types, builds derivations and instantiates schemas. Like the Penn Lambda Calculator (Champollion et al. 2007), the Haskell interpreter also reduces terms. We can enter any syntactic derivation at the interpreter prompt and see its inferred type and its normal-form denotation.

The choice of variable names, dictated by the requirement that sentence denotations be closed, in turn describes quantifier scopes, as we shall see next.

4 The Inverse-Scope Problem

If we compute the denotation of $[_{\mathsf{M}}$ Someone VP.] by the rules of Sect. 2.2, we obtain

(17) $$[\![\text{Someone VP.}]\!] = (\lambda k_0.\,(\lambda k.\,\exists y.ky)(\lambda x.\,[\![\text{VP}]\!]\,(\lambda f.\,k_0(f \cdot x))))\,(\lambda v.\,v) \rightsquigarrow \exists y.\,[\![\text{VP}]\!]\,(\lambda f.\,(f \cdot y))$$

No matter what VP is, the existential always scopes over it. Thus, we invariably get the linear-scope reading for the sentence. Obtaining the inverse-scope reading is the problem. One suggested solution (Barker 2002; de Groote 2001) is to introduce nondeterminism into semantic composition rules. We do not find that approach attractive because of over-generation: we may end up with a great number of denotations, not all of which correspond to available readings. Explaining different scope-taking abilities of existentials and universals (see Sect. 5) also becomes very difficult.

Our solution to inverse scope is the *continuation hierarchy* (Danvy and Filinski 1990). Like Russian dolls, contexts nest. Plugging a term into a context gives a bigger term, which can be plugged into another, wider context, and so on. This hierarchy of contexts is reflected in the continuation hierarchy. Quantifiers gain access not only to their immediate context but also to a higher-up context, and may hence quantify over outer contexts. We build the hierarchy from the CPS denotations of Sect. 2.2, to be called CPS^1 denotations (with the annotated types of Sect. 3). We introduce the corresponding direct style of the hierarchy in Sect. 4.2.

Before we begin, let us quickly skip ahead and peek at the final result, to see the difference that the continuation hierarchy makes. Equation (17) will look somewhat like

(17a)
$$\begin{aligned}[\![\mathsf{Someone\ VP.}]\!] = (\lambda k_0.\, (\lambda k_1.\, \lambda k_2.\, k_1\ y\ (\lambda v.\, k_2(\exists y.v)))\\ (\lambda x.\, [\![\mathsf{VP}]\!]\ (\lambda f.\, k_0\ (f \cdot x))))\\ (\lambda v.\, \lambda k_2.\, k_2\ v)(\lambda v.\, v)\\ \rightsquigarrow [\![\mathsf{VP}]\!]\ (\lambda f.\, \lambda k_2.\, k_2(f \cdot y))(\lambda v.\, (\exists y.v))\end{aligned}$$

(see Eq. (25) for the complete example). VP will now have a chance to introduce a quantifier to scope over $\exists y.\cdot$.

We build the hierarchy by iterating the CPS transformation. An expression may be re-written in CPS multiple times. Each re-writing adds another continuation representing a higher (outer) context (Danvy and Filinski 1990). Let us take an example. A term **john** written in CPS takes the continuation argument representing the term's context, and plugs itself into that context: $\lambda k.\, k$ **john**. Mechanically applying to it the rules of transforming terms into CPS (Danvy and Filinski 1990) gives $\lambda k_1.\, \lambda k_2.\, (k_1$ **john**$)\ k_2$. This CPS^2 term receives two continuations and plugs **john** into the inner one, obtaining the CPS^1 term k_1 **john** that computes the result to be plugged into the outer context k_2. We may diagram the CPS^1 term $\lambda k_1.\, k_1$ **john** as $[_{k_1} \ldots$ [**john**] $\ldots]$, that is, **john** filling in the hole in a context represented by k_1. Likewise we diagram the CPS^2 term $\lambda k_1.\, \lambda k_2.\, (k_1$ **john**$)\ k_2$ as $[_{k_2} \ldots [_{k_1} \ldots$ [**john**] $\ldots] \ldots]$. In the CPS^2 case, if k_2 represents the outer context, the application $k_2\ e$ represents plugging e into that context. If k_1 is an inner context, $k_1\ e\ k_2$ corresponds to plugging e into it and the result into an outer context k_2. We shall see soon that types make it clear which context, outer or inner, a continuation represents and what needs to be plugged into what.

The CPS^2 term $\lambda k_1.\, \lambda k_2.\, k_1$ **john** k_2 is however extensionally equivalent to the CPS^1 term $\lambda k.\, k$ **john** we started with. In general, if a term uses its continuation 'trivially',[4] further CPS transformations leave the term intact. Thus, after quantifier-free lexical entries are converted once into CPS, they can be used as they are at any level of the CPS hierarchy.

[4] We say that a term uses its continuation argument k trivially if k is used exactly once in the term, and each application in the term is the entire body of a λ-abstraction.

Although the CPS^2 term of **john** is same as the CPS^1 term, the types differ. The CPS^1 type is $(e \to t^n) \to t^n$, telling us that **john** receives a context to be plugged with a term of the type e giving a term of the type t^n. The CPS^2-term receives another continuation k_2, representing the outer context $t^n \to t^{l_1}$. Thus the type of $\lambda k_1.\, \lambda k_2.\, k_1$ **john** k_2 is $(e \to ((t^n \to t^{l_1}) \to t^{l_2})) \to ((t^n \to t^{l_1}) \to t^{l_2})$. This type is schematic, written with schematic meta-variables n, l_1 and l_2 standing for some variable names to be determined when building a derivation, as described in Sect. 3.

In general, types in the CPS hierarchy have a regular structure and can be described uniformly. The key observation is recurrence of the pattern $(t^n \to t^{l_1}) \to t^{l_2}$ that can be represented by its sequence of annotations n, l_1, l_2. Therefore, we introduce the notation

(18)
$$\begin{aligned} \{n\} &= t^n \\ \{nl_1l_2\} &= (t^n \to \{l_1\}) \to \{l_2\} \\ \{nl_1l_2l_3l_4l_5l_6\} &= (t^n \to \{l_1l_2l_3\}) \to \{l_4l_5l_6\} \\ &\vdots \end{aligned}$$

where all ns and ls are schematic meta-variables. Since these sequences can become very long, we use Greek letters α, β, γ to each stand for a schematic *sequence* of variable names. All occurrences of the same Greek letter bearing the same superscripts and subscripts refer to the same sequence. We will state the length of the sequence separately or leave it implicit in the CPS level under discussion. Thus the type of $\lambda k.\, k$ **john** for any CPS level has the form $(e \to \{\alpha\}) \to \{\alpha\}$. Juxtaposed Greek letters and schematic variables signify concatenated sequences. For example, (18) is compactly written as follows.

(19)
$$\begin{aligned} \{n\} &= t^n \\ \{n\alpha\beta\} &= (t^n \to \{\alpha\}) \to \{\beta\} \end{aligned}$$

4.1 CPS-Hierarchy Semantics

The CPS^2 semantics for our language fragment is shown in Fig. 10. Except for the quantifiers, the figure looks like the ordinary CPS semantics, Fig. 5, with the wholesale replacement of the type t by $\{\alpha\}$. The interesting part is quantifier words. There are now two sets of them, indexed with 1 and 2: the quantifier words become polysemous, with two possible denotations. Postulating the polysemy of quantifiers is similar to generalizing the conjunction schema (Partee and Rooth 1983), or assuming the free indexing in LF.

The quantifiers $\mathsf{everyone}_1$ and $\mathsf{someone}_1$ are the quantifiers from Sect. 2.2, whose denotations are re-written in CPS. For example, the denotation of $\mathsf{everyone}$ from Fig. 9 (which is the precise version of that from Fig. 5) is $\lambda k.\, \forall_n (k\, n)$; re-writing it in CPS gives $\lambda k_1.\, \lambda k_2.\, k_1\; n\; (\lambda v.\, k_2(\forall_n v))$. It plugs the variable n into the (inner) context k_1, then plugs the result into $\forall_n[]$ and finally into the outer context k_2. Thus, $\mathsf{everyone}_1$ quantifies over the immediate, inner context k_1, as in Sect. 2.2 above.

Syntax	Semantic type	Denotation $[\![\cdot]\!]$
M $\to$ S .	t^0	$(\![[\![\mathsf{S}]\!]]\!)$
S $\to$ NP VP	$(t^n \to \{\alpha\}) \to \{\beta\}$	$[\![\mathsf{NP}]\!]\ ^{<}\ [\![\mathsf{VP}]\!]$
VP $\to$ Vt NP	$((et)^n \to \{\alpha\}) \to \{\beta\}$	$[\![\mathsf{Vt}]\!]\ ^{>}\ [\![\mathsf{NP}]\!]$
NP $\to$ John	$(e \to \{\alpha\}) \to \{\alpha\}$	$\lambda k.\, k\ \mathbf{john}$
NP $\to$ Mary	$(e \to \{\alpha\}) \to \{\alpha\}$	$\lambda k.\, k\ \mathbf{mary}$
VP $\to$ left	$((et) \to \{\alpha\}) \to \{\alpha\}$	$\lambda k.\, k\ \mathbf{leave}$
Vt $\to$ saw	$((e(et)) \to \{\alpha\}) \to \{\alpha\}$	$\lambda k.\, k\ \mathbf{see}$
NP $\to$ $\mathsf{everyone}_1$	$(e^{n+1} \to \{(n+1)\gamma\}) \to \{n\gamma\}$	$\lambda k_1.\, \lambda k_2.\, k_1\ n\ (\lambda v.\, k_2(\forall_n v))$
NP $\to$ $\mathsf{someone}_1$	$(e^{n+1} \to \{(n+1)\gamma\}) \to \{n\gamma\}$	$\lambda k_1.\, \lambda k_2.\, k_1\ n\ (\lambda v.\, k_2(\exists_n v))$
NP $\to$ $\mathsf{everyone}_2$	$(e^{n+1} \to \{\gamma(n+1)\}) \to \{\gamma n\}$	$\lambda k_1.\, \lambda k_2.\, \forall_n (k_1\ n\ k_2)$
NP $\to$ $\mathsf{someone}_2$	$(e^{n+1} \to \{\gamma(n+1)\}) \to \{\gamma n\}$	$\lambda k_1.\, \lambda k_2.\, \exists_n (k_1\ n\ k_2)$

Fig. 10 Syntax and the CPS^2 semantics for the small fragment. α and β are sequences of schematic meta-variables of length 3, and γ is a sequence of length 2. See the text for expressions and types of the semantic composition operators $^{>}$, $^{<}$ and $(\!|\cdot|\!)$

The continuation arguments to $\mathsf{everyone}_1$ are used trivially, so the denotation can be used as it is not only for CPS^2 but also for CPS^3 and at higher levels.

The second set of quantifiers quantify over the outer context, as their denotation says. For example, $\lambda k_1.\, \lambda k_2.\, \forall_n (k_1\ n\ k_2)$ plugs the variable n into the inner context k_1, plugs the result into k_2 and quantifies over the final result. The inner and the outer contexts are uniquely determined, as shall see shortly.

The semantic combinators $^{>}$ and $^{<}$ in (10) use their continuation argument trivially; therefore, they also work for CPS^2 and for all other levels of the hierarchy. We need to give them more general schematic types, extending Fig. 8 so it works at any level of the hierarchy:

(20) $$\frac{m_1 : (\sigma_2^{n_1} \to \{\alpha\}) \to \{\beta\} \quad m_2 : ((\sigma_2\sigma_1)^{n_2} \to \{\gamma\}) \to \{\alpha\}}{m_1\, ^{<} m_2 : (\sigma_1^{\max(n_1,n_2)} \to \{\gamma\}) \to \{\beta\}}$$

(21) $$\frac{m_1 : ((\sigma_2\sigma_1)^{n_1} \to \{\alpha\}) \to \{\beta\} \quad m_2 : (\sigma_2^{n_2} \to \{\gamma\}) \to \{\alpha\}}{m_1\, ^{>} m_2 : (\sigma_1^{\max(n_1,n_2)} \to \{\gamma\}) \to \{\beta\}}$$

We only need to change $(\!|\cdot|\!)$ to account for the two continuation arguments, and hence, two initial continuations:

(22) $$(\!|m|\!) \stackrel{\text{def}}{=} m(\lambda v.\, \lambda k_2.\, k_2\ v)(\lambda v.\, v)$$

The initial CPS^1 continuation $(\lambda v.\, \lambda k_2.\, k_2\ v)$ plugs its argument into the outer context; the initial outer context is the empty context. Schematically, $(\!|m|\!)$ may be diagrammed as $[_{k_2}[_{k_1} m]]$.

The two sets of quantifiers, level-1 and level-2, treat the inner and outer contexts differently. The remainder of this subsection presents several examples of computing denotations of sample sentences by using the lexical entries and the composition rules of Fig. 10 and performing simplifications by β-reductions. As we shall see, the sequence of reductions for, say, Someone$_1$ VPcan be diagrammed at a high level as follows:

(23)
$$\begin{aligned} & [\![\text{Someone}_1\ \text{VP.}]\!] \\ &\quad = (\![\, [\![\text{Someone}_1\ \text{VP}]\!] \,]\!) \\ &\quad = [_{k_2}[_{k_1}[\![\text{Someone}_1\ \text{VP}]\!]]] \\ &\quad \rightsquigarrow [_{k_2}\exists_n[_{k_1}n^{<}[\![\text{VP}]\!]]] \end{aligned}$$

We hence see that it is the level-1 quantifiers that wedge themselves between the inner context k_1 and the outer context k_2. We also see that, if the VP contains only level-1 QNPs, they would quantify over $[_{k_1}n^{<}\ldots]$ giving the linear-scope reading. On the other hand, if the VP has a level-2 QNP, it will quantify over the outer context $[_{k_2}\exists_n[_{k_1}n^{<}\ldots]]$ yielding the inverse-scope reading. After this preview, we describe the computation of denotations in detail.

It is a simple exercise to show that [$_\text{M}$ Someone$_1$ [$_\text{VP}$ saw everyone$_1$].] has the same linear-scope reading $\exists_0\forall_1\mathbf{see}\cdot 1\cdot 0$ as computed with the ordinary CPS, Sect. 2.2—with essentially the same β-reductions shown in that section. It is also easy to see that [$_\text{M}$ Someone$_2$ [$_\text{VP}$ saw everyone$_2$].] also has exactly the same denotation. The interesting cases are the sentences with different levels of quantifiers. For example,

(4)
$$\begin{aligned} & [\![[_\text{M}\ [_\text{S}\ [_\text{NP}\ \text{Someone}_2]\ [_\text{VP}\ [_\text{Vt}\ \text{saw}]\ [_\text{NP}\ \text{everyone}_1]]].]]\!] \\ &\quad = (\lambda k_0.\,(\lambda k_1.\,\lambda k_2.\,\exists_0(k_1\ 0\ k_2))(\lambda x. \\ &\qquad\quad (\lambda k_3.\,(\lambda k.\,k\ \mathbf{see})(\lambda f'.\,(\lambda k_1.\,\lambda k_2.\,k_1\ 1\ (\lambda v.\,k_2(\forall_1 v)))(\lambda x'.\,k_3\ (f'\cdot x')))) \\ &\qquad\quad (\lambda f.\,k_0\ (f\cdot x)))) \\ &\qquad (\lambda v.\,\lambda k_2.\,k_2\ v)(\lambda v.\,v) \\ &\quad \rightsquigarrow (\lambda k_1.\,\lambda k_2.\,\exists_0(k_1\ 0\ k_2))(\lambda x. \\ &\qquad\quad (\lambda k_2.\,(\lambda v.\,k_2(\forall_1 v))(\mathbf{see}\cdot 1\cdot x))) \\ &\qquad (\lambda v.\,v) \\ &\quad \rightsquigarrow (\lambda k_1.\,\lambda k_2.\,\exists_0(k_1\ 0\ k_2))(\lambda x. \\ &\qquad\quad (\lambda k_2.\,k_2(\forall_1(\mathbf{see}\cdot 1\cdot x)))) \\ &\qquad (\lambda v.\,v) \\ &\quad \rightsquigarrow (\lambda k_2.\,\exists_0(\forall_1(\mathbf{see}\cdot 1\cdot 0)))(\lambda v.\,v) \\ &\quad \rightsquigarrow \exists_0(\forall_1(\mathbf{see}\cdot 1\cdot 0)) \end{aligned}$$

The result still shows the linear-scope reading, because someone$_2$ quantifies over the wide context and so wins over the narrow-context quantifier everyone$_1$. One may wonder how we chose the names of the quantified variables: 0 for someone$_2$ and 1 for everyone$_1$. The choice is clear from the final denotation: since it should have the type t^0 (that is, be closed), the schema for the corresponding someone$_2$ must have been instantiated with $n = 0$. Therefore, $\forall_1(\mathbf{see}\cdot 1\cdot 0)$ must have the type t^1, which determines the schema instantiation for everyone$_1$. One may say that 'names follow scope'. The variable names can also be chosen before β-reducing, while building the typing derivation, as demonstrated in Sect. 3.

We now make a different choice of lexical entries for the same quantifier words in the running example:

(5) $[\![[_{\mathsf{M}}\ [_{\mathsf{S}}\ [_{\mathsf{NP}}\ \mathsf{Someone}_1]\ [_{\mathsf{VP}}\ [_{\mathsf{Vt}}\ \mathsf{saw}]\ [_{\mathsf{NP}}\ \mathsf{everyone}_2]]].]]\!]$

$$\begin{aligned}
&= (\lambda k_0.\, (\lambda k_1.\, \lambda k_2.\, k_1\ 1\ (\lambda v.\, k_2(\exists_1 v)))(\lambda x.\\
&\qquad (\lambda k_3.\, (\lambda k.\, k\ \mathbf{see})(\lambda f'.\, (\lambda k_1.\, \lambda k_2.\, \forall_0(k_1\ 0\ k_2))(\lambda x'.\, k_3\ (f' \cdot x'))))\\
&\qquad (\lambda f.\, k_0\ (f \cdot x))))\\
&\quad (\lambda v.\, \lambda k_2.\, k_2\ v)(\lambda v.\, v)\\
&\rightsquigarrow (\lambda k_1.\, \lambda k_2.\, k_1\ 1\ (\lambda v.\, k_2(\exists_1 v)))(\lambda x.\\
&\qquad (\lambda k_2.\, \forall_0(k_2\ (\mathbf{see} \cdot 0 \cdot x))))\\
&\quad (\lambda v.\, v)\\
&\rightsquigarrow (\lambda k_2.\, (\lambda k_2.\, \forall_0(k_2\ (\mathbf{see} \cdot 0 \cdot 1)))(\lambda v.\, k_2(\exists_1 v)))\\
&\quad (\lambda v.\, v)\\
&\rightsquigarrow (\lambda k_2.\, \forall_0((\lambda v.\, k_2(\exists_1 v))(\mathbf{see} \cdot 0 \cdot 1)))\\
&\quad (\lambda v.\, v)\\
&\rightsquigarrow \forall_0(\exists_1(\mathbf{see} \cdot 0 \cdot 1))
\end{aligned}$$

We obtain the inverse-scope reading: $\mathsf{everyone}_2$ quantified over the higher, or wider, context and hence outscoped $\mathsf{someone}_1$. This outscoping is noticeable already in the result of the first set of β-reductions, which may be diagrammed as $\forall_0[_{k_2}\exists_1[_{k_1}\mathbf{see} \cdot 0 \cdot [1]]]$. Since the universal quantifier eventually got the widest scope, the schema for $\mathsf{everyone}_2$ must have been instantiated with $n = 0$. Again, the choice of quantifier variable names is determined by quantifiers' scope.

Thus the continuation hierarchy lets us derive both linear- and inverse-scope readings of ambiguous sentences. The source of the quantifier ambiguity is squarely in the lexical entries for the quantifier words rather than in the rules of syntactic formation or semantic composition.

4.2 Continuation Hierarchy in Direct Style

Like the ordinary CPS, the CPS hierarchy can also be built on demand. Therefore, we do not have to decide in advance the highest CPS level for our denotations, and be forced to rebuild our fragment's denotations should a new example call for yet a higher level. Rather, we build sentence denotations by combining parts with different CPS levels, or even not in CPS. The primitive parts, lexical entry denotations, may remain not in CPS (which is the case for all quantifier-free entries) or at the minimum needed CPS level, regardless of the level of other entries. The incremental construction of hierarchical CPS denotations—building up levels only as required—makes our fragment modular and easy to extend. It also relieves us from the tedium of dealing with unnecessarily high-level CPS terms.

	Syntax	Semantic type	Denotation $[\![\cdot]\!]$
M	→ S .	t^0	$(\!\vert[\![\mathsf{S}]\!]\vert\!)$
S	→ NP VP	t^n or $(t^n \to \{\alpha\}) \to \{\beta\}$	$[\![\mathsf{NP}]\!]\ ^{<}\ [\![\mathsf{VP}]\!]$
VP	→ Vt NP	et^n or $((et)^n \to \{\alpha\}) \to \{\beta\}$	$[\![\mathsf{Vt}]\!]\ ^{>}\ [\![\mathsf{NP}]\!]$
NP	→ John	e	**john**
NP	→ Mary	e	**mary**
VP	→ left	et	**leave**
Vt	→ saw	$e(et)$	**see**
NP	→ everyone$_1$	$(e^{n+1} \to \{(n+1)\alpha\}) \to \{n\alpha\}$	$\lambda k_1.\, \lambda k_2.\, k_1\ n\ (\lambda v.\, k_2(\forall_n v))$
NP	→ someone$_1$	$(e^{n+1} \to \{(n+1)\alpha\}) \to \{n\alpha\}$	$\lambda k_1.\, \lambda k_2.\, k_1\ n\ (\lambda v.\, k_2(\exists_n v))$
NP	→ everyone$_2$	$(e^{n+1} \to \{\beta(n+1)\gamma\}) \to \{\beta n\gamma\}$	$\uparrow [\![\mathsf{everyone}_1]\!]$
NP	→ someone$_2$	$(e^{n+1} \to \{\beta(n+1)\gamma\}) \to \{\beta n\gamma\}$	$\uparrow [\![\mathsf{someone}_1]\!]$

Fig. 11 Syntax and the multi-level direct-style continuation semantics for the small fragment: the merger of Figs. 3, 10. Lexical entries other than the quantifiers keep the simple denotations from Fig. 3. Here α, β and γ are sequences of schematic meta-variables whose length is determined by the CPS level; β is two longer than γ

Luckily, the semantic combinators $^<$ and $^>$ capable of combining the denotations of different CPS levels have already been defined. They are (14) and (15) in Sect. 2.3. The luck comes from the fact that the composition of CPS^1 denotations uses its continuation argument trivially, and therefore, works at any level of the CPS hierarchy. We only need to extend the schema for $(\!|\cdot|\!)$, in a regular way:

(6)
$$(\!|m|\!) \stackrel{\text{def}}{=} \begin{cases} m & \text{if } m : \{0\} \\ m(\lambda v.\, v) & \text{if } m : \{nn0\} \\ m(\lambda v.\, \lambda k.\, kv)(\lambda v.\, v) & \text{if } m : \{nnl_1l_1l_2l_20\} \\ \ldots \end{cases}$$

Applying the schematic definition (26) requires a bit of explanation. If the term m has the type with no arrows, we should compute $(\!|m|\!)$ according to the first case, which requires m be of the type t^0. If m has the type that matches $\{nn0\}$, that is, $(t^n \to t^n) \to t^0$ for some n, we should use the second case, and so on. A term like $\lambda k.\, k$ (**leave** $\cdot$ **john**) of the schematic type $\{0\alpha\alpha\}$ may seem confusing: its type matches $\{nn0\}$ (with α instantiated to $\{0\}$ and n to 0) as well as the type $\{nnl_1l_1l_2l_20\}$ (with $\alpha = \{000\}$ and $n = l_1 = l_2 = 0$) and all further CPS types. We can compute $(\!|\lambda k.\, k$ (**leave** $\cdot$ **john**)$|\!)$ according to the second or any following case. The ambiguity is spurious however: whichever of the applicable equations we use, the result is the same—which follows from the fact that a CPS^i term which uses its continuation argument trivially is a CPS$^{i'}$ term for all $i' \geq i$ Danvy and Filinski (1990). As a practical matter, choosing the lowest-level instance of the schema (26) produces the cleanest derivation.

Figure 11 shows our new fragment.

The quantifier-free lexical entries have the simplest denotations and can be combined with CPS^n terms, $n \geq 0$. The quantifiers everyone$_1$ and someone$_1$ have the schematic denotations that can be used at the CPS^n level $n \geq 1$. The higher-level quantifiers are systematically produced by applying the $\uparrow$ combinator of the type $((e^{n+1} \to \{\alpha\}) \to \{\beta\}) \to ((e^{n+1} \to \{\gamma\alpha\}) \to \{\gamma\beta\})$ (where α and β have the same length and γ is one longer).

(27) $\uparrow m \stackrel{\text{def}}{=} \lambda k.\, \lambda k'.\, m(\lambda v.\, k\ v\ k')$

With the entries in Fig. (11), all sample derivations from Sect. 4 can be repeated in direct style with hardly any changes.

Our direct-style multi-level continuation semantics is essentially the same as that presented in Shan (2004). We do not account for directionality in semantic types (since we use CFG or potentially CCG, rather than type-logical grammars) but we do account for the levels of quantified variables in types (whereas in Shan (2004), quantification was handled informally).

We have thus shown that the CPS hierarchy just as the ordinary CPS can be built on demand, without committing ourselves to any particular hierarchy level but raising the level if needed as a denotation is being composed. The result is the modular semantics, and much simpler and more lucid semantic derivations. From now on, we will use this multi-level direct style.

5 Scope Islands and Quantifier Strength

We have used the continuation hierarchy to explain quantifier ambiguity between linear- and inverse-scope readings. We contend that the ambiguity arises because quantifier words are polysemous: they have multiple denotations corresponding to different levels of the CPS hierarchy. The higher the CPS level, the wider the quantifier scope.

We turn to two further problems. First, just quantifiers' competing with each other on their strength (CPS level) does not explain all empirical data. Some syntactic constructions such as embedded clauses come into play and restrict the scope of embedded quantifiers. That restriction however does not seem to spread to indefinites: "the varying scope of indefinites is neither an illusion nor a semantic epiphenomenon: it needs to be 'assigned' in some way" (Szabolcsi 2000). We shall use the CPS hierarchy to account for scope islands and to assign the varying scope to indefinites.

5.1 Scope Islands

Like our running example "Someone saw everyone", two characteristic examples (4) and (5), repeated below, also have two quantifier words.

Syntax			Semantic type	Denotation $[\![\cdot]\!]$
VP	$\to$	Vs that S	et^n or $((et)^n \to \{\alpha\}) \to \{\beta\}$	$[\![\mathsf{Vs}]\!]^{>} (\![[\![\mathsf{S}]\!]]\!)$
NP	$\to$	that S	e	$\mathbf{That}\,.\,(\![[\![\mathsf{S}]\!]]\!)$
NP	$\to$	Det N	e^n or $(e^n \to \{\alpha\}) \to \{\beta\}$	$[\![\mathsf{Det}]\!]\ [\![\mathsf{N}]\!]$
N	$\to$	teacher	et	**teacher**
N	$\to$	boy	et	**boy**
VP	$\to$	disappeared	et	**disappear**
Vt	$\to$	upset	$e(et)$	**upset**
Vs	$\to$	report	$t(et)$	**report**
Det	$\to$	every$_1$	$(et) \to (e^{n+1} \to \{(n+1)\alpha\}) \to \{n\alpha\}$	$\lambda z.\, \lambda k_1.\, \lambda k_2.\, k_1\ n\ (\lambda v.\ k_2\ (\forall_n (z\,.\,n \Rightarrow v)))$
Det	$\to$	some$_1$, a$_1$	$(et) \to (e^{n+1} \to \{(n+1)\alpha\}) \to \{n\alpha\}$	$\lambda z.\, \lambda k_1.\, \lambda k_2.\, k_1\ n\ (\lambda v.\ k_2\ (\exists_n (z\,.\,n \wedge v)))$
Det	$\to$	every$_2$	$(et) \to (e^{n+1} \to \{\beta(n+1)\gamma\}) \to \{\beta n\gamma\}$	$\lambda z.\, \uparrow ([\![\mathsf{every}_1]\!]\ z)$
Det	$\to$	some$_2$, a$_2$	$(et) \to (e^{n+1} \to \{\beta(n+1)\gamma\}) \to \{\beta n\gamma\}$	$\lambda z.\, \uparrow ([\![\mathsf{some}_1]\!]\ z)$

Fig. 12 Syntax and the multi-level direct-style continuation semantics for the additional fragment

(28) That every boy left upset a teacher.

(29) Someone reported that John saw everyone.

These examples are not ambiguous however: (28) (the same as (4)) has only the inverse-scope reading, whereas (29) (the same as (5)) has only the linear-scope reading. The common explanation (see survey Szabolcsi (2000)) is that embedded tensed clauses are *scope islands*, preventing embedded quantifiers from taking scope wider than the island.

To analyze these examples, we at least have to extend our fragment with more lexical entries and with syntactic forms for clausal NPs, with the corresponding semantic combinators.[5] Figure 12 shows the additions. Most of them are straightforward. In particular, we generalize quantifying NPs like everyone to quantifying determiners like every. The determiner receives an extra (et) argument for its restrictor property, of the type of the denotation of a common noun.[6] Unlike Barker (2002), we do not use choice functions in the denotations for the quantifier determiners. Instead, the denotation of the NP is obtained from the denotations of the Det and N by ordinary function application.

Just as quantifying NPs are polysemous, so are quantifying Dets on our analysis: there are weak (or level-1) forms every$_1$ and a$_1$ and strong (or level-2) forms every$_2$ and a$_2$. Stronger quantifiers outscope weaker ones. For example,

[5] If the domain of the semantic type t only contains the two truth values, we clearly cannot give an adequate denotation to embedded clauses: the domain is too small. Therefore, we now take the domain of t to be a suitable complete Boolean algebra.

[6] This is a simplification: generally speaking, the argument of a Det is not a bare common noun but a noun modified by PP and other adjuncts. Until we add PP to our fragment in Sect. 5.3, the simplification is adequate.

$[_{\mathsf{M}}\ [_{\mathsf{S}}\ [\mathrm{a}_1\ \mathsf{boy}]\ [\mathsf{upset}\ [\mathsf{every}_2\ \mathsf{teacher}]]].]$ determines the inverse-scope reading $\forall_0(\mathbf{teacher}\cdot 0 \implies \exists_1(\mathbf{boy}\cdot 1 \wedge \mathbf{upset}\cdot 0\cdot 1))$.

Recall from Fig. 11 how the matrix denotation $\mathsf{M} \to \mathsf{S}.$ is obtained from the denotation of the main clause: $[[\mathsf{M}]] = (\!|[[\mathsf{S}]]|\!)$. We see exactly the same pattern for the clausal NPs in the semantic operations corresponding to $\mathsf{Vs\ that\ S}$ and $\mathsf{that\ S}$: in all the cases, the denotation of a clause is enclosed within $(\!|\cdot|\!)$, which is the semantic counterpart of the syntactic clause boundary. The typing rules for $(\!|\cdot|\!)$ in Fig. 8 specify its result have the type t^0, as befits the denotation of a clause. The type t^0 is not a CPS type and hence $(\!|[[\mathsf{S}]]|\!)$ cannot get hold of its context to quantify over. Therefore, if S had any embedded quantifiers, they can quantify only as far as the clause. The operation $(\!|\cdot|\!)$ thus acts as the scope delimiter, delimiting the context over which quantification is possible. (Incidentally, the same typing rules of $(\!|\cdot|\!)$ severely restrict how this scope-delimiting operation may be used within lexical entries. For example, $(\!|[[\mathsf{VP}]]|\!)$ is ill-typed since VP does not have the type t^n or $(t^n \to \{\alpha\}) \to \{\beta\}$.)

In case of (28), we obtain the same denotation (30) no matter which lexical entry we choose for the embedded determiner, every_1 (31) or every_2 (32). The quantifier remains trapped in the clause and the sentence is not ambiguous. Incidentally, since all quantifier variables used within a clause will be quantified within the clause, their names can be chosen regardless of the names of other variables within the sentence. That's why the name 0 is reused in (30). Again, names follow scope. A similar analysis applies to (29).

(30) $\exists_0(\mathbf{teacher}\cdot 0 \wedge \mathbf{upset}\cdot 0\cdot(\mathbf{That}\cdot\forall_0(\mathbf{boy}\cdot 0 \implies \mathbf{leave}\cdot 0)))$

(31) $[[[_{\mathsf{M}}\ [_{\mathsf{NP}}\ \mathsf{That}\ [_{\mathsf{S}}\ \mathsf{every}_1\ \mathsf{boy\ left}]]\ [_{\mathsf{VP}}\ \mathsf{upset}\ [_{\mathsf{NP}}\ \mathsf{a}_1\ \mathsf{teacher}]].]]]$

(32) $[[[_{\mathsf{M}}\ [_{\mathsf{NP}}\ \mathsf{That}\ [_{\mathsf{S}}\ \mathsf{every}_2\ \mathsf{boy\ left}]]\ [_{\mathsf{VP}}\ \mathsf{upset}\ [_{\mathsf{NP}}\ \mathsf{a}_1\ \mathsf{teacher}]].]]]$

We have demonstrated that a scope island is an effect of the operation $(\!|\cdot|\!)$, which is the semantic counterpart of the syntactic clause boundary. In our analysis, each surface syntactic constituent still corresponds to a well-formed denotation, and each surface syntactic formation rule still corresponds to a semantic combinator. Our approach hence is directly compositional.

5.2 Wide-Scope Indefinites

Given that enclosing all clause denotations in $(\!|\cdot|\!)$ traps all quantifiers inside, how do indefinites manage to get out? And they do get out: "Indefinites acquire their existential scope in a manner that does not involve movement and is essentially syntactically unconstrained" (Szabolcsi 2000, Sect. 3.2.1). For example:

(33) Everyone reported that [Max and some lady] disappeared.
(34) Most guests will be offended [if we don't invite some philosopher].
(35) All students believe anything [that many teachers say].

Szabolcsi argued (Szabolcsi 2000) that all these examples are ambiguous. In particular, in (33) (the same as (7)), either different people meant a different lady disappearing along with Max, or there is one lady that everyone reported as disappearing along with Max. Interestingly, the example

(36) Someone reported that [Max and every lady] disappeared.

is not ambiguous: there is a single reporter of the disappearance for Max and all ladies. The unambiguity of (36) is explained by the embedded clause's being a scope island, which prevents the universal from taking wide scope. The ambiguity of (33) leads us to conclude that indefinites, in contrast to universals, can scope out of clauses, complements and coordination structures. Szabolcsi (2000) gives a large amount of evidence for this conclusion. Accordingly, our theory must first explain how anything can get out of a scope island, then postulate that only indefinites have this escaping ability.

The operation $(\!|\cdot|\!)$ that effects the scope island has the schematic type that can be informally depicted as $\mathrm{CPS}^i[t] \to \mathrm{CPS}^0[t]$ where

$$\mathrm{CPS}^i[t] = \{\alpha\} \quad \text{where the length of } \alpha \text{ is } 2^{i+1} - 1$$

Since the result of $(\!|m|\!)$ has a CPS^0 type, that is t, the result cannot get hold of any context. Hence we need a less absolutist version of $(\!|\cdot|\!)$ which merely lowers rather than collapses the hierarchy. We call that operation $(\!|\cdot|\!)_2$, of the informal schematic type $\mathrm{CPS}^{\leq 2}[t] \to \mathrm{CPS}^0[t]$ and $\mathrm{CPS}^{i+2}[t] \to \mathrm{CPS}^i[t]$ where $i \geq 1$. Whereas $(\!|m|\!)$ delimits all the contexts of m, $(\!|m|\!)_2$ delimits only the first two contexts of the hierarchy. Quantifiers within m of level 3 and higher will be able to get hold of the context of $(\!|m|\!)_2$. One may think of $(\!|\cdot|\!)_2$ as the inverse of $\uparrow\uparrow$. The following example illustrates the lowering:

(37a) $$[\![\textsf{someone}_1\ \textsf{left}]\!] = \lambda k_1.\, \lambda k_2.\, k_1\ (\mathbf{leave} \cdot n)\ (\lambda v.\, k_2 \exists_n v)$$

(37b) $$(\!|[\![\textsf{someone}_1\ \textsf{left}]\!]|\!)_2 = \exists_0(\mathbf{leave} \cdot 0)$$

(37c) $$(\!|\uparrow [\![\textsf{someone}_1\ \textsf{left}]\!]|\!)_2 = \exists_0(\mathbf{leave} \cdot 0)$$

(37d) $$(\!|\uparrow\uparrow [\![\textsf{someone}_1\ \textsf{left}]\!]|\!)_2 = \lambda k_1.\, \lambda k_2.\, k_1\ (\mathbf{leave} \cdot n)\ (\lambda v.\, k_2 \exists_n v)$$

In (37a) and the identical (37d), the existential quantifies over the potentially wide context k_1. In (37b) and (37c), whose denotations are again identical, $\exists_0$ scopes just over **leave** $\cdot$ 0 and extends no further.

Why did we choose 2 as the number of contexts to delimit at the embedded clause boundary? Any number $i \geq 2$ will work, to explain the quantifier ambiguity within the embedded clause and wide-scope indefinites. We chose $i = 2$ for now pending analysis of more empirical data.

If (37a)–(37d) is the specification for $(\!|\cdot|\!)_2$, then (38) below is the implementation. It is derived from the schema (26) by cutting it off after the third line and inserting the generic lowering-by-two operation as the final default case.

Syntax			Semantic type	Denotation $[\![\cdot]\!]$
VP	$\to$	Vs that S	et^n or $((et)^n \to \{\alpha\}) \to \{\beta\}$	$[\![\text{Vs}]\!]^{>} (\![[\![\text{S}]\!]]\!)_2$
NP	$\to$	that S	e^n or $(e^n \to \{\alpha\}) \to \{\beta\}$	$\textbf{That} \mathbin{.} (\![[\![\text{S}]\!]]\!)_2$
NP	$\to$	NP_1 and NP_2	e^n or $(e^n \to \{\alpha\}) \to \{\beta\}$	$\textbf{alongWith} \mathbin{.} [\![\text{NP}_1]\!] \mathbin{.} [\![\text{NP}_2]\!]$
N	$\to$	max	e	**max**
N	$\to$	lady	et	**lady**
Det	$\to$	some_3, a_3	$(et) \to (e^{n+1} \to \{\beta_3(n+1)\gamma\}) \to \{\beta_3 n\gamma\}$	$\lambda z. \uparrow [\![\text{some}_2]\!]\, z$
Det	$\to$	some_4, a_4	$(et) \to (e^{n+1} \to \{\beta_4(n+1)\gamma\}) \to \{\beta_4 n\gamma\}$	$\lambda z. \uparrow [\![\text{some}_3]\!]\, z$

Fig. 13 Adjustments to the syntax and the multi-level direct-style continuation semantics for the additional fragment, to account for wide-scope indefinites. If the size of the sequence γ is j, the size of β_3 is $3(j+2)$ and of β_4 is $7(j+2)$

(38)
$$(\![m]\!)_2 \stackrel{\text{def}}{=} \begin{cases} m & \text{if } m : \{0\} \\ m(\lambda v.\, v) & \text{if } m : \{nn0\} \\ m(\lambda v.\, \lambda k.\, kv)(\lambda v.\, v) & \text{if } m : \{nnl_1 l_1 l_2 l_2 0\} \\ m(\lambda v.\, \lambda k.\, kv)(\lambda v.\, \lambda k.\, kv) & \text{otherwise} \end{cases}$$

It is easy to show that the definition (38) indeed satisfies (37a)–(37d). A useful lemma is the identity $(\uparrow m)(\lambda v.\, \lambda k.\, kv) = m$, easily verified from the definition (27) of $\uparrow$.

To make use of this lowering operation $(\![\cdot]\!)_2$, we adjust the lexical entries in Fig. 12 as shown in Fig. 13. The main change is replacing $(\![\cdot]\!)$ in the semantic composition rules for embedded clauses with $(\![\cdot]\!)_2$. In other words, we now distinguish the main clause boundary from embedded clause boundaries. Figure 13 also reflects our postulate: only indefinites may be at the CPS level 3 and higher—not universals.

The typical example (33) can now be analyzed as follows (see Fig. 12 for the denotations of disappeared and report):

(39) $[_{\text{M}}$ Everyone$_1$ reported that $[_{\text{S}}$ Max and some$_{\text{i}}$ lady disappeared$]$.$]$

When the level i of some$_{\text{i}}$ is 1 or 2, the indefinite is trapped in the scope island.

(40a) $\forall_0 \textbf{report} \cdot (\exists_0 (\textbf{lady} \cdot 0 \wedge \textbf{disappear} \cdot (\textbf{alongWith} \cdot \textbf{max} \cdot 0))) \cdot 0$

At the level $i = 3$, the indefinite scopes out of the clause but is defeated by the universal in the subject position, giving us another linear-scope reading, along the lines expounded in Sect. 2.2.

(40b) $\forall_0 \exists_1 \textbf{lady} \cdot 1 \wedge \textbf{report} \cdot (\textbf{disappear} \cdot (\textbf{alongWith} \cdot \textbf{max} \cdot 1)) \cdot 0$

Finally, some$_4$, lowered from level 4 to level 2 as it crosses the embedded clause boundary, has sufficient strength left to scope over the entire sentence.

(40c) $\exists_0 \textbf{lady} \cdot 0 \wedge \forall_1 \textbf{report} \cdot (\textbf{disappear} \cdot (\textbf{alongWith} \cdot \textbf{max} \cdot 0)) \cdot 1$

Syntax	Semantic type	Denotation $[\![\cdot]\!]$
$\mathsf{N'} \to \mathsf{N}$	$e^n \to n\alpha$	$\lambda x.\,\lambda k.\,k\ ([\![\mathsf{N}]\!] \bullet x)$
$\mathsf{N'} \to \mathsf{N'}\ \mathsf{PP}$	$e^n \to n\alpha$	$\lambda x.\,(\wedge)^{>}([\![\mathsf{N'}]\!]\ \ x)^{>}([\![\mathsf{PP}]\!]\ \ x)$
$\mathsf{PP} \to \mathsf{from}\ \mathsf{NP}$	$e^n \to n\alpha$	$\lambda x.\,(\!\lvert \mathbf{from}^{>}\ [\![\mathsf{NP}]\!]\ ^{>}x \rvert\!)_2$
$\mathsf{NP} \to \mathsf{Det}\ \mathsf{N'}$	$(e^n \to \{\alpha\}) \to \beta$	$[\![\mathsf{Det}]\!]\ [\![\mathsf{N'}]\!]$
$\mathsf{Det} \to \mathsf{every}_1$	$(e^{n+1} \to \{(n+1)n\alpha\beta\gamma\})$ $\to (e^{n+1} \to \{(n+1)\alpha\beta\}) \to \{\gamma\}$	$\lambda z.\,\lambda k_1.\,z\ n\ (\lambda x.\,\lambda k_2.$ $k_1\ n\ (\lambda v.\,k_2\ (\forall_n(x \Rightarrow v))))$
$\mathsf{Det} \to \mathsf{some}_1,\ \mathsf{a}_1$	$(e^{n+1} \to \{(n+1)n\alpha\beta\gamma\})$ $\to (e^{n+1} \to \{(n+1)\alpha\beta\}) \to \{\gamma\}$	$\lambda z.\,\lambda k_1.\,z\ n\ (\lambda x.\,\lambda k_2.$ $k_1\ n\ (\lambda v.\,k_2\ (\exists_n(x \wedge v))))$
$\mathsf{Det} \to \mathsf{no}_1$	$(e^{n+1} \to \{(n+1)n\alpha\beta\gamma\})$ $\to (e^{n+1} \to \{(n+1)\alpha\beta\}) \to \{\gamma\}$	$\lambda z.\,\lambda k_1.\,z\ n\ (\lambda x.\,\lambda k_2.$ $k_1\ n\ (\lambda v.\,k_2\ (\neg \bullet \exists_n(x \wedge v))))$

Fig. 14 Adjustments to the syntax and the multi-level direct-style continuation semantics for the additional fragment, to account for prepositional phrases. The higher-level quantificational determiners are produced with the $\uparrow$ operations; see Fig. 13 for illustration. If the size of the sequence α is j, the size of β is also j and the size of γ is $2j+1$

5.3 Inverse Linking

Our analysis of inverse linking turns out quite similar to the analysis of wide-scope indefinites. We take the argument NP of a PP to be a scope island, albeit it is evidently a weaker island than an embedded tensed clause. We realize the island by an operation similar to $(\!\lvert \cdot \rvert\!)_2$. Therefore, a strong enough quantifier embedded in NP can escape and take a wide scope. That escaping from the island corresponds to inverse linking.

To demonstrate our analysis, we extend our fragment with prepositional phrases; see Fig. 14. We add a category of N' of nouns adjoined with PP. We generalize Det to take as its argument N' rather than bare common nouns. For simplicity, we use the same $(\!\lvert \cdot \rvert\!)_2$ operation for the PP island as we used for the embedded-clause island. Recall that $(\wedge)$ is a constant of the type $t(tt)$ and we write the $\mathcal{D}_Q$ expression $(\wedge) \cdot d_1 \cdot d_2$ as $d_1 \wedge d_2$.

The type of the quantificational determiners shows that a determiner takes a restrictor and a continuation, which may contain n other free variables. The determiner adds a new one, which it then binds. Although the denotations of determiners in Fig. 14 bind the variables they themselves introduced, that property is not assured by the type system. For example, nothing prevents us from writing 'bad' lexical entries like $\lambda z.\,\forall_n z$ or 1. Although the type system will ensure that the overall denotation is closed, what a binder ends up binding will be hard to predict. It is an interesting problem to define 'good' lexical entries (with respect to scope) and codify the notion in the type system. This is the subject of ongoing work Kameyama et al. (2011).

We analyze inverse linking thusly.

(41a) $[_{\mathsf{NP}}$ No$[_{\mathsf{N'}}\,[_{\mathsf{N'}}$ man$][_{\mathsf{PP}}$ from a foreign country$]]]$was admitted.

(41b) $neg \cdot \exists_0\mathbf{man} \cdot 0 \wedge (\exists_1\mathbf{country} \cdot 1 \wedge \mathbf{from} \cdot 1 \cdot 0) \wedge \mathbf{admitted} \cdot 0$

(41c) $exists_0$**country** $\cdot\, 0 \wedge \neg \cdot \exists_1$**man** $\cdot\, 1 \wedge$ **from** $\cdot\, 0 \cdot 1 \wedge$ **admitted** $\cdot\, 1$

The PP in (41a) contains an ambiguous quantifier. If the quantifier is weak, it is trapped in the PP island and gives the salient reading (41b). If the quantifier is strong enough to escape, the inverse-linking reading (41c) emerges. We thus reproduce quantifier ambiguity for QNP within NP and explain inverse linking.

6 Conclusions

We have given the first rigorous account of linear- and inverse-scope readings, scope islands, wide-scope indefinites and inverse linking based on the D&F continuation hierarchy. Quantifier ambiguity arises because quantifier words are polysemous, with multiple denotations corresponding to different levels of the hierarchy. The higher the level, the wider the scope. Embedded clauses and PPs create scope islands by lowering the hierarchy and trapping low-level quantifiers. Higher-level quantifiers (which we postulate only indefinites possess) can escape the island and take wider scope. The continuation hierarchy lets us assign scope to indefinites and universals and explain their differing scope-taking abilities.

Our analysis is directly compositional: each surface syntactic constituent corresponds to a well-formed denotation, and each surface syntactic formation rule corresponds to a *unique* semantic combinator.

We have shown how to build the continuation hierarchy modularly and on-demand, without committing ourselves to any particular hierarchy level but raising the level if needed as a denotation is being composed. In particular, quantifier-free lexical entries have unlifted types and simple denotations.

We look forward to extending our analysis to other aspects of scope—how quantifiers interact with coordination (as in (1.1)), pronouns and polarity items—and to distributivity in universal quantification. We would also like to investigate if hierarchy levels can be correlated with Minimalism features or feature domains. Finally, we plan to extend our analyses of single sentences to discourse.

Acknowledgments We are very grateful to Chris Tancredi for many helpful suggestions and a thought-provoking conversation. We thank anonymous reviewers for their comments.

References

Barker, C. (2002). Continuations and the nature of quantification. *Natural Language Semantics*, *10*(3), 211–242.

Barker, C., & Shan, C.-c. (2006). Types as graphs: Continuations in type logical grammar. *Journal of Logic, Language and Information*, *15*(4), 331–370.

Barker, C., & Shan, C.-c. (2008). Donkey anaphora is in-scope binding. *Semantics and Pragmatics1*(1), 1–46.

Bekki, D., & Asai, K. (2009). Representing covert movements by delimited continuations. *Proceedings of the* 6th *International Workshop on Logic and Engineering of Natural Language Semantics*, Japanese Society of, Artificial Intelligence (November 2009).

Bernardi, R., & Moortgat, M. (2010). Continuation semantics for the Lambek-Grishin calculus. *Information and Computation*, *208*(5), 397–416.

Champollion, L., Tauberer, J., & Romero, M. (2007). The Penn Lambda Calculator: Pedagogical software for natural language semantics. In T. H. King & E. M. Bender (Eds.), *Proceedings of the Workshop on Grammar Engineering Across Frameworks* (pp. 106–127). Stanford: CA, Center for the Study of Language and Information.

Danvy, O., & Filinski, A. (1990). Abstracting control. *Proceedings of the 1990 ACM Conference on Lisp and Functional Programming* (pp. 151–160), New York, ACM Press (27–29 June 1990).

de Groote, P. (2001). Type raising, continuations, and classical logic. In R. van Rooy & M. Stokhof (Eds.), *Proceedings of the* 13th *Amsterdam Colloquium* (pp. 97–101). Institute for Logic: Language and Computation, Universiteit van Amsterdam.

Kameyama, Y., Kiselyov, O., & Shan, C.-c. (2011). Computational effects across generated binders: Maintaining future-stage lexical scope. Technical Report CS-TR-11-17, Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba.

Montague, R. (1974). The proper treatment of quantification in ordinary English. In R. H. Thomason (Ed.), *Formal philosophy: Selected papers of Richard Montague* (pp. 247–270). New Haven: Yale University Press.

Partee, B. H., & Rooth, M. (1983). Generalized conjunction and type ambiguity. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning use and interpretation of language* (pp. 361–383). Berlin: Walter de Gruyter.

Reinhart, T. (1979). Syntactic domains for semantic rules. In F. Guenthner & S. J. Schmidt (Eds.), *Formal semantics and pragmatics for natural languages* (pp. 107–130). Dordrecht: Reidel.

Rompf, T., Maier, I., & Odersky, M. (2009). Implementing first-class polymorphic delimited continuations by a type-directed selective CPS-transform. In G. Hutton & A. P. Tolmach (Eds.), *ICFP '09: Proceedings of the ACM International Conference on Functional Programming* (pp. 317–328). New York: ACM Press.

Shan, C.-c. (2004). Delimited continuations in natural language: Quantification and polarity sensitivity. In H. Thielecke (Ed.), *CW'04: Proceedings of the* 4th *ACM SIGPLAN Continuations Workshop* (pp. 55–64). Number CSR-04-1 in Technical Report, School of Computer Science, University of Birmingham.

Shan, C.-c. (2007a). Linguistic side effects. In C. Barker & P. Jacobson (Eds.). *Direct compositionality* (pp. 132–163). New York: Oxford University Press.

Shan, C.-c. (2007b). Inverse scope as metalinguistic quotation in operational semantics. In: K. Yoshimoto (Ed.), *Proceedings of the* 4th *International Workshop on Logic and Engineering of Natural Language Semantics* (pp. 167–178), Japanese Society of Artificial Intelligence (18–19 June 2007).

Szabolcsi, A. (2000). The syntax of scope. In M. Baltin & C. Collins (Eds.), *Handbook of contemporary syntactic theory* (pp. 607–634). Oxford: Blackwell.

Szabolcsi, A. (2009). *Quantification*. Cambridge: Cambridge University Press.

Japanese Reported Speech: Towards an Account of Perspective Shift as Mixed Quotation

Emar Maier

Abstract English direct reported speech is easily distinguished from indirect reported speech by, for example, the lack of a complementizer (*that*), the quotation marks (or the accompanying prosody), and/or verbatim ('shifted') pronouns. By contrast, Japanese employs the same complementizer for all reports, does not have a consistent intonational quotation marking, and tends to drop pronouns where possible. Some have argued that this shows no more than that many Japanese reports are ambiguous. They claim that, despite the lack of explicit marking, the underlying distinction is just as hard in Japanese as it is in English. On the basis of a number of 'mixed' examples, I claim that the line between direct and indirect *is* blurred and I propose a unified analysis of speech reporting in which a general mechanism of mixed quotation replaces the classical two-fold distinction.

Keywords Reported speech · Direct–indirect discourse · Japanese · Quotation · Semantics

I would like to thank the Japanese friends I used as informants for valuable suggestions along with their judgments: Mana Kobuchi-Philip, Yurie Hara, Hidetoshi Shiraishi, and Katsuhiko Yabushita. Further thanks go to Jennifer Spenader and the participants and organizers of LENLS 2008 in Asahikawa, as well as the anonymous referees for this volume. The research is supported by the the EU under FP7 (ERC Starting Grant 263890-BLENDS).
This chapter is an updated and extended version of a chapter that appeared in the Proceedings of LENLS 2008 (Maier 2009). While the data and general line of argumentation is the same, the current version makes the formal machinery of pure and mixed quotation more precise, among other things

E. Maier (✉)
Department of Philosophy, University of Groningen, oude boteringestraat 52,
9712GL Groningen, Netherlands
e-mail: emar.maier@gmail.com

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*,
Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_7,

1 Introduction

There is an obvious contrast between:

(1) a. Taro said that I would go to Tokyo
    b. Taro said: "I will go to Tokyo"

The first is an example of indirect speech, in which I report what Taro said on an earlier occasion in my own words; the second is a direct report, where I report Taro by quoting his words verbatim. From this informal characterization it follows that in (1a) the pronoun *I* is used by (and refers to) me, the reporter, whereas in (1b) it refers to Taro, which explains the truth-conditional divergence between the two sentences.

There is a deep disagreement about the nature of the direct–indirect distinction. On the one side are formal semanticists and philosophers, who see a rigid use–mention distinction (Kaplan 1989); on the other are people working in cognitive linguistics and pragmatics who emphasize a continuum (Voloshinov 1973). As a case study, in this chapter I focus on one specific subtopic within this general debate: the distinction between direct and indirect discourse in Japanese. Japanese reported speech is interesting in this regard, as there is relatively little marking of the distinction, i.e. in a colloquial setting, a sentence like (2) has two interpretation, one in which the embedded first person refers to me, as in (1a), and one in which it refers to Taro, as in (1b) (Hirose 1995, p. 224):

(2) Taroo-wa boku-ga Tookyoo e iku to itta
    Taro-Top I-Nom Tokyo to go Comp said

The received view, especially among formal semanticists, now seems to be that Japanese reports like (2) are simply ambiguous between direct and indirect. The current chapter offers more Japanese data to argue against the notion of a categorical direct–indirect distinction. I present an alternative in which mixed quotation allows one to 'defer' any part of any report complement. The resulting general framework can be seen as a formal semanticist's first stab at capturing the grey area between the extremes of fully direct and fully indirect discourse.

But first we take a closer look at the motivation for the traditional distinction between direct and indirect speech.

2 Distinguishing Indirect and Direct Speech

The difference between direct and indirect speech is marked in a number of different ways in different languages. Let's go through a couple of the better known ones (cf. e.g. Banfield 1973 for a thorough overview).

Syntax/semantics English indirect discourse is usually marked by a complementizer *that*; in Dutch, such a complementizer and an additional change in word order are obligatory; in German, indirect discourse requires changes in both word order and mood of the verb. A distinguishing feature of direct speech syntax is its ‘syntactic opacity’ (Oshima 2006), i.e. it blocks movement, (3), quantifying in, (4), and NPI licensing, (5) (Anand and Nevins 2004):

(3) a. What did Taro say he had seen
b. *What did Taro say: “I have seen”?

(4) a. There’s something that Taro says he likes
b. *There’s something that Taro says: “I like”

(5) a. Nobody said they had seen anything
b. ??Nobody said “we saw anything?[1]

Orthography/prosody In written languages, direct speech is usually marked with quotation marks. In spoken language this direct speech marking tends to surface as a distinct intonational contour (Klewitz and Couper-Kuhlen 1999; Jansen et al. 2001).

Semantics/pragmatics As noted above, reporting someone’s words in indirect speech requires adjusting the original utterance’s indexicals to the reporting context. To report the same as (1b) in indirect speech, Taro’s *I* would have to be changed to *he*. In English, the same holds for indexicals like *tomorrow* and the present tense. Note however the cross-linguistic variation: in Russian, the present tense is not adjusted, while in Amharic even first person forms can apparently be retained (Schlenker 2003; Anand 2006).

These and other characteristics indeed give the impression of a “binary, categorical distinction” where “a direct report is about a relation between an agent and a linguistic object while an indirect report is about a relation between an agent and a proposition” (Oshima 2006, p. 23). This traditional explanation of the direct–indirect distinction seems to rest on a fundamental distinction between two functions of language: words can be used to refer to the world (*use*), but also to refer to words and other linguistic items (*mention*). Before arguing against it, let me first clarify the supposed link between indirect–direct and use–mention. In letting language refer to itself, mentioning poses a serious, and somewhat neglected, challenge to compositionality and hence to formal semantics. For this reason I will be rather explicit in my formalization below.

3 Modeling the Indirect–Direct Distinction as Use Versus Mention

In this section I introduce some formal machinery to model the traditional, rigid direct indirect distinction. We need: (i) a logic of indirect discourse as an intensional operator (3.1); a logic of mentioning, the use of language to refer to linguistic expressions, with quotation marks (3.2); and a (preliminary) account of direct discourse as pure quotation (3.3).

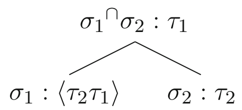
[1] The sentence as a whole is grammatical, and likely true. It does not, however, report the same as (5a).

3.1 Indirect Discourse

Model-theoretically, language *use* is what's captured by the familiar Fregean semantics. A proper analysis of indirect speech reporting and indexicality requires Kaplan's two-dimensional version, which analyzes indirect *saying that* as an intensional operator. Roughly, x *says that* φ means that x uttered some sentence that expressed the same proposition in its original utterance context as that expressed by φ in the current report context.[2]

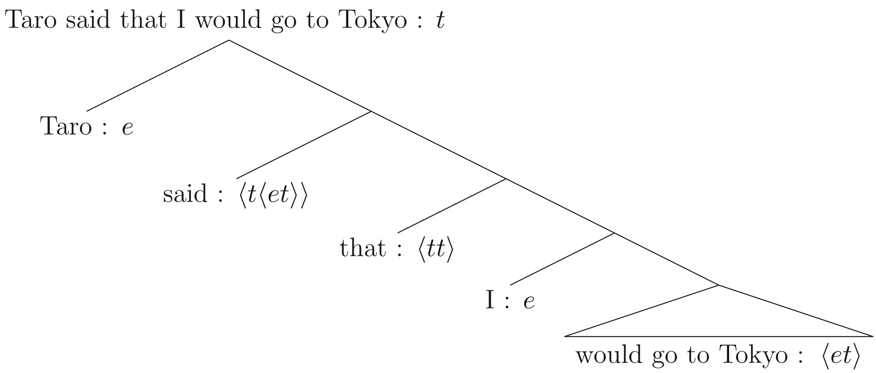
For concreteness, let's set up a simplified formal syntax and semantics to be used throughout the chapter. The syntax analyses sentences as trees, whose nodes contain linguistic material (say, finite strings of letters) coupled with a category. For simplicity, we'll just use the familiar semantic types as our syntactic categories. In other words, syntax is entirely type-driven,[3] i.e. the types determine whether two subtrees can be combined or not, via the following rule:

(6)



For now, the basic types are e for simple entity referring expressions, and t for sentences expressing propositions (=sets of possible worlds).[4] To illustrate this perhaps somewhat unusual presentation of a rather familiar system, I'll derive the indirect speech report in (1a), that is, using (6) we construct a tree whose top node contains the sentence, of type t. The leaves will be lexical entries, like 'Taro : e', and 'said : $\langle t\langle et\rangle\rangle$', which gives something like this:

(7)



---

[2] The operator that this amounts to is intensional because to evaluate whether x *says that* φ is true we need to consider the proposition expressed by φ, rather than φ's truth value, or its exact wording (or Kaplanian "character").

[3] This oversimplification leads to massive overgeneration, but that will not concern us here.

[4] I'm assuming that t is the type for referring to sets of possible worlds rather than truth values. Nothing hinges on this; we could use type s and $^\wedge$, or an intensional functional application rule instead.

Call this simple tree grammar G. Trees in G can be interpreted compositionally, but, because we will be needing the representational framework of DRT to model presupposition resolution later on, we will map them onto an intermediate logical language first, viz. the language of *preliminary DRSs*. I'll suppress types in DRSs when possible, and I treat names and indexicals as *in situ* constants rather than presupposition triggers for now. Translations of our terminal nodes look something like this. As for notation: boxes denote DRSs, which are a special type of expressions of type t; typically they contain a set of regular type t expressions (DRS Conditions) and a set of variables (DRS Universe), although I leave out the latter if it's empty.

(8) a. $\mathbb{T}(\text{Taro}: e) = \texttt{t}$
b. $\mathbb{T}(\text{would go to Tokyo}: \langle et \rangle) = \lambda \texttt{x} \boxed{\text{go} - \text{tokyo(x)}}$
c. $\mathbb{T}(\text{said}: \langle t \langle et \rangle \rangle) = \lambda \texttt{X} \lambda \texttt{x} \boxed{\text{say(x)} : \text{X}}$

For complex expressions, composition is function application:

(9)
$$\mathbb{T}\left(\begin{array}{c} \sigma_1{}^\frown\sigma_2 : \tau_1 \\ \diagup \quad \diagdown \\ \sigma_1 : \langle \tau_2\tau_1 \rangle \quad \sigma_2 : \tau_2 \end{array} \right) = \mathbb{T}(\sigma_1 : \langle \tau_2\tau_1 \rangle)(\mathbb{T}(\sigma_2 : \tau_2))$$

Our example indirect report translates as:

(10) $\boxed{\text{say(t)} : \boxed{\text{go} - \text{tokyo(i)}}}$

The logical expressions of DRT are, in turn, interpreted model-theoretically. Every type has an associated domain into which expressions of that type are mapped by the interpretation function, $[\![]\!]$. To deal with the semantic context dependence of indexicals we relativize semantic interpretation to a context parameter c. Thus, we can use Kaplan's (1989) logic of demonstratives to interpret *I* and indirect speech *say*. Later on, we'll add a discourse context, to deal with presupposition resolution in a dynamic way—only then will the added benefit of using DRSs become apparent.

(11) a. $[\![\texttt{i}]\!]^c$ = the speaker of c
b. $[\![\texttt{say}(\alpha) : \varphi]\!]^c$ = the proposition that $[\![\alpha]\!]^c$ utters some sentence that in her utterance context c′ expresses proposition$[\![\varphi]\!]^c$

Complex expressions in DRT can be read as a notational variant of static type logic. Line breaks indicate conjunction, boxes indicate a term of type t (a DRS). So, to finish our example, the proposition expressed by *Taro said that I would go to Tokyo* in context c, is: $[\![\mathbb{T}$ (Taro said that I would go to Tokyo)$]\!]^c$ = $[\![(10)]\!]^c$ = the proposition that Taro utters some sentence that, in its original utterance context, expressed the proposition that I would go to Tokyo.

3.2 Mention

I'm following the so-called disquotational theory of pure quotation, by which an expression enclosed in quotation marks refers to the enclosed expression: *sweet*, a predicate of type *et*, refers to the set of sweet things, but *'sweet'*, the same expression flanked by quotation marks, refers to a word, viz. the word *sweet*. To formalize the disquotational semantics, we first need to extend our semantic domains to include linguistic material, i.e. we need to add a new type *u* for expressions that refer not to objects or properties in the world, but to linguistic entities. We can turn every expression in *G* into a type *u* term by putting quotation marks around it. More precisely, we add the following composition rule (cf. Potts 2007a):

(12)

$$\begin{array}{c} `^{\cap}\sigma^{\cap}\text{'} : u \\ | \\ \sigma : \tau \end{array}$$

So, since *sweet* : $\langle et \rangle$ is a well-formed expression in *G*, so is *'sweet'* : *u*. Semantically, the former refers to a set of entities (type $\langle et \rangle$), the latter to a linguistic entity (type *u*). More specifically, a quotation refers to the quoted expression, i.e. the typed expression in *G* underneath it, in this case *sweet* : $\langle et \rangle$. Again, we'll have to go through the intermediate translation into the language of DRT (in which I'll use $\ulcorner\urcorner$ as quotation marks):

(13)

a. $$\mathbb{T}\left(\begin{array}{c} `^{\cap}\sigma^{\cap}\text{'} : u \\ | \\ \sigma : \tau \end{array}\right) = \ulcorner \sigma : \tau \urcorner$$

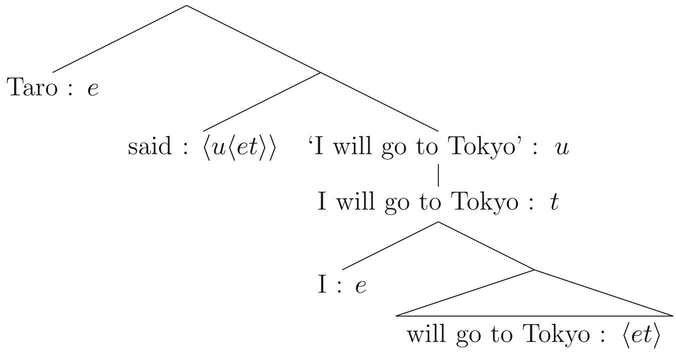
b. $$[\![\ulcorner \sigma : \tau \urcorner]\!]^c = \sigma : \tau \ (\in D_u = G)$$

Thus, despite the detour, a quoted term really does refer to the expression in quotes. This quoted expression itself must belong to *G*, i.e. a string of letters paired with a semantic type, but it need not have a semantic interpretation. After all, the meaning of a quoted expression does not depend on the meaning of its constituent, but rather on the expression itself. This is precisely the sense in which quotation is not fully compositional (Werning 2005). I consider this an important feature of quotation, as it explains, for example, why grammatical errors can be felicitously quoted, and why you can't usually "quantify into" pure quotations. A Bushism like 'misunderestimate' will be modeled as semantically vacuous but well-formed expressions of type $\langle e \langle et \rangle \rangle$ (i.e. transitive verb/two-place relation) and can therefore only be used meaningfully in quotes (cf. Maier 2008).

3.3 Direct Speech

The logic of mentioning above extends to a straightforward semantics of direct speech: Simply analyze 'say' in its direct discourse sense as a transitive verb that takes as direct object a term of type *u*, and analyze direct quotation marks as pure quotes, capturing the traditional view of direct discourse being a relation between an individual and an utterance:

(14) a. Taro said, "I will go to Tokyo" : t



b. $\mathbb{T}((14a)) = \boxed{\texttt{say}_{dd}\texttt{(t,}\ulcorner\text{I will go to Tokyo} : t\urcorner\texttt{)}}$

c. $[\![(14b)]\!]^c$ = the set of worlds in which Taro utters "I will go to Tokyo : t"

I should stress that the DRS translation and hence the interpretation ignores the semantics of 'I will go to Tokyo : t' and everything underneath it in the tree (14a) completely. This is ultimately unsatisfactory: ellipsis and anaphora, among other things, show that direct quotation is not entirely opaque in this sense. I will not pursue this matter here, but refer to Maier (2008) for an analysis.

Taking Stock

The direct–indirect discourse distinction can be cashed out formally in an intensional logic with a mention operator. Indirect discourse saying translates as an intensional operator, i.e. a relation of type tet, while direct discourse saying translates as a relation of type uet.

The distinguishing characteristics of direct and indirect speech listed in the first section all follow from this semantics.

Syntax/semantics Direct speech's 'verbatimness' with respect to clause structure and word order, among other things, follows from the fact that it is the original utterance itself that is the object of the say (uet) relation. The fact that mentioning turns the quote into a (self-)referential term of type u with, semantically speaking, no internal structure, explains the opacity with respect to movement and NPI licensing at the syntax–semantics interface.

Orthography/prosody The various forms of quotation marking in direct speech fall out simply as surface realizations of the pure quotation operator. Crucially, in our grammar, the quotation marks of direct discourse or pure quotation are not "just" punctuation. They have a genuine semantic and syntactic impact.

Semantics/pragmatics Indexical adjustment in indirect speech follows from the Kaplanian semantics of indirect speech where we have to match the proposition that was expressed in the reported context with the proposition expressed by the complement clause in the reporting context.

4 Challenging the Indirect–Direct Distinction: The Case of Japanese

Despite this apparent success of a rather simple semantics, Maier (2008) challenges the strict indirect–direct distinction by pointing out that even English direct discourse is semantically somewhat transparent. This claim is backed by the observations that (i) anaphoric and elliptical dependencies can cross direct discourse boundaries (as in *"My girlfriend bought me this tie," said John, but I don't think she did*, cf. Partee 1973), and (ii) a direct report triggers a rather strong (default) inference that the corresponding indirect version is also true (for example, the direct (1b) implies that Taro said that he would go to Tokyo).

For so-called mixed quotation (Cappelen and Lepore 1997), consisting of an indirect report in which only a certain part is directly quoted, Maier's (2008) case is strengthened by additional syntactic/semantic evidence. But, focusing on genuine direct discourse, it may well be possible to get around both of the transparancy arguments by adding a distinct pragmatic mechanism that leaves the separation of direct and indirect discourse intact at the semantic level.[5] In the remainder of this chapter I present some further evidence against a rigid direct–indirect distinction.

4.1 A Rumor About Japanese Speech Reporting

As "rumor has it that there is no such [direct–indirect] distinction in Japanese" (Coulmas 1985, p. 53) I turn to that language in hope to seal the fate of the classical report distinction. My ultimate goal is to replace it with an analysis of speech reports as indirect discourse (analyzed in terms of 3.1's "say $:tet$") with optional mixed quotation of any notable parts. Unfortunately, some work remains to be done as Coulmas continues the sentence quoted above by remarking that the rumor about Japanese is "obviously not true."

Let's reconstruct how the rumor might have started originally. Recall our enumeration of the ways in which direct and indirect discourse can be kept apart. Now note

[5] I know of no actual proposal to this effect, but I envisage a kind of system that takes the strictly separatist semantics of direct speech as mention and combines it with a strengthening mechanism that adds the corresponding indirect version of a direct report, the *use* inference, to the semantic representation. Assuming that the resolution of ellipsis and anaphora triggered by the following discourse apply after pragmatic strengthening of a direct report, would derive (i) as well.

that, syntactically, Japanese does not distinguish direct and indirect discourse by a special complementizer. The marker *to* is used for all speech reporting. Tense and word-order are consistently retained in speech reports, nor is there a special mood for indirect discourse. Then, ortographically, direct discourse in written text may often be recognizable from the quotation marks, but in colloquial spoken language these may go unnoticed.[6]

So, of the previously listed tests for distinguishing the two modes we are left with indexical adjustment and syntactic transparency as indicators of indirectness. Unfortunately, these characteristics are invisible in a single given sentence itself, so less useful for the task of classifying reports that are not otherwise marked. In addition, the clearest examples of indexicals, person pronouns, tend to be dropped in colloquial Japanese. For these reasons our current test battery will indeed fail to classify many reports as either direct or indirect. According to Coulmas, this is the source of our rumor.

So what does this mean for the interpretation of Japanese reports? Given a strict, logical direct–indirect separation (Coulmas, Hirose, Oshima *op. cit.*), many reports must be simply ambiguous between the two distinct derivations demonstrated in Sect. 3.1 and 3.3. So, even with overt pronouns, we will often have to rely on the context to disambiguate. A case in point is (2), where taken on its own perhaps no more than a slight pause distinguishes the readings (1a) and (1b).[7] Presumably, the context will favor one of these readings, so, as Coulmas rightly observes, this syntactic/semantic ambiguity need not hinder communication, yet a genuine ambiguity it is nonetheless.

Separatists, like Coulmas, Hirose and Oshima, point out that, to facilitate contextual disambiguation, Japanese can rely on a very rich repertoire of what Hirose (1995) calls "addressee-oriented expressions." These include particles like *yo* and *ne*, imperatives, and honorifics like the polite *-masu* verb forms. Like traditional indexicals, the meanings of such expressions are tied to the actual utterance context (Potts and Kawahara 2004) and "semantically presuppose the existence of an addressee" (Hirose 1995) in that context. For speech reporting this means that such expressions can only occur in direct speech, or else, when they do occur embedded in an indirect report, apply only to the actual reporter and her relation to her addressee. Unfortunately for the separatist's cause, this prediction is not borne out, as I show next.

[6] One informant speaks of a distinct quotation intonation, another of a short pause after the quote should clarity demand it. Further research is required to establish whether the intonational clues in Japanese are more subtle than in English.

[7] If the report was made in Tokyo, *kuru* ('come') could be used to indicate indirect discourse, though *iku* ('go') would still be compatible with indirect discourse too, as indirect discourse is known to shift the indexical goal parameter of *come/go* in Japanese. More on this below.

4.2 Neither Direct Nor Indirect: The Data

Take the embedded honorific *-masu* (+past = *-mashita*) form in:

(15) kare wa watashi ga matta machigaimashita to iimashita
he Top I Nom again was.wrong-Polite Comp said-Polite
a. 'He said [polite]: "I was wrong again [polite]"'
b. 'He said [polite] that I was wrong again [polite]' (Coulmas 1985:57)

The embedded first person pronoun could well be the reported speaker's, as in the direct reading (15a), but, according to Coulmas, it could also refer to the reporter, in which case we should be dealing with indirect discourse, (15b). The question is, who is being polite to whom with *machigaimashita*? Unless it's a direct quote it must refer to the context of the report, but the reporter has already expressed his politeness to his addressee sufficiently in the matrix verb. Coulmas claims that even in the indirect reading, (15b), it could indicate politeness of the reported speaker, apparently contradicting the indexical addressee-orientation of *-masu*. For now let's use the term 'shifting' for the phenomenon of (arguably) addressee-oriented expressions used in (arguably) indirect speech and interpreted with respect to the reported context/speech act.

Shifted addressee-orientation in indirect speech is not restricted to the occasional embedded *-masu* form (as Coulmas seems to suggest). Here is an example of what Kuno (1988) would call 'blended quasi-direct discourse' with an imperative. My boss tells me:

(16) asatte made ni kono shigoto-o yare
day after tomorrow until this work-Acc do-Imp-Familiar
'Finish [familiar] this work in two days!'

If I want to report this to you the next day, I might say:

(17) ashita made ni sono shigoto-o yare to jooshi-ni
tomorrow until that work-Acc do-Imp-Fam Comp boss-by
iwaremashita
was told-Polite
'I was told [polite] by the boss that I should finish [familiar] that work by tomorrow'

The adjustment of the indexicals (*asatte* ('day after tomorrow') to *ashita* ('tomorrow'); *kono* ('that') to *sono* ('this')) clearly indicate indirect speech. Discussing a similar example, Oshima further defends the indirect analysis on the basis of wh-movement:

Except for the imperative form, what Kuno calls blended discourse has all the characteristics of indirect discourse. For example, a wh-phrase in a 'quasi-direct quote' can take matrix scope:

(18) Taro$_i$-wa [yatu$_i$-no uti-ni nanzi-ni
Taro-Top he-Gen house-Dat what.time-Dat
ko-i] to it-ta no ka?
come-Imp Comp say-Past Q Q
'What time did Taro$_i$ say, [come to his$_i$ house]?'

(Oshima 2006:13)

On the other hand, the familiar imperative form *yare!* ('finish/do! [familiar]') in (17) has an addressee-oriented honorific and performative force that suggests direct speech. To see this, note first that in Japanese, as in English, imperatives don't seem to be embeddable under clearly marked indirect reports at all (* *The boss said that finish that work!*). Nonetheless, it has been argued that imperatives *can* be used in indirect speech, both in English (Crnič and Trinh 2009) and in Japanese (Schwager 2005). For a full discussion I refer to Maier (2010), but for now the central point is this: Even if imperatives could be embedded syntactically, we'd still have to explain the semantic shifting of all the addressee-oriented aspects in *yare*. Both the performative aspect and the honorific aspect are naturally analyzed as addressee-oriented: *yare!* $\approx$ I command you to do it, and I in fact am entitled to address you in a familiar manner. However, in this case, the main verb's *-mashita* shows that I take a rather formal, polite stance to you. Moreover, it's clear from the context that it's the boss commanding me, rather than me you.

Now, it may be technically possible to devise a system that allows indirect discourse to shift the relevant addressees and evaluative judges for the examples in (15) en (17). For the imperative, we can add some shiftable context parameters to the semantics of the imperative form (Schwager 2005). I submit that such a move is *ad hoc* rather than explanatory. Shifting phenomena in reports are really pervasive, so a more general shiftin g or mixing mechanism would be desirable. As noted above, for instance, it's not just the imperative force that is shifted in (17), the honorific marking of *yare* is also shifted. Moreover, Oshima himself provides two more classes of speaker/addressee-oriented expressions that retain their original form inside an otherwise indirect report: deictic predicates and empathy-loaded expressions.

As an example of a deictic predicate, take *iku* 'go', indicating movement away from the context's speaker:

(19) kinoo, Matsushima-kun-wa [kyoo boku-no uti-ni ik-u]
yesterday Matsushima-Top today I-Gen home-Dat go-Pres
to it-ta
Comp say-Past
'Yesterday, Matsushima said that he would go to my home today.'
(Oshima 2006:15)

As the reported movement is toward the speaker's own house, we'd expect *kuru* ('come'), so again, we're dealing with an apparent perspective shift here.

As an example of an empathy-loaded expression, finally, take *yaru* 'give', indicating the speaker empathizes more with the giver than with the receiver:

(20) kinoo, Matsushima-kun-wa boku-ni [kyoo boku-ni purezento-o
yesterday Matsushima-Top I-Dat today I-Dat gift-Acc
yaru] to itta
give-Pres Comp say-Past
'Yesterday, Matsushima said to me that he would give me a gift today.'

[(Oshima 2006, p. 16)]

Yet again, we have an indexical, speaker-oriented expression embedded in an indirect report, and interpreted with respect to the reported rather than the actual speech context. After presenting the semantics mixed quotation, I show that similar perspective shifting has been discussed already in English, before returning to the Japanese examples of this section.

5 Towards a Unified Analysis: Mixed Quotation in Speech Reporting

The problem separatists have in dealing with the examples above is an apparent shifting and mixing of perspectives in indirect speech. There are ways to deal with such indirect shifting, but they involve a substantial overhaul of the semantics of indirect speech reporting or of indexicality/addressee-orientation (cf. Schlenker's analyses of indexical shifting in Amharic). I claim that we need not go there; we already have everything we need with (i) Kaplan's (1989) classic semantics of indexicals and indirect speech and (ii) an account of mixed quotation. Both of these mechanisms are independently motivated and relatively uncontroversial, but the second one may need some explanation.

5.1 The Presuppositional Account of Mixed Quotation

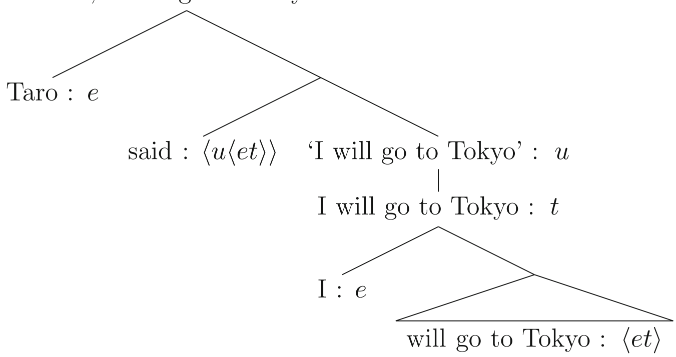
My preferred analysis of mixed quotation is Geurts and Maier's (2003) presuppositional account. In that framework, quotation marks trigger the presupposition that someone used precisely the words mentioned within them (necessitating an underlying mention logic, as developed in 3.2 already) to express something, while that underspecified something is left embedded in an indirect report, as in (21) (the ∂ symbol marks a presupposition trigger).

(21) Quine says that quotation "has a certain anomalous feature"
$\rightsquigarrow$ Quine said that quotation has ∂[the property he expresses with the phrase 'has a certain anomalous feature']

First, the syntax. Mixed quotation will be modeled as a unary operation, like pure quotation, but now retaining the type of its daughter node rather than turning it into u. This reflects a key feature of mixed quotation, described already by Davidson (1979): a mixed quoted phrase fills the same syntactic/semantic slot in the sentence as it would without quotation marks ("syntactic incorporation", cf. Maier (2008)). I'll use double quotes for mixed quotation, single for pure:

(22)

a. Taro said, "I will go to Tokyo" : t



b. $\mathbb{T}$((14a)) = $\boxed{\texttt{say}_{dd}\texttt{(t,}\ulcorner\text{I will go to Tokyo : } t\urcorner\texttt{)}}$

c. $[\![(14b)]\!]^c$ = the set of worlds in which Taro utters "I will go to Tokyo : t"

Now the semantics. Following Geurts and Maier (2003), mixed quoting triggers a presupposition. To formalize the presupposition *the property he expresses as 'has a certain anomalous feature'*, we need not only a device to mention the quoted expression, but also a representation of an "express" relation that "defers" its interpretation to its source. In our formal language, let's say:

(23) a. If $\alpha$is of type e, υa quote (type u), and ψ an expression of arbitrary type τ, then $\texttt{express}(\alpha, \upsilon, \psi)$is a term of type t, and

b. $[\![\texttt{express}(\alpha, \upsilon, \psi)]\!]^c$ = the set of worlds where $[\![\alpha]\!]^c$ uses $[\![\upsilon]\!]^c$to express $[\![\psi]\!]^c$

We can use this `express` to reformulate the presupposition mentioned above as 'the `X` such that $\texttt{express}(\texttt{x}, \ulcorner\text{has a certain anomalous feature : } et\urcorner, \texttt{X})$'. Now, by syntactic incorporation, we want the presupposed `X` to have the type of the quoted expression, i.e. $\langle et\rangle$. In Van der Sandt's (1992) version of DRT, where a presupposition looks like a dashed box with a non-empty top compartment (universe) containing the presupposed variables:

(24) kare wa watashi ga matta machigaimashita to iimashita
he Top I Nom again was.wrong-Polite Comp said-Polite

a. 'He said [polite]: "I was wrong again [polite]"'
b. 'He said [polite] that I was wrong again [polite]' (Coulmas 1985:57)

The rule for mixed quotation triggers the introduction of such a presupposition:

(25) asatte made ni kono shigoto-o yare
day after tomorrow until this work-Acc do-Imp-Familiar
'Finish [familiar] this work in two days!'

The direct, compositional contribution of this DRS to the truth conditions is just the variable X, of type τ. The dashed presupposition box below it indicates that this X is 'what x uses the phrase σ to mean', and furthermore indicates that such X and x are to be found in the discourse context. In other words, it tells us to search the context for a source x and the property that x expresses with the quoted phrase. We'll get to this process of contextually resolving a presupposition, after we finish the DRS construction.

Above the mixed quote in the tree (22), we have a standard indirect discourse, so translation in DRT proceeds as in Sect. 3.1, yielding the preliminary DRS (26) as output (analyzing *quotation* as a term of type *e*, and *Quine* as a type *e* presupposition):

(26) ashita made ni sono shigoto-o yare to jooshi-ni
tomorrow until that work-Acc do-Imp-Fam Comp boss-by
iwaremashita
was told-Polite
'I was told [polite] by the boss that I should finish [familiar] that work by tomorrow'

We will interpret such a preliminary DRS only after resolving all its presuppositions relative to a DRS representation of the discourse context. This happens as follows. Say, we're in a context where Quine is already established as a discourse referent in the common ground. That is, before we interpret the sentence, we already have a DRS with a non-empty universe introducing an individual q named Quine (quine(q)). We merge this context DRS with our preliminary DRS and bind the presuppositions z (triggered by the proper name *Quine*) and x (triggered by the mixed quotation as the source of the quoted words) to q (the Quine from the context).

(27) Taro$_i$-wa [yatu$_i$-no uti-ni nanzi-ni
Taro-Top he-Gen house-Dat what.time-Dat
ko-i] to it-ta no ka?
come-Imp Comp say-Past Q Q
'What time did Taro$_i$ say, [come to his$_i$ house]?'

(Oshima 2006:13)

Now, the remaining presupposition searches the global context for a meaning that Quine attaches to the quoted phrase. This is indeed a rather specific entity that will not be easily found, unless we have just been talking about deviant ways in which an individual uses a certain term. Real presupposition binding in this case is out, which brings us to accommodation: unless there is evidence against Quine using the phrase at all, we simply add to the global discourse representation that he used it to refer to some property X. Formally, we merge the presupposition with the global DRS:

(28)
kinoo, Matsushima-kun-wa [kyoo boku-no uti-ni ik-u]
yesterday Matsushima-Top today I-Gen home-Dat go-Pres
to it-ta
Comp say-Past
'Yesterday, Matsushima said that he would go to my home today.'
(Oshima 2006:15)

This final result of contextual, pragmatic resolution of the grammatically introduced presuppositions, has a straightforward classical model-theoretic interpretation, which comes down to: Quine uttered "has a certain anomalous feature" to express some property X (not otherwise specified), and he says that quotation has that property.

In this way we get an account of the hybrid use/mention character of mixed quotation. The analysis suggests an extension to direct discourse, analyzing it as mixed quotation of an entire sentence (i.e. type t rather than $\langle et \rangle$). This would effectively blur the line between direct and indirect discourse. The following picture emerges: to report another's speech there is only indirect discourse, within which the device of mixed quotation can be used to mimic a particular phrase of the reported speech act verbatim.[8] Direct discourse, in this picture, is merely a limiting case of mixed quotation.

Although inspired by this general rejection of the rigid direct–indirect distinction, the concrete aim of the current chapter is somewhat more modest: I present evidence for the suggested blurring of the direct–indirect distinction in Japanese speech reporting. More specifically, I will show how to accommodate the above Japanese data using mixed quotation rather than some kind of context shifting. I will not here pursue the strong claim that all forms of classic direct discourse in Japanese and beyond are best analyzed in terms of presuppositional mixed quotation.[9]

---

[8] The reporter can have a variety of reasons for wanting to do this: he may not have understood the original words, the words may be meaningless, the reporter may be uncomfortable using the phrase, may want to liven up his whole report, may consider that phrase exceptionally well put, etc.

[9] The reason being that there is some evidence that full-fledged, classical direct discourse behaves somewhat differently from standard mixed quotation. For instance, as a reviewer points out, it's not obvious that direct discourse quotations show the projection behavior characteristic of presuppositions. One reason for this may be the fact that direct discourse is often syntactically unembedded, the say-frame attached only parenthetically (*"I did it," said John*). These parenthetically framed uses resist embedding, e.g. under negation (Recanati 2001): **"I did it," John didn't say*. So standard presupposition tests don't even apply here. The question is then, insofar as non-parenthetical direct quotations can be embedded, do the alleged quotational presuppositions project? The answer seems

5.2 From Mixed Quotation to Blended Discourse

The presuppositional semantics of mixed quotation can be and has been applied to account for some aspects of shiftiness in indirect discourse. Maier (2007), for instance, analyzes shifted indexicals like Amharic *I* as mixed quoted,[10] rather than meddling with Kaplan's semantics, and Maier (2012) posits mixed quotation in ancient Greek indirect-to-direct *fade-in* and so-called recitative *hoti* (≈'that'+direct speech).

Closer to the current data set is Anand's (2007) suggestion to treat apparently shifted expressives like *that bastard* in (29) as mixed quoted (as it's the father, rather than the actual speaker who thinks Webster is a bastard):

(29) My father screamed that he would never allow me to marry that bastard Webster [Kratzer (1999)]
⇝ My father screamed that he would never allow me to marry "that bastard Webster" [≈(Anand 2007)]
⇝ My father screamed that he would never allow me to marry ∂[the individual he refers to as 'that bastard Webster']

Anand argues that the quotational shift analysis of 'non-speaker-oriented expressives' is empirically superior to Potts' (2007b) analysis that meddles with Kaplan's contexts by adding a 'expressive judge' parameter.

I claim that in all these cases of shiftiness in reports the mixed quotation analysis is simpler and more compatible with tried and tested semantic theory than the alternative: Schlenker's context shifting monsters, which overturn Kaplan's famous prohibition thereof and even threaten the notion of rigidity,[11], and/or the *ad hoc* addition of shiftable expressive judges to the utterance context (cf. Anand 2007).

Note that these cases, like the Japanese ones discussed here, lack overtly realized mixed quotation marking. This can be no counterargument against the presence of mixed quotation in the logical form, as we have already seen that overt quotation marking may be absent even in full-blown direct discourse, in colloquial spoken Japanese at least. Moreover, at the subclausal level we also find naming constructions where overt (pure) quotation marks are lacking consistently, even in writing:

(Footnote 9 continued)
to be no: *John didn't say, "I did it"* obviously doesn't presuppose that John uttered the words 'I did it'. Although my hunch is that, in contrast to the indirect version *John didn't say that he did it*, the direct version does come with a presupposition that *someone* said those words. Another possible difference, also noted by the reviewer, is that direct quotations sometimes behave like pure quotes, for instance in allowing totally ungrammatical, even non-linguistic material to be quoted. I leave a further investigation of these facts and the required theoretical adjustments for future research.

[10] However, Anand (2006) objects that, in Amharic at least, there are syntactic restrictions on shifting that are not straightforwardly captured by the mixed quotation approach. Other languages, such as Catalan Sign Language (Quer 2005), diagnosed with shifting pronouns seem to violate these constraints and may still be amenable to a mixed quotational treatment.

[11] Schlenker argues that his system upholds Kaplan's fundamental distinction between rigid/directly referential and descriptive terms, but this is not altogether clear—and much less so in e.g. Von Stechow's (2002) related account where shifted indexicals correspond to mere bound variables.

(30) My name is Emar

Nonetheless, it's worth pointing out that the theory does place constraints on the scope of mixed quotation. As it stands, only linguistic strings with a semantic type can be quoted. In other words, we predict that the quoted phrase must make a compositional contribution to the truth-conditions. This does not necessarily restrict us to grammatical, meaningful constituents, though. I've mentioned the idiolectic 'misunderestimate', which is arguably a constituent (it has a syntactic/semantic type in *G*) although it need not have a semantic interpretation in our system. With some extensions to the basic framework, we can also allow for subconstituent quotations such as *John said the stalag"mites" were falling down* (cf. Maier 2007) and superconstituent quotations such as *Mary said the dog ate "strange things, when left to its own devices"* (from Abbott (2003), analyzed with the help of the additional principle of 'quote-breaking' by Maier (2008) and 'unquotation' by Shan (2011)).

6 Blended Discourse in Japanese

My analysis of the Japanese data is now easily stated. The examples in 4.2 appear to mix direct and indirect discourse because they do; they are indirect reports with a mixed quoted phrase. Let's go through our examples.

6.1 Quoting Honorifics

The intended 'indirect' reading (15b) of (15), the report with the embedded *-masu* politeness form, corresponds to a logical form where that form (and perhaps some more, but not the first person pronoun) is mixed quoted:

(31)
kare wa watashi ga "matta machigaimashita" to
he Top I Nom again was.wrong-Polite Comp
iimashita
said-Polite
≈ 'He said that I was "wrong again"'[12] cf. (15)
⇝ he said [polite] that I was ∂[what he expressed as 'wrong again [polite]']

Note again that this involves quotation marks that are invisible on the surface. I actually share the assumption of invisible quotes with direct–indirect separatists like Coulmas, who appeals to them to get the other reading, (15a). For us, that so-called direct discourse reading brings nothing new, the only difference with (31) is that the first person pronoun is now also part of the mixed quote, which presumably then covers the whole clause:

(32) He said [polite] that "I was wrong again [polite]"

6.2 Quoting Imperatives

The next two examples, (17) and (18), feature (invisible) mixed quoted imperatives:

(33) ashita made ni sono shigoto-o "yare" to jooshi-ni
tomorrow until that work-Acc do-Imp Comp boss-by
iwaremashita
was told-Polite
≈ 'I was told by the boss that I should "finish!" that work by tomorrow' cf. (17)

⇝ I was told [polite] by the boss to ∂[what he expressed as "finish! [familiar]"] that work by tomorrow

The quotation marks here correctly defer the impolite imperative force to the reported speaker, the boss. However, this is not the final story yet. There are two major problems with the analysis sketched in (33): (i) What is the semantics (more specifically, what is the semantic type) of an imperative? This is still a hotly debated issue, but whatever the exact semantic objects involved, (ii) it's probably something that attaches to an entire imperative clause, not just the verb without its direct object, as required here. Although I cannot present these matters in any more detail here, Maier (2010) addresses precisely these concerns, using Schwager's (2005) semantics of (unembedded) imperatives and Shan's (2011) unquotation.

6.3 Empathy-Loaded and Deictic Predicates

To get a fully unified account of shifting through mixed quotation, the logical form of (19), finally, requires mixed quotation of *iku* ('go'), which yields the interpretation *Matsushima said he would do ∂[what he referred to as 'go'] to my house*. And similarly for (20). Of course, on the basis of only these particular examples we cannot discard the possibility that *iku* and *yaru* are simply descriptive terms that can be freely shifted by binding to any salient discourse entity. The empirical cross-linguistic diversity is rather subtle, and beyond the scope of this chapter. I'm simply assuming some kind of true, indexical context-orientation.

7 Conclusions

I have provided a principled account of shifting without complicating our contexts or the semantics of indexicals and reports. We have essentially given up the categorical direct–indirect distinction. In fact, we have given up the whole notion of direct discourse: speech reporting follows Kaplan's semantics of indirect discourse except for the parts (in some cases the whole clause, or more) that are mixed quoted. These quoted parts are automatically (by presupposition resolution) deferred to the reported speaker. For Japanese in particular, this means we can keep the intuitive analysis of speaker/addressee-oriented expressions like imperatives, honorifics, and deictic verbs, as indexicals. In other words, I agree with Hirose that in reported speech "addressee-oriented expressions are, by definition, used only as public expressions [= direct discourse/quotation]", but what I reject is the, often implicit, assumption that "phrases and sentences containing addressee-oriented expressions are also addressee-oriented, functioning as public expression [= direct discourse]" (Hirose 1995, p. 227).

References

Abbott, B. (2003). Some notes on quotation. *Belgian Journal of Linguistics*, *17*(1), 13–26.

Anand, P. (2006). *De De Se*. Ph.D. thesis, MIT.

Anand, P. (2007). Re-expressing judgment. Theoretical Linguistics, *33*, 199208.

Anand, P., & Nevins, A. (2004). Shifty operators in changing contexts. In *Proceedings of SALT XIV* (pp. 1–18). Ithaca, NY: CLC Publications.

Banfield, A. (1973). Narrative style and the grammar of direct and indirect speech. *Foundations of Language*, *10*(1), 1–39.

Cappelen, H., & Lepore, E. (1997). Varieties of quotation. *Mind*, *106*, 429–50.

Coulmas, F. (1985). Direct and indirect speech: general problems and problems of japanese. *Journal of Pragmatics*, *9*, 41–63.

Crnič, L., & Trinh, T. (2009). Embedding imperatives in english. In A. Riester & T. Solstadt (Eds.), *Proceedings of Sinn und Bedeutung 13* (Number 1), Stuttgart.

Davidson, D. (1979). Quotation. *Theory and Decision*, *11*, 27–40.

Geurts, B., & Maier, E. (2003). Quotation in context. *Belgian Journal of Linguistics*, *17*(1), 109–128.

Hirose, Y. (1995). Direct and indirect speech as quotations of public and private expression. *Lingua*, *95*, 223–238.

Jansen, W., Gregory, M., & Brenier, J. (2001). Prosodic correlates of directly reported speech: evidence from conversational speech. In *Proceedings of the ISCA Workshop on Prosody in Speech Recognition and Understanding* (pp. 77–80), Citeseer.

Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry & H. Wettstein (Eds.), *Themes from kaplan* (pp. 481–614). Oxford: Oxford University Press.

Klewitz, G., & Couper-Kuhlen, E. (1999). Quote—unquote? the role of prosody in the contextualization of reported speech sequences. *Interaction and Linguistic Structures (InLiSt)*, *12*, 1–34.

Kratzer, A. (1999). Beyond 'ouch' and 'oops': how descriptive and expressive meaning interact. 'Unpublished Ms., UMass, AmherstMA.

Kuno, S. S. (1988). Blended quasi-direct discourse in Japanese. In W. Poser (Ed.), *Papers from the Second International Workshop on Japanese Syntax* (pp. 75–102). Stanford: CSLI.

Maier, E. (2007). Quotation marks as monsters, or the other way around? In M. Aloni, P. Dekker & F. Roelofsen (Eds.), *Proceedings of the Sixteenth Amsterdam Colloquium* (pp. 145–150). Amsterdam: ILLC.

Maier, E. (2008). Breaking Quotations. In K. Satoh, A. Inokuchi, K. Nagao & T. Kawamura (Eds.), *New frontiers in artificial intelligence* (Vol. 4914, pp. 187–200). *Lecture Notes in Computer Science*, Berlin/Heidelberg: Springer.

Maier, E. (2009). Japanese reported speech: against a direct-indirect distinction. In H. Hattori, T. Kawamura, T. Idé, M. Yokoo & Y. Murakami (Eds.), *New frontiers in artificial intelligence* (Vol. 5447, pp. 133–145). *Lecture Notes in Computer Science*, Heidelberg: Springer.

Maier, E. (2010). Quoted imperatives. In M. Prinzhorn, V. Schmitt & S. Zobel (Eds.) *Proceedings of Sinn und Bedeutung 14* (pp. 289–304), Vienna.

Maier, E. (2012). Switches between direct and indirect speech in ancient greek. *Journal of Greek Linguistics*, *12*(1), 118–139.

Oshima, D. Y. (2006). *Perspectives in reported discourse*. Ph.D. thesis, Stanford.

Partee, B. (1973). The syntax and semantics of quotation. In S. Anderson & P. Kiparsky (Eds.), *A festschrift for morris halle* (pp. 410–418). New York: Holt, Rinehart and Winston.

Potts, C. (2007a). The dimensions of quotation. In C. Barker & P. Jacobson (Eds.), *Direct compositionality* (pp. 405–431). New York: Oxford University Press.

Potts, C. (2007b). The expressive dimension. *Theoretical Linguistics*, *33*(2), 165–198.

Potts, C., & Kawahara, S. (2004). Japanese honorifics as emotive definite descriptions. In K. Watanabe & R. B. Young (Eds.), *Proceedings of the 14th Conference on Semantics and Linguistic Theory* (pp. 235–254). Ithaca, NY: CLC Publications.

Quer, J. (2005). Context shift and indexical variables in sign languages. In *Semantics and Linguistics Theory (SALT) 15* (Vol. 15, pp. 152–168), eLanguage.

Recanati, F. (2001). Open quotation. *Mind*, *110*, 637–687.

Schlenker, P. (2003). A plea for monsters. *Linguistics and Philosophy*, *26*, 29–120.

Schwager, M. (2005). *Interpreting imperatives*. Ph.D. thesis, University of Frankfurt/Main.

Shan, C.-C. (2011). The character of quotation. *Linguistics and Philosophy*, *33*(5), 417–443.

van der Sandt, R. (1992). Presupposition projection as anaphora resolution. *Journal of Semantics*, *9*(4), 333–377.

Voloshinov, V. N. (1973). *Marxism and the philosophy of language* (Vol. I). New York: Seminar Press.

von Stechow, A. (2002). Binding by verbs: Tense, person and mood under attitudes. *In Proceedings of NELS*, *33*, 379–403.

Werning, M. (2005). Right and wrong reasons for compositionality. *The Compositionality of Meaning and Content: Foundational Issues*, *1*, 285–309.

What is Evidence in Natural Language?

Elin McCready

Abstract This chapter tries to understand the proper notion of evidence to use in the semantic analysis of natural language evidentials. I review various notions of justification from the epistemological literature, and consider how they relate to the use of evidentials and related constructions. I then consider how (some) evidentials behave under Gettier scenarios. The conclusion is that the required notion of evidence is one which is weaker than (many accounts of) knowledge, involves increase of speaker credence, but which is necessarily first-person. I thus settle on a view based on a self-ascription of probability increase due to knowledge of propositions that increase credence after conditionalization.

Keywords Evidentials · Evidence · Probabilities · Justification · Epistemology · Semantics · Indexicality

1 The Problem

Evidentials have been widely studied in linguistics for many years. They can be defined roughly as expressions which indicate the basis of the claim made by a speaker: in short, the source of the evidence on which the speaker's claim is based. In early descriptive work the existence of evidentials was clearly acknowledged, though not given much concentrated attention (e.g. Sapir 1922). Subsequently, researchers have addressed themselves to complex questions about the nature of evidentials and how to define the category: for example, much attention has been given to the distinction between evidentials and miratives (e.g. DeLancey 1997) and between evidentials and epistemic modals (de Haan 1999; Matthewson et al. 2007). Most recently of all, scholars working in formal semantics and pragmatics have worked

E. McCready (✉)
Department of English ,Aoyama Gakuin University, Shibuya-ku, Tokyo, Japan
e-mail: mccready@cl.aoyama.ac.jp

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*, Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_8,

to place evidentials into the mathematical analysis of linguistic meaning; after the seminal work of Garrett (2001) and Faller (2002), a rather large literature has developed (e.g. Ogata 2005; Chung 2005; McCready and Asher 2006; McCready and Ogata 2007; Matthewson et al. 2007; Faller 2007; Davis et al. 2007; Murray 2010; Bronnikov 2011, among many others). From this work, a great deal has been learned about evidentials: for instance, that evidentials in different languages can show different properties with respect to scope and 'level of application', some evidentials act much like (epistemic) modals, while others appear more like elements which alter the speech acts performed with the sentences that include them, and that different evidentials show distinct preferences for what kinds of content they can apply to.

However, some issues remain open. One involves what hearers ought to do with evidentially marked content after processing it. Should it be updated with in the usual way? Or should it be treated more like content modified by e.g. modal operators? The vast literature on dynamic semantics tells us that the two options yield very different results (cf. Veltman 1996). Some authors have taken explicit positions on this issue; for instance, McCready and Ogata (2007) claim that Japanese evidentials behave like modals, and so update with evidentially marked content should act as global tests on hearer information states, but simultaneously introduce new discourse referents for information sources. On the other hand, Davis et al. (2007) take evidentials to modify a contextual standard for assertability, according to which update on the evidentially marked content proceeds as usual. Part of the reason for this difference in focus presumably involves the different sorts of evidential systems examined by the two chapters. But the general problem remains. In some cases, it can be shown to lead to other complex issues: McCready (2011) considers the general situation for update with hearsay evidentials in light of widely accepted pragmatic principles, showing that it is not at all simple, regardless of the evidential system in question. A more general treatment of this issue is needed.

This chapter, however, focuses on another, still more pressing problem. What is evidence, actually? Put in another way, what is the content introduced, or manipulated, by evidentials? A great deal of the formal literature on evidentiality takes the notion of evidence as a pure primitive. We often find definitions like the following one (slightly modified; emphasis mine)[1]:

(1) $[\![k'a(f)(B)(w)(\varphi)]\!]$ is only defined if for all worlds $w, w' \in B(w)$ **the inferential evidence in** w **holds in** w', and f is a choice function of type $\langle st, st\rangle$ such that $f(B(w)) \subseteq B(w)$.
If defined, $[\![k'a(f)(B)(w)(\varphi)]\!] = 1$ iff $\forall w' \in f(B(w)) : [\![\varphi]\!](w') = 1$.
(Matthewson et al. 2007)

Here, it is assumed that we can identify whatever evidence is being presupposed to exist. But this is not a trivial task. Compare a standard case of presupposition like (2): here there are no problems finding out whether the presupposition is or is not satisfied. We merely check for the existence of a King of France.

[1]The use of the choice function here is meant to account for quantificational variability in the St'at'imcets evidentials. See Matthewson et al. (2007) for details.

(2) The King of France enjoys oatmeal with blueberries.

Life is not so easy for (1). Without a definition of evidence, how can we be sure that we have found the piece the speaker had in mind, or a substantial enough piece of evidence, or indeed any evidence at all? It seems to me that we can't. This, in my view, constitutes a serious problem for (most) current theories of evidentiality.[2]

The issue becomes even more worrying when one considers accounts of evidentiality like that of Chung (2005), for whom evidence is (as I read things) something to be found in the external physical environment. She defines a notion *v-trace* along the lines of the 'temporal trace function' of Krifka (1992; further modified by Faller 2004 to a function on spatiotemporal traces for use in the analysis of a particular kind of past tense marker). Such a function is sensible: we know what it is for an event or an individual to occupy (space) time, and we know how to identify the spatiotemporal region that event or individual has occupied. But consider the *v-trace* function, defined as (3) in the version of Chung (2007).

(3) $v\text{-}trace(e) = \{\langle t, l\rangle | \exists v[\mathit{EVIDENCE\text{-}FOR}(v, e) \wedge AT(v, t, l)]\}$, where $AT(v, t, l)$ is true iff the evidence v for the occurrence of the eventuality e appears at a location l at time t.

So the 'evidential trace' of an event corresponds to the spatiotemporal coordinates of all evidence for that event. Clearly, it is difficult or impossible to identify such evidence without knowing what evidence is meant to be. The essential difference between the cases is that, unlike the King of France, the notion of evidence is not epistemologically innocent. There are many assumptions lurking in the background, and, without clarifying those assumptions, it does not seem to me that we can claim to have an actual theory.

Further, I think that leaving crucial terms undefined leads to strange predictions in our theories. Considering (3) again, what happens in the case of inferential evidence? Suppose that I see you at home, and I know that you ordinarily stay home if it is raining, and otherwise don't. Together this is enough to infer that it must be raining now. But what is the spatiotemporal trace of this evidence? It seems odd to say that it is the same as the trace of your being at home, for my background knowledge also comes into play. But where is the spatiotemporal trace of this knowledge? The obvious answers are highly counterintuitive, and I doubt that anyone would really want to accept them; further, we are already into dubious philosophical territory with the assumption that evidence corresponds to *events*. While this might make sense for many kinds of evidence, such as the observable fact that you are at home, an existentially quantified statement about an event, it makes much less sense for propositions such as the second premise above, that you ordinarily are at home only if it is raining: how can one find the spatiotemporal coordinates of a generic statement? Clearly, the lack of a proper definition of evidence is leading to some pernicious

[2] A reviewer observes that similar worries arise for theories of modality that assume an ordering source or modal base (e.g. Kratzer 1981): without a definition of priorities, preferences, or laws, how can we have a completely satisfying theory of deontic modality?

consequences at this point, which I believe can be avoided by being more explicit about what evidence is.

So far my argument has been limited mostly to what might be described as ontological qualms. One might respond: why don't we just leave it to the epistemologists? Why do linguists have to worry about this issue? The answer is: we don't, if we are willing to let the epistemologists define for us how our theories are supposed to work. More precisely, there is no obvious reason to assume that the notions of evidence relevant to work in epistemology or philosophy of science are identical to those we require for doing semantics. Why should natural language care about scientific evidence? There seems no a priori reason why it should, and perhaps many reasons why we should expect that a different kind of evidence is at issue. The goal of this chapter is therefore to find the right notion of evidence for work on evidentiality. This task is of independent interest beyond the task of, as it were, picking up after the formal semanticists: I hope to show that we learn something substantial about the nature of evidentiality in natural language from the attempt.

The structure of the chapter is as follows. In Sect. 2 I will lay out some background relating to the main topic of the chapter: first, I will summarize some existing positions in the linguistics literature about the nature of evidence (some implicit in the formal analyses that have been proposed), and then turn to some facts that have been used to argue for particular views of evidence in natural language (McCready 2008, 2010). These facts primarily relate to two factors: the thesis that evidence must be knowledge (cf. Williamson 2000), and Gettier scenarios (Gettier 1963), broadly construed as involving the distinction between subjective and objective understandings of what counts as evidence and knowledge. Section 3 discusses some distinctions made in epistemology relevant to the issues, along with some extra necessary conditions on an account of linguistic evidence. Section 4 takes the facts adduced in Sect. 2.2. and considers various possible stories about linguistic evidence in their light. The conclusion is that natural language evidence is a concept which involves speaker beliefs about whether or not the truth of one proposition increases the probability of another in a necessarily *de se* manner. In Sect. 4, also, several phenomena are exhibited which are of interest with respect to the proposal, involving externalist versus internalist concepts of evidence and true causality. I show that in each case some light is shed by the proposed framework. Section 5 concludes the chapter.

2 Subjectivity, Skepticism, and Gettier Scenarios

In this section I examine accounts of evidence from the linguistic literature and show that they leave certain questions unanswered. I focus on the account of McCready and Ogata (2007), who give the most explicit story of which I am aware; looking at other available analyses, I show that even the most (formally) explicit have as their main goal explaining the assertability conditions of evidential sentences, taking the notion of evidence essentially for granted. This is done in 2.1. Section 2.2 discusses some data that bear on the kind of evidence used in evidentials having to do with the use of evidential sentences in skeptical and Gettier scenarios.

2.1 Implicit Accounts of Evidence

What does evidence do? The obvious answer is that it provides justification for certain beliefs. One way to think about this justification is by means of changes in the probabilities assigned to the content of those beliefs. Those authors working on evidentiality who are explicit about the concept of evidence all take this perspective, as far as I know. Here I will briefly review the most explicit theory I am aware of in linguistics, that of McCready and Ogata (2007). I will then indicate what points are left untouched, underspecified, or insufficiently argued for by these authors.

McCready and Ogata (2007) defined the function of evidence in (roughly) a Bayesian manner. This paper provides an analysis of certain Japanese evidentials, of two types: inferential evidentials, and hearsay evidentials.[3] The inferential evidentials were modeled using an operator $\triangle_a^i$, where i indexes an evidence source and a is an agent, whose effect can be stated informally as follows:

(4) $\triangle_a^i\phi$ is true given a world w, time s, agent a and probability function μ iff:

  a. ϕ was less likely according to a as determined by μ at some time preceding s (before introduction of some piece of evidence i),
  b. ϕ is still not completely certain for a at s (given i), and
  c. the probability of ϕ for a never decreased between the time a became aware of the evidence i and s as a result of the same piece of evidence i (i.e., the probability of ϕ given i is upward monotonic).

Even more informally, then, observation of i made ϕ likely but not certain, and i never subsequently had the effect of lowering the likelihood of ϕ. The observation of the evidence itself was modeled with a predicate E. This predicate also serves a complex function. Informally, $\mathsf{E}_a^i\varphi$ (i) changes the probabilities assigned to every proposition ψ (excluding φ itself) in the current information state σ by replacing them with the conditional probability of ψ given φ, if defined, and (ii) replaces the modal accessibility relation with one restricted to worlds in which φ holds. This account is meant as a treatment of what evidence does in a context; it changes the probability of other propositions that are related to it, and revises the set of accessible

[3] I will not have much to say about hearsay evidentials here; McCready and Ogata (2007) analyzed them as simple tests for the existence of an 'observation sentence' in the Quinean sense (Quine 1960) of hearsay type, with content identical to that in the scope of the evidential. See McCready (2011) for some additional discussion.

possibilities to one containing only those possibilities that make the content of the evidence true.[4]

In the definitions of these two operators, we have an implicit account of what a piece of evidence does: it alters the probability of certain propositions holding true, and also induces changes in the general epistemic space of the agent acquiring the evidence. These aspects seem uncontroversial, at least in an intuitive sense. A number of questions remain unanswered, however. We might ask: is mere increase in subjective probability sufficient for use of evidentials? Whose subjective probability function is at issue? Need all evidence be propositional, as the account implies? Are there no other requirements on evidence than that it monotonically increases the probability of the evidenced? The further question of how to give a general identification of the propositions which, by this definition, count as evidence (those which raise the probability of whatever we are interested in) I will have to put aside in this chapter, though it might be statable in terms of defeasible reasoning: evidence talk is statable in terms of defeasible reasoning : ψ is evidence for ϕ iff $\psi > \phi$ as normality conditional (Pollock and Cruz 1999). Obviously, the connections between reasoning in terms of probabilities and using defeasible inference go deep (Halpern 2003). I will not explore this connection here, but it isn't obvious to me how distinct the two kinds of theory really are in terms of empirical predictions (as opposed to philosophical commitments).

Let us now consider the theory of Davis et al. (2007). This theory, like those discussed in the first section, takes evidence to be a primitive; however, it is quite explicit about what the evidential itself is meant to do with respect to justificational requirements for assertion of sentences including evidentials. On the natural assumption that evidence ties directly to justification, the theory therefore makes indirect claims about what evidence *does* (as opposed to what it *is*).

What are the justificational norms of assertion? This topic remains controversial in philosophy. Some possibilities: assertions might require simple belief of the speaker about the asserted proposition (Bach and Harnish 1979), they might require knowledge (Williamson 2000), or they might require something weaker.[5] Davis et al. (2007) assume a version of the last view on which the context makes available an assertability threshold; this is a real number in [0,1] corresponding to a probability. Individuals are associated with (subjective) probability functions. If the probability the speaker assigns to φ is higher than the threshold, then φ is assertable for that speaker. The basic strategy is borrowed from Lewis (1980); the realization here is similar to work on gradable adjectives by Kennedy (2007) and others.

What is the function of evidentials in this theory? According to Davis et al. (2007), evidentials serve to shift the assertability threshold. Direct evidentials leave the threshold unchanged; hearsay evidentials and other kinds of evidentials marking

[4] McCready and Ogata (2007) did this by revising the accessibility relation; it could also be done by revising the set of available possibilities *qua* set, as was done in the analysis of modality of Asher and McCready (2007).

[5] See the chapters in Brown and Herman (2011) for extensive and interesting recent discussion of this issue.

indirect information sources lower it. The upshot is that evidentials can strengthen or weaken justificational requirements on assertion; this part of their meaning can be characterized as pragmatic. Semantically, they simply 'commit the speaker to' the existence of evidence of the relevant type. Exactly what this commitment amounts to is not clear (presupposition? assertion? conventional implicature?), but this is perhaps as desired by the authors, since these aspects of evidential meaning seem to vary from language to language, and indeed from evidential to evidential within a single language. With this move, the authors avoid the central question of this chapter, as they take evidence to be primitive. However, their proposal is of interest as it concerns directly the question (raised in the introduction) of how evidential information is to be processed. Fuller discussion can be found in McCready (2011).

One can find other notions of evidence in the literature on evidentials that take more controversial philosophical positions. For instance, Bronnikov (2011) provides a general model of cognition and inference; this model assumes a version of the representational theory of mind (cf. Fodor 1987, 2000) and characterizes semantic content based on its behavior with respect to the proof theory of such representations. Evidentials then are dealt with in one of two ways: either (a) as existence checks on the statement of the evidence-relevant content in the (representational) mental model, or (b) as attempts to prove that the sentence representations are present in distinct 'boxes' corresponding to different means of acquiring information. In effect, there is no longer a need for a separate theory of evidence, only for a theory of how bits of content behave in the proof theory of mental representations: a fairly radical move.Whether this looks desirable will depend on one's attitudes toward the representational theory of mind in general, and how exactly one spells out the proof theory of evidence, something about which Bronnikov (2011) is not very specific. For present purposes, the proposal is not spelled out enough to be viewed as a proposal about the nature of the evidence manipulated by evidentials.

2.2 Some Relevant Facts

In this section I would like to exhibit some data that speaks to the kind of evidence required for use of natural language evidentials. This data is of two types. The first involves situations where an agent alters subjective probabilities based on purely mental phenomena. It shows that simple subjective probability increase is not enough to ground use of evidentials. The second type involves calling into question the justificatory base itself in various ways. It shows that the putative evidence must be taken by the evaluator to properly match the external world in order to be a basis for use of evidentials.

2.2.1 Subjective Probability is Not Enough

So far we have seen views on which evidence corresponds to changes in subjective probability. An obvious question that arises with respect to such views is this: do all changes in probabilities follow from observation of evidence? Clearly not: consider, for instance, a situation in which Jerome hopes to get a particular job, and, although he is unqualified for it, allows his hope to influence his perceived likelihood of employment to the degree that it affects his actions. Another instance, even more irrational, involves a person who lets her desire to win big slowly raise the likelihood she assigns to winning the lottery to a point where she actually believes that she will win, after her purchase of one ticket.[6] Such cases easily lead to irrational behavior.[7] One response to this sort of case is to identify evidence with knowledge: one certainly would not like to say that Jerome knows he will get the job, though he may believe it with a high degree of credence. This is the $E = K$ thesis of Williamson (2000). This means that, if evidentials do indeed look for evidence in the usual epistemological sense, the relevant evidence must be known by the speaker. This position was taken by McCready (2008, 2010). It is supported by the data adduced in the next section, at least on one interpretation. However, $E = K$ by itself means little without an account of what knowledge *is*, itself a contentious issue; as we will see, given the facts about use of evidentials, $E = K$, while it may be correct, does not do much to advance our understanding of the issue.

2.2.2 Evidentials in Skeptical Scenarios

How to tell knowledge from belief? Clearly, for the truth of both $B_s\phi$ and $K_s\phi$ the speaker must assign an extremely high degree of credence to ϕ.[8] Here is a traditional answer from epistemology: knowledge is justified true belief. I can be said to know p if I believe p, p is true, and I have good reason to believe p. This answer looks reasonable, and many people have espoused some version of it. But it is wrong, as the epistemologists well know. Gettier (1963) discovered examples in which all the conditions above are met, but still there is no knowledge. Here is a scenario in the Gettier style. Johnny is traveling in the country when he sees what looks to him like a

---

[6]These cases are modeled after an example in Williamson (2000).

[7]See Fantl and McGrath (2009) for extensive discussion of the relation between knowledge and practical reason, as well as Hawthorne (2004).

[8]I assume a probabilistic account for consistency with the previous (Bayesian) discussion; one could also transpose this view to a more linguistically-traditional possible worlds-based picture of attitudes, so that

$$[\![B_a\phi]\!] = 1 \text{ iff } \frac{card(\{w' : R_a(w, w') \;\&\; \phi(w')\})}{card(\{w' : R_a(w, w')\})} \succ s,$$

i.e. the proportion of the agent's epistemically accessible worlds verifying ϕ exceeds some contextually specified degree. The requirement for knowledge would then be to further increase the required proportion (to something approximating 1) or to add extra conditions, as in the main text below. I do not see much to choose between the two pictures, at least for the present application.

horse on top of a hill and hear a horse neigh. However, what he sees is a horse-shaped rock, and the neigh is just the wind whistling through that pipe over there. But there is—coincidentally—a horse standing behind the rock. Now consider this sentence:

(5) Johnny knows there is a horse on top of the hill.

This statement seems false–though the conditions listed are satisfied: Johnny believes that there is a horse on top of the hill, there is in fact a horse there, and Johnny has good reason—in fact two good reasons—to believe there is one there, at least from his own perspective.

The above considerations suggest a way to distinguish knowledge from belief: if one can undermine the justification for the putative piece of knowledge, yet there is no change in the (subjective) cognitive status of the object of the attitude, then it is belief.[9] If the cognitive status of the putative knowledge changes together with the change in justification—if it becomes uncertain or eliminated—then the putative knowledge is knowledge indeed. Obviously, this won't work for every sort of knowledge (for instance, it has nothing to say about a priori knowledge, by definition). Still, this way of distinguishing things seems a good first approximation, and we will see the result of working with this assumption in this section. It is a bit too simple, though: I have obviously left out the problem of determining what the relevant justification is supposed to be, and for what agent or agents its status changes. As we'll see, both these points turn out to be crucial for the full story. But this characterization allows us to proceed to get some initial data to go ahead with.[10]

The strategy, then, is to call into question the justification for the evidence. We will first use the most extreme form of this general strategy: the *skeptical argument*. Skeptical arguments call into question the foundations of all our knowledge (for some given area). They have the following general form: one introduces possibilities which falsify all—or some relevant portion of—our putative knowledge and cannot be conclusively eliminated. Because we cannot eliminate them, possible flaws in the foundations of our knowledge enter our awareness. In view of these potential errors, we become uncertain about the solidity of our knowledge. As a result, our

---

[9]Here, I intend a situation in which the justification is undermined but the attitude holder remains willing to assent to the proposition, or willing to act as if it were true, on the assumption that guiding action is a primary role of knowledge as claimed by Fantl and McGrath (2009). The precise characterization of this undermining is nontrivial; in what follows, I will talk about knowledge being destroyed or eliminated by the undermining of evidence, but, as a reviewer notes, given that the basis for assuming that knowledge is eliminated is an unwillingness to assent to a knowledge attribution or to use the term 'know', the possibility that there is only a change in linguistic behavior cannot be eliminated. This observation raises the question of the relationship between linguistic evidence and epistemological conclusions, a highly fraught issue which I cannot address in the present chapter.

[10]The linguistically minded reader may now be wondering why we need to go to all this trouble. After all, isn't knowledge factive, and belief not? That means that the object of knowledge is presupposed, but not so in the case of belief. If this is so, then why must we worry about justification and the foundations of knowledge? There is some initial plausibility to this objection, but it rests on a confusion. The verbs *know* and *believe* are factive and not factive respectively, but here we are not interested in knowledge or belief as it is linguistically expressed. Rather, we are interested in evidence, as the object required for the felicitous use of evidentials. This content is not explicitly expressed in language. To find out its properties, we must take a more indirect route.

knowledge disappears. One can think of this effect in various ways, for instance as a change in the contextual standards for knowledge attribution (e.g. DeRose 1992; Lewis 1996; DeRose and Warfield 1999 provides an overview of other possibilities; general background on epistemological stances can be found in Pollock and Cruz 1999). For our purposes we need not take a stand on which, if any, of the currently available positions is correct.

Skeptical scenarios usually look highly implausible. Some traditional examples include the possibility that you might be deceived by an evil demon into believing that you are receiving certain kinds of perceptual input, such as that you are drinking a cup of coffee; that you might be a brain in a vat, with your perceptual centers stimulated by electric impulses, or in a Matrix-like situation; that you might be dreaming everything you are perceiving, or be in a catatonic state. The common characteristic to these scenarios is that, in each, the sensory data you receive is not trustworthy as a guide to what actually is. Note the similarity to the Gettier scenarios. The difference between skeptical scenarios and Gettier cases is that, in the skeptical scenarios, there is no possibility for the individual in the scenario to either learn that he is in fact in such a scenario or to conclusively prove that he is not in one, because all of his sensory input is open to question, while in the Gettier cases, a world-internal observer could make the Gettiered individual aware of his error. To anticipate the later discussion, this distinction turns out to play a role in the use and evaluation of evidentials.

We can also find scenarios that look more common-sensical, especially when we confine ourselves to scenarios that only cast doubt on certain types of knowledge or knowledge in certain domains. For instance, consider a scenario on which you fell down a moment ago and hit your head, and the resulting damage caused you to hallucinate your current state—you appear to be reading this chapter, but in fact you are lying on the floor outside your office viewing an internal projection of what you had planned to do before your injury. This situation seems quite normal compared to those above, but only calls into question your knowledge of your present activities, rather than of your entire set of memories.

What all skeptical scenarios have in common is the property that—if taken seriously—they destroy knowledge. For any p that one putatively knows (or for any salient p, for limited skeptical scenarios like the above), one may retain the belief that p but this belief can no longer be conclusive. There is always a possibility of error. Such beliefs are thus no longer knowledge in the strict sense. As a result, skeptical arguments can be viewed as tests for knowledge, when used on susceptible speakers. By running a skeptical argument on someone who is willing to consider them seriously, one can test whether a particular bit of their cognitive state is knowledge or belief, in the following sense: if the skeptical argument has no effect on the cognitive status of the content of interest, that content is merely believed.

To believe something, one must assign it a degree of subjective probability higher than whatever the threshold for belief is taken to be. In general, this threshold is contextually determined in the usual way familiar from degree predicates (cf. Kennedy 2007 on degree predicates and Stanley 2005 on belief in particular); skeptical arguments are implausible enough that they will not (barring an extremely high contextual standard) rule out belief, for they lessen degrees of subjective probability in a very minor way. Thus beliefs can survive skeptical scenarios, but knowledge cannot. We thus have a way to distinguish knowledge from belief. One application of this tool, the one I am concerned with in this chapter, is in determining whether mere belief is acceptable for the use of evidentials. I now turn to this application.

How can one use skeptical arguments in the desired way? The idea is straightforward. First, give a speaker a piece of evidence supporting some conclusion φ in the intuitive sense. As before, I temporarily sidestep issues concerning exactly what should count as evidence for some conclusion; I will discuss some of these issues later. After providing the evidence, ask whether $Evid(\varphi)$ is true (or assertable, depending on the language). This step ensures that the piece of evidence is the right kind to license the evidential in general. Here is an application of this test to the case of the Japanese evidentials discussed in the last section. Recall that such evidentials require (a) the existence of evidence of a certain type and (b) a certain degree of credence in their prejacent propositions for felicity. Under ordinary circumstances, the observation that the street is wet outside in the morning leads to a rise in the probability that it rained the night before. So, by the definitions above, it should count as evidence, and be sufficient to license the inferential evidential *mitai*.

(6) michi-ga nureteiru. kinoo-no ban ame-ga futta mitai
street-Nom wet yesterday-Gen night rain-Nom fell INF.EVID

'The street is wet. It must$_{\text{inf}}$ have rained last night.' (spoken in the morning)

(7) kinoo ame-ga futta soo-da
yesterday rain-Nom fell EVID.HRSY-Cop

'[I heard that] it rained yesterday.' (spoken after John said that it rained yesterday.)

This is correct. In this case, the evidential sentence is assertable. So the sequence $E; Evid(S)$, where E is the evidence and $Evid(S)$ the sentence containing the evidential, is a felicitous one.

The test for knowledge comes when we introduce a skeptical scenario after the evidence. Here is an English version.

(8) The street is wet. But perhaps there is no street—perhaps I am just dreaming. (Anyway,) It rained last night–Evid$_{inf}$.
(9) John said that it rained yesterday. But maybe I was just hallucinating when I saw John. (Anyway,) It rained yesterday–Evid$_{hrsy}$.

Now ask the speaker to consider the new sequence E; S; $Evid\varphi$. Could this sequence of sentences be assented to or asserted? Is this sequence acceptable, where S is the skeptical scenario and E provides the evidence on which the evidential sentence depends? Or, for languages where we can consider the evidentials primarily truth-conditional, is the sentence containing the evidential judged true in this new context? If the new sequence is acceptable, and the sentence containing the evidential is true, then the evidence required does not need to be actual knowledge: belief is sufficient. We know this because the skeptical scenario, if taken seriously, destroys knowledge; so if the evidence must be knowledge, then the sentence with the evidential would be bad. Conversely, if the new sequence is not acceptable, or the sentence with the evidential is false, then knowledge is required.[11] So that is the test. What are its results? I have tried this test on a number of Japanese speakers. A few 'skeptical' individuals were unwilling to take the skeptical arguments seriously. Disregarding these subjects, no speaker allows sentences with evidentials after the skeptical scenario is introduced. This suggests very strongly that the evidence needed for Japanese evidentials is not as simple as plain vanilla subjective belief.

Here is a possible objection to the test. It might be suggested that my informants are just balking at asserting anything about the world, given that I have called into question all their knowledge of it, and its very existence. This objection has some initial plausibility, but when examined closely, lacks force. It contains two subarguments. The first involves assertion: the unstated assumption is that, without full confidence, one cannot assert anything. This unstated assumption is false. To assert, knowledge is not necessary—we do not even need total belief. Belief beyond reasonable doubt is sufficient, where the level depends on context (again, see Stanley 2005 or Davis et al. 2007 for more discussion).[12] In any case, the objection depends on the particular skeptical scenario chosen above, which did in fact call into question everything about the world. But it is easy enough to change the scenario in such a way that we limit its application to the case at hand. Here is an instance. I give only the English version for readability.

(10) The street is wet. [But you may have a brain tumor that causes all streets to look wet, even though they are not. You cannot be sure if the street is truly wet or not.] It rained-Evid$_{inf}$.

This new scenario only calls into question the speaker's knowledge of street wetness. The rest of the world remains untouched. Nonetheless, speakers are reluctant to use evidentials in scenarios like these as well. I conclude that the apparent flaw in the test is only apparent, and that evidence for evidentials—in Japanese at least—must be knowledge.

2.2.3 Gettiered Evidentials

We are now evaluating the claim that the evidence needed for evidentials must be stronger than increase in subjective probability. When the evidence was called into

[11] An alternate version of the test, suggested by a reviewer, involves a sequence of statements in the second person, followed by a question about acceptability of the use of an evidential sentence:

- Scenario: You just went outside and saw that the street is wet. But you then wonder if maybe you are dreaming and didn't (really) see a street at all.
  In this scenario, can you truthfully(/reasonably/felicitiously) say: It rained last night–Evid$_{inf}$?

The results of this test seem the same as what we find with the version in the main text.

[12] This position is in opposition to the knowledge-norm view of Williamson (2000) and others. On this view, speakers should, normatively, only assert the things they know. I believe that this view is far too strong—it is tantamount to forcing people to have full credence in any content they assert, which is virtually impossible in the real world.

question via the application of skeptical arguments, evidentials could no longer be used (in Japanese). A natural question to ask now is whether the same thing happens in Gettier cases. This section explores the facts in this domain.

Recall one primary difference between Gettier cases and skeptical scenarios: the Gettiered individual is Gettiered because of facts about the world, whereas the victim of a skeptical argument has his knowledge destroyed precisely because the facts about the world became uncertain. The crucial point is that, while the justification the Gettiered individual (ie., in our cases, the individual with the mis-evidence) has for his beliefs is not well-founded, this lack of justification can be apparent to other individuals in the Gettier case. Thus we see that being Gettiered is a perspective-dependent problem: only the Gettiered individual is necessarily Gettiered. In skeptical scenarios, however, there is no way to determine whether the skeptical argument is true; all individuals have an equal lack of access to the 'actual' situation. (In the case of skeptical scenarios limited to cases like the brain tumor scenario above, this assessment remains valid: the individual who sees the street as not wet in (10) might equally be the victim of a brain tumor of a different kind.) This discussion suggests that we may expect to find differences between felicitous uses of evidentials in the two kinds of cases.

This expectation is fulfilled. Unsurprisingly, the Gettiered individual can assert an evidential with respect to his putative knowledge:

(11) ano oka-no ue-ni uma-ga iru mitai da
that hill-Gen top-Dat horse-Nom exists EVID Cop
'There appears to be a horse on top of that hill.' (said by the Johnny of (5))

For the outside observer the situation is a bit more complex. We can distinguish two cases involving only failure of warrant. I leave aside cases where the object of belief is in fact false.

1. The observer knows that Johnny's warrant for belief is no good, but does not know whether there is actually a horse.
2. The observer knows both that Johnny's warrant is no good and and that there is a horse.

In both of these cases, (11) is problematic. But it is problematic for different reasons. In Case 1, it is infelicitous, because of the first clause (i) of the definition of the inferential evidential; the outside observer has no piece of evidence—that is, no piece of *knowledge*, since Johnny's putative evidence is useless—that increased the probability that there is a horse on the hill to the necessary level. In Case 2, the observer runs afoul of the second clause (ii); since the observer *knows* that there is a horse, the probability she assigns to there being a horse approximates 1; she is completely certain that there is a horse, and the evidential sentence cannot be used. The situation of Case 2 is not particularly relevant for the present discussion as it involves something closer to a Gricean violation, modeled in the theory of McCready and Ogata (2007) as something akin to Veltman (1996) examples with epistemic modalities:

(12) # It is sunny . . . It might be sunny.

If we know that it is sunny, it is not helpful to assert the possibility. The evidential case is analogous.

A question immediately arises when one considers the behavior of evidentials in Gettier scenarios, concerning the distinction between assertability and truth evaluation. The issue turns on the question of whose perspective is taken, both to the (putative) evidence itself and to the evaluation of the evidential-containing sentence. Consider again (11). Is this sentence assertable? By Johnny, yes; in his Gettiered state, he believes that he has evidence enough to make it true, so he can utter it sincerely. By a non-Gettiered observer, however, it is not assertable: in Case 1, the observer knows the putative evidence to be incorrect, so the evidential is false; in Case 2, the evidential statement is just inappropriate given the observer's knowledge. So we see that the perspective taken matters for assertability. Now consider what happens when we evaluate the truth of (11). Here again, Johnny himself will take (11) to be true—as will anyone Gettiered along with Johnny—but the outside, omniscient observer will take it to be false. So perspective matters for truth evaluation as well. I will return to this issue in much more detail in Sect. 3.

2.3 Summary

In this section we have established the following. Japanese (inferential) evidentials cannot be used when skeptical scenarios call the evidence they rely on into question. They can be used by Gettiered individuals, but not by individuals who observe the Gettier scenario from an external perspective. These facts taken together suggest that the evidence required for evidentials is stronger than increase in subjective probability, and is indeed a kind of knowledge. However, we can't make it too strong: if we do, we eliminate the possibility for evidentials to be used by confused or Gettiered individuals, who clearly can use them; though such uses might be judged inappropriate or false by observers with the relevant discriminatory powers. But what other possibilities exist? To answer this question, we need to look at accounts of knowledge and justification that have been proposed in the epistemological literature. This is the task of the next section.

3 Evidence in Epistemology

The way I will approach the problem of characterizing 'evidence for evidentials' is as follows: the question of what evidence is for natural language amounts to the question of how natural language makes use of the notion of justification, i.e. the justification for assertion, and the evaluation of that justification. Thus we are interested in what notion of justification is at play in natural language. In this section, I will have a look at some existing views (in broad outline) with an eye toward determining what might be appropriate for natural language analysis.

3.1 Justification and Knowledge

One major line of demarcation between theories of evidence and justification is that between internalist and externalist theories. Internalist theories, roughly speaking, are those which take justification to depend exclusively on the state of the agent, i.e. on her mental states; externalist theories conversely put some part of the responsibility for justifying in the external environment. Which kind of theory looks preferable for natural language evidentials?

We can start by considering Gettier cases. Here, Gettiered individuals are willing to assert evidentials. This means that we cannot have a purely externalist conception of evidence, for if we do, we predict that such assertions would be unwarranted. However, external observers judge such uses of evidentials false. One possible conclusion to draw is that two distinct conceptions are at work: assertions are performed according to an internalist conception of evidence, and evaluation is done according to an externalist conception. This sounds a bit odd. Can it be right?

Let us first consider what we might want for evaluation on this kind of view. We can imagine several different kinds of externalism that might work for evidentials. Two are identified by Fantl and McGrath (2009): 'radical' and 'moderate' externalism.[13] The two types differ in the degree to which they require external verifiability. Radical externalism requires the existence of a genuine reliable method for knowledge acquisition. This reliability also must be externally determined, which of course means that Gettiered contexts are irrelevant; the method of justification used by a Gettiered individual is by assumption unreliable. The result is that evidential statements based on evidence gathered according to the methods used by Gettiered individuals will be evaluated as false in general. This prediction is obviously wrong. A second possible position is moderate externalism. On this view, knowledge requires justification, but justification is understood externally. This position might look correct for evaluation; but consider a small complication of the case. Suppose we have two Gettiered individuals. One makes an evidential assertion, and the other evaluates it: here, the evidential sentence is judged true or appropriate, but from an external perspective there is obviously no proper evidence. The conclusion is that, in fact, both use and evaluation of evidential sentences use the same metric, which then leads to a requirement for an internalist view of justification (and knowledge, if we adopt the $E = K$ thesis).

There is a third option: the knowledge-level justification of Fantl and McGrath (2009:97). Having knowledge-level justification means having a (justified) belief that you are justified enough to know, though you may in fact be mistaken. This amounts

[13]These varieties of externalism are set up by Fantl and McGrath specifically for the analysis of knowledge, and so are adapted slightly to the equivalents for justification, our concern here.

to having higher-order beliefs about the reliability of one's own justification. This view also comes in internalist and externalist versions; here, I will adopt an externalist version. However, knowledge-level justification is a weak enough notion that it will still apply in to Gettiered agents: such agents believe that they are justified enough to have knowledge.

On this view, though, how can we account for the difference in judgments about felicity of agents inside and outside Gettier scenarios? Assertion and evaluation are both perspective-dependent. We assert what we believe, or think we know; but doing so depends on what we know at the time of assertion. However, evaluation can take place with respect to a different perspective, in much the way explored for epistemic modals by Egan (2006) and MacFarlane (2011), or indeed the entire body of work on epistemic logic (Fagin et al. 1995; van Ditmarsch et al. 2007). Does this mean that we must follow the relativists and take modal, evidential, or indeed plain vanilla statements to be assessor-sensitive? I think not: evaluating something as true does not mean that it is true. The difference between objective truth and subjective truth evaluation seems to be a point of confusion in this entire literature. But this is not the place to pursue this point in detail. In the rest of the chapter, I'll focus on the proper conditions for assertion and evaluation of evidentials. What we've seen so far is that, for assertion, an internalist model is nearly sufficient, i.e. Fantl and McGrath's 'knowledge-level justification', which, given that it is determined by agent beliefs about the reliability of their warrants, cannot but be essentially internalist in conception. However, the relevant notion must also involve some higher-order beliefs about the match between justification and external world. The same holds, mutatis mutandis, for the evaluation of evidential sentences for truth; but, since such evaluation is necessarily agent-dependent (since no evaluation happens except by the agency of some individual), this evaluation need not track actual truth at all. I will discuss the relation between these two aspects of linguistic practice in Sect. 4.

3.2 Awareness

Before moving to a discussion of the options we have for modeling natural language evidence, there is one more point to consider, related intimately to perspective-dependence. A piece of evidence cannot count as evidence for an agent unless that agent is aware that the evidence is indeed evidence. In other words, if I cannot recognize something as evidence for something else, it fails to be so. This situation differs from evidence in e.g. philosophy of science, where we are more interested in evidence as an absolute; here, awareness of the evidence *qua* evidence is strictly required.

Failure to recognize evidence as evidence can be rooted in at least two sources. One might fail to recognize the relationship between the evidential proposition and the proposition which it is evidence for. A simple case of such a situation might involve a person in a new country who is not aware of all the relevant social conventions. For example, suppose a European moves to Japan to work in a Japanese company.

She sees one employee bow deeply to another. However, she is not aware that the bow is *too* deep and has a sarcastic character indicating lack of respect; instead, she concludes that the first employee respects the other.[14] The problem here is that she was not sufficiently familiar with the relevant conventions and so was unable to recognize the particularized evidential relationship.

The other obvious source of failure is the lack of a relevant concept. This situation is brought out clearly by an example from Audi (2002), who writes "if a child has no concept of an insurance adjuster, then seeing one examining a damaged car and talk to its owner about deductibles will not function as a source of justification for the proposition that this is an insurance adjuster" (2002:89). We could view this as a special case of failure of recognition. Still, this issue is not one that will arise often in the linguistic case, simply because lacking the relevant concept is enough to (virtually) guarantee that the speaker will not utter any evidentially marked sentence which contains terms denoting the content in question.

Plainly, if an agent fails to recognize something as evidence she should not be licensed to use an evidential. We would therefore like to build this restriction into the concept of evidence we use for evidential terms. This need ties closely together with the view of evidence as perspective-based that Gettier scenarios already make clear that we require. I will explore ways to spell this out in the next section.

4 Options for Evidential Evidence

4.1 Desiderata

We are now in the following situation. We require a theory of evidence on which evidence is sensitive to the awareness and perspectives of agents, in order to account for cases of failure to recognize evidence and the behavior of evidentials in internal Gettier cases. Still, as we have seen, in skeptical scenarios use of evidentials is not so good, and they are also judged false or inappropriate by external observers of Gettier scenarios; this means that we need some means of tracking the external environment as well—or, on the internalist view, tracking individuals' beliefs about their relation to the external environment. In the rest of this section I will explore a couple of ways of resolving this tension, and finally settle on a specific proposal.

4.2 Knowledge

The first option is to accept Williamson's $E = K$ thesis and take the evidence required for evidential use to necessarily be knowledge. This move is straightforward and,

[14] As in other societies, excessive use of honorific speech or 'honorific behavior' is naturally interpreted ironically.

in conjunction with a theory of the effect of evidence (e.g. raising subjective probabilities) yields a reasonable looking picture. In earlier work, I adopted this view *simpliciter* (McCready 2008, 2010). However, I now believe that more is required.

The main reason involves the relation between this view and evidence. If one follows virtually all work in epistemology and assumes that knowing something requires more than simply believing it, difficulties arise for the fit between felicitous use of evidentials and true knowledge of the required evidence. Consider internal Gettier scenarios again. What does the agent know in cases like these? Recall the horse scenario from (5). Here, Johnny's evidence is the content of his perceptual field: hearing what he takes to be the sounds of a horse, and seeing what he takes to be the form of a horse. Does he know these propositions? I suppose it depends on what exactly they are. Is his evidence that he sees a horse, or that he sees a horse-like form? If the former, he does not know it, if the latter, clearly, he does.[15] The conclusion of McCready (2008, 2010) that $E = K$ rested on missing this particular point. I now think the right answer is that his evidence is the latter proposition. His belief that there is a horse follows from a Millarian 'quasi-inference', which amounts to making the assumption that perception is reliable (Millar 1991). If this is correct, then assuming that evidence must be knowledge does not help. Johnny does know his evidence. The problem lies in the reliability of the inference from perception of the horse-like form to the conclusion that there is a horse. This problem is completely separate from the question of whether or not the $E = K$ thesis is true. If the above reasoning is correct, we must look elsewhere for the characterization we need.

For completeness, we should consider the other possibility: that what Johnny knows is that there is a horse. If this is the evidence at issue, then, again, knowledge is not going to help us much: in this case, it will amount to mere belief, since there is no horse at all. In other words, the required characterization of belief is nothing more than Fantl and McGrath's 'knowledge-level justification', sufficient justification for the agent to believe that he knows and nothing more. But if this is all we have, then knowledge won't do much to help decide the case. The $E = K$ thesis thus seems to be completely inert in the present case.

4.3 Higher-Order Beliefs About Probabilities

Here is a possibly better way to go. Suppose that we take the acquisition of evidence to be the self-ascription of an increase in subjective probability based on the putative evidence. (Let me put aside for a moment possible incoherencies in the above notion.)

---

[15]This issue is a difficult one for theories of direct evidentiality as well; how reliable is perception? I think the right move here is to assume something like Faller's 'best possible grounds', by which we can evade the problem. Surely perceiving a horse(-like form) is the best possible grounds one could have for believing that there is a horse, regardless of whether or not there actually is one. One also wonders in this context about the evidential basis of Wittgensteinian hinge propositions (Wittgenstein 1991), and what evidentials are used with them. I do not have data that speaks to this issue.

We then get the assertability of evidentials in internal Gettier cases: based on the putative evidence, the agent self-ascribes a rise in probability (which she would not do if she had all the relevant information). Conversely, from an external perspective, it is clear that the putative evidence is merely a confusion, so the external observer does not self-ascribe any changes in probability distributions. As a result, the evidential sentence will be assertible for the Gettiered individual, yet the uttered sentence will be judged false by the external observer. This is precisely as desired.

However, is the proposed definition coherent? The possible issue is this: subjective probability distributions are based on the degrees of belief agents assign to various propositions (whatever those may be; see Hajek and Eriksson 2007). Is it conceptually possible to have a subjective probability distribution without being aware of the probabilities that are assigned? In other words, are we always aware by definition of our subjective probability distribution? If we are so aware, then self-ascription follows and the theory fails to discriminate between cases.

There are a number of possible fixes for this problem. The most obvious is to simply assume that agents are not necessarily aware of what probabilities they assign to various propositions. This amounts to abandoning positive introspection for beliefs. The details of how this would go depend on the background theory of belief, but in standard Kripke semantics we have to give up the 4 axiom (transitivity on states). While this is nonstandard, there is not anything necessarily wrong with it. Still, one would prefer a theory of evidentiality which did not require making assumptions about the nature of belief; if anything, the implications ought to go the other way.

A second option would be to introduce more complex notions of introspection. For example, one might suppose that agents are introspective, but that there is some higher-order uncertainty about their belief states. We might model this as uncertainty about possible subjective probability distributions; agents could be associated with a (convex) range of such distributions, in terms of which introspective knowledge could be defined, for example in terms of the mean of a sampling value (cf. Kruschke 2010). The downside to this move is the now highly complex nature of belief states and subjective probabilities. Perhaps such states are empirically necessary, but, again, one would prefer to do without them if possible.

A third option would be to move away from subjective probabilities and have agents self-ascribe properties involving probabilities of other kinds. For instance, one might self-ascribe being in a world where the objective probability of ϕ increased on the basis of the truth of ψ, ie. a world in which ψ is evidence for ϕ. We now have something resembling the picture developed by David Lewis in Lewis (1979), though he did not talk about probabilities there. This seems like a reasonable possibility. I will return to it shortly, but first let us consider what might appear to be a simpler and more attractive way of ensuring self-ascription of probability increases.

4.4 Subjective Probabilities with Judges

Suppose that we assume subjective probability distributions as usual, but take the relevant agents to be, not simply individuals, but *judges*. The idea of a judge has been used pretty extensively in recent work on such phenomena as predicates of personal taste (Lasersohn 2005). The basic idea is that terms like these have truth conditions that are, essentially, relativized to individual interpreters; thus, we can disagree about whether (for example) horse sashimi is tasty without either of us being strictly speaking wrong. This view has been extended to epistemic modals by Stephenson (2007), motivated by two factors: the apparent agent-dependence of the truth of statements including epistemic modals, and the availability of shifts in agent in certain contexts. As I said above, the former consideration does not strike me as compelling, due to the difference between truth proper and truth evaluation. The latter parallel is telling, though. Both predicates of personal taste and epistemic modals shift in terms of 'judge' under attitude predicates and in questions; we can find similar facts with Japanese experiencer predicates and certain indexicals (McCready 2007).

(13) a. Horse sashimi is tasty.
b. It might be raining. (judge = speaker)

(14) a. Robert thinks horse sashimi is tasty.
b. Robert thinks it might be raining. (judge = Robert)

(15) a. Is horse sashimi tasty?
b. Might it be raining? (judge = hearer)

With embeddable evidentials in e.g. Japanese, we find precisely similar shifts: under attitudes, the source is the attitude holder, and in questions the hearer (in general), but in simple sentences, it is the speaker. (I omit specific examples for space reasons.) The parallel indicates that—if it is reasonable to assume judge sensitivity in personal taste predicates and modals—it is reasonable to assume judge sensitivity for evidentials as well.[16]

If so, we can derive certain aspects of the preferred interpretation of evidentials. But recall that the original motivation was to simplify the self-ascriptive aspect of linguistic evidence. The question to ask therefore is: do we get *de-se*-ness for free on this theory? I think that the answer must be negative. For an analogy, we may ask whether people are necessarily aware of their tastes. Since taste predicates (on these theories) involve judge-sensitivity, we would anticipate that the same sort of self-ascriptive quality found with evidence should be found with taste predicates as well. If it is not, much of the reason to adopt a judge-based theory of evidence disappears.

[16] Is it actually reasonable at all? That is a different question entirely. The fact that shift is not obligatory in questions, only preferred, begins to make it appear that it might *not* be desirable, given that one must assume something like optional application of some monstrous operator to derive the facts. Better might be a fully pragmatic story: from aspects of the meaning of the 'judge-sensitive' expressions, derive a preferred interpretation on the basis of independently motivated rules or knowledge of the language and how it is typically used. See McCready (2012) for a framework that might be applicable in this context, given the right setup of lexical entries.

Note that this is not to say that evidentials themselves do not make use of whatever mechanism determines the perspective holder in taste predicates or epistemic modals, only that judge sensitivity is not in play at the deeper level of (subjective) evidence itself.

Can one be unaware of one's tastes? Clearly, the answer is yes. Imagine a woman or a man who likes only men who are bad for her/him but doesn't realize it,[17] or someone who finds herself eating junk food all the time despite believing that she prefers healthy food. The behavior of such individuals shows that they have certain tastes, though consciously these tastes are not accessible to them, and, if asked, they might well deny those tastes. The view that people's behavior shows their tastes better than their overt opinions is the theory of revealed preference (Samuelson 1938), which is widely accepted in economics, though not completely uncontroversial. Tastes, then, do not require self-ascription, though (I have argued, with others) evidence does. It thus seems to be an error to identify self-ascription with judge-sensitivity.

The reader might be worrying at this point about the connection between assertion and self-ascription. Someone who is not aware of a given preference would not assert that preference, just as someone who is not aware of some evidence would not assert the existence of that evidence. This is certainly true. However, as I said above, I am not disputing that both evidentials and taste predicates exhibit something like judge-sensitivity in interpretation.[18] The question here is the proper way to characterize the required notion of evidence. I have argued that, for this, we need something like the self-ascription of probability increases. The fact that self-ascription is not required for preference shows that the parallel with taste predicates cannot be the proper way to approach this particular question.

4.5 Proposal

Accordingly, let me make a proposal which directly references self-ascription. The idea will be to take the usual account of evidence as increase in probability via conditionalization, and to self-ascribe the property of being in a world in which the required increase occurs. The proposal thus comes in two parts: the change in probabilities, and the self-ascription of that change.

Describing evidence in terms of probability is completely straightforward. What we want is just the following:

(16) φ is evidence for χ iff $\mathsf{P}(\chi|\varphi) > \mathsf{P}(\chi|\neg\varphi)$, where $(\chi|\varphi)$ is the conditionalization of χ on φ.

Should we think of the above conditionalization in terms of subjective probabilities or something else, say objective or logical probabilities? As discussed in the last sections, using subjective probabilities here probably requires a highly nonstandard

[17] Thanks to Jason Quinley (p.c) for this example.

[18] Or whatever the proper way is to characterize the relevant dependencies; see footnote 16.

logic of belief, and should be avoided if possible. However, using other theories of probability may have strange consequences, as it is perfectly possible for ψ to be evidence for ξ without any human being aware of this fact. If we tied the felicity of using evidentials to this probability change, wrong predictions about truth evaluation would result.

However, this problem can be avoided by introducing self-ascription. We need some notion of self-ascription (or reflexive awareness) anyway in order to account for other intuitive requirements on evidence, as with the Audi (2002) example of the insurance adjustor. A straightforward technical implementation of this idea is due to Stalnaker (2008). On this theory, propositional content consists of sets of world-individual pairs,[19] rather than the familiar sets of worlds: so, roughly, $[\![\phi]\!] = \{\langle w, i\rangle : \phi(w)\}$. This corresponds to one way of understanding Lewis (1979) view of 'propositions as properties,' which allows self-location in worlds in which certain propositions hold, as required for modeling the cognitive status of indexical content (Perry 1979). Stalnaker ensures such self-location by modeling belief by a strengthened version of the usual accessibility relations on worlds, so that $\langle w, i\rangle R \langle w', i'\rangle$ holds only if i believes in w that she is in world w' and that she is i'.

These two elements can be combined as follows. Suppose that subjective evidence, as required for use of evidentials, amounts to self-ascription in a world where such evidence exists. Let us define a predicate *EVID* for convenience as follows:

(17) $$EVID(\chi, \varphi) \longleftrightarrow \varphi \text{ is evidence for } \chi \text{ (cf.16)}$$

Then, for φ to count as subjective evidence for χ with respect to agent a, we require:

(18) $$SE(a,\chi,\varphi) \longleftrightarrow$$
$$\forall\langle w,a\rangle[\langle w_{@},a\rangle R_{dox}\langle w,a\rangle \rightarrow \langle w,a\rangle \in \{\langle w,i\rangle : EVID(\chi,\varphi)(w)\} \cap \{\langle w,i\rangle : K_i\varphi(w)\}]$$

The above simply says that φ is subjective evidence for χ for agent a just in case, in all of the agent's belief-accessible worlds (which are also required to be worlds in which she self-locates), φ is indeed evidence for χ (defined in terms of probability increase), and a knows, in that world, that φ. The latter amounts to implementing knowledge-level justification.

Does this analysis meet the desiderata presented at the beginning of this section? Obviously it accounts for awareness: it is specifically designed to do so. Does it give the right results for the cases we considered? For basic cases of use of evidentials, it will give the right results: if the piece of evidence is actually true and actually a piece of evidence by the criterion in (16), use of the evidential will both be licensed and deemed true. For internal Gettier scenarios, the evidential user will be justified in use of an evidential sentence, given belief in the apparent evidence φ; this is as desired. Conversely, for an external observer, φ will be false in that observer's doxastically accessible worlds, if resolved to the proposition believed by the Gettiered individual;

19 Officially, Stalnaker's formulation uses pairs of *centers* and worlds, where centers are pairs of individuals and times. In this chapter, I am not concerned with temporally dependent propositions, so I will redact this aspect of the theory.

if resolved to the actual, misperceived fact, it will be true, but no longer evidence. This is also the right prediction. Finally, for the case of skeptical scenarios, the evidential user will no longer be able to self-ascribe the proper evidential relation. All in all, this notion of evidence looks the proper one for the analysis of natural language evidentials.

5 Conclusion

This chapter has proposed an analysis of the evidence required for use of natural language evidentials. According to this proposal, a proposition is evidence for χ for an agent only if, after learning it, the agent self-ascribes an increase in the likelihood of χ. I showed that this view is capable of making sense of complex data involving skeptical scenarios and Gettier cases, both internal and external. That said, it is not the case that no problems remain, or that the above analysis exhausts all complications related to the notion of evidence. For example, the following dialogue is a translation from the Japanese; 'must$_{inf}$' picks out an inferential evidential like *mitai*, which, as we have seen, is relatively weak. Still, infelicity arises in cases like this one, where the causal nature of the connection between the evidence itself and what it is supposed to be evidence for is called into question.

(19) a. A: It must$_{inf}$ have rained last night.
b. Why do you say so?
c. Well, the ground is wet.
d. But that's from the sprinkler.
e. Then I was wrong./# Still, it must$_{inf}$ have rained.

Note that in this case all requirements I specified are satisfied: the speaker has self-ascribed an increase in probability, which is, prior to learning the actual cause of the ground being wet, completely justified as far as we can tell. Does this mean that more is required for evidence in natural language, perhaps even genuine causality? I do not think so. This issue is closely related to Gettier cases. It is just that here the Gettiered individual becomes 'un-Gettiered' after learning a piece of relevant information. After so learning, he is in a position to evaluate his previous statement as involving a spurious correlation, and consequently no longer self-ascribes a probability increase. Note the similarity to the cases of changes in judgements on epistemic modal statements discussed by e.g. von Fintel and Gillies (2008); again, we have not so much a change in truth-value as a change in truth-value judgement. Note though that cases like these require us to deploy the full Stalnaker system, as we require the ability to model temporal dependence of (self-locating) belief.

Is the notion proposed in this chapter the only one at issue for natural language evidentials? Further research is required to determine whether the concept proposed is a universal one or whether it is specific to the particular language studied, and even to the particular evidentials examined within that language. My suspicion is that it is indeed universal, but that certain evidentials (like *rashii*) may have more stringent additional requirements, given how they behave with respect to certain kinds of assertions and under certain kinds of modifiers, issues about which I cannot go into detail here. Still, it is my hope that the present proposal puts the study of evidentiality onto a somewhat firmer footing than previously, and that the results of the chapter about subjectivity and evaluation have helped to show something about the nature of evidential knowledge.

Acknowledgments I owe thanks to very many people over the course of this project. First is Norry Ogata, with whom I started working on this topic. It is too bad we could not have written this chapter together. I would also like to thank Chris Davis, David Etlin, Magda Kaufmann, Chris Kennedy, Linton Wang, the anonymous reviewers, and audiences at LENLS, GLOW, and the University of Chicago.

References

Asher, N., & McCready, E. (2007). Were, would, must and a compositional account of counterfactuals. *Journal of Semantics*, *24*(2), 93–129.
Audi, R. (2002). The sources of knowledge. In *The Oxford handbook of epistemology* (pp. 71–95). Oxford: Oxford University Press.
Bach, K., & Harnish, R. (1979). *Linguistic communication and speech acts*. Cambridge: MIT Press.
Bronnikov, G. (2011). Representation of Inference in the Natural Language. Ph.D. thesis, University of Texas at Austin.
Brown, J., & Herman, C. (Eds.). (2011). Assertion. Oxford: Oxford University Press.
Chung, K.-S. (2005). Space in tense: The interaction of tense, aspect, evidentiality and speech act in Korean. Ph.D. thesis, Simon Fraser University.
Chung, K.-S. (2007). Spatial deictic tense and evidentials in Korean. *Natural Language Semantics*, *15*, 187–219.
Davis, C., Christopher, P., & Peggy, S. (2007). The pragmatic values of evidential sentences. In M. Gibson & T. Friedman (Eds.), *Proceedings of SALT 17*. CLC Publications.
de Haan, F. (1999). Evidentiality and epistemic modality: Setting boundaries. *Southwest Journal of Linguistics*, *18*, 83–101.
DeLancey, S. (1997). Mirativity: The grammatical marking of unexpected information. *Linguistic Typology*, *1*, 33–52.
DeRose, K. (1992). Contextualism and knowledge attributions. *Philosophy and Phenomenological Research*, *52*, 913–929.
DeRose, K., & Warfield, T. (Eds.). (1999). *Skepticism: A contemporary reader*. Oxford: Oxford University Press.
Egan, A. (2006). Epistemic modals, relativism and assertion. *Philosophical Studies*, *133*, 1–22.
Fagin, R., Halpern, J., Moses, Y., & Vardi, M. (1995). *Reasoning about knowledge*. Cambridge: MIT Press.
Faller, M. (2002). Semantics and pragmatics of evidentials in Cuzco Quechua. Ph.D. thesis, Stanford University.
Faller, M. (2004). The deictic core of 'non-experienced past' in Cuzco Quechua. *Journal of Semantics*, *21*, 45–85.
Faller, M. (2007). The Cuzco Quechua reportative evidential and rhetorical relations. In A. Simpson & P. Austin (Eds.), *Linguistische Berichte* (Vol. Sonderhafte 14, pp. 223–252). Hamburg: Helmut Buske Verlag.
Fantl, J., & McGrath, M. (2009). *Knowledge in an uncertain world*. Oxford: Oxford University Press.
Fodor, J. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

Fodor, J. (2000). *The mind doesn't work that way*. Cambridge: MIT Press.
Garrett, E. (2001). Evidentiality and assertion in Tibetan. Ph.D. thesis, UCLA.
Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, *23*, 121–123.
Hajek, A., & Eriksson, L. (2007). What are degrees of belief? *Studia Logica*, *86*, 185–215.
Halpern, J. Y. (2003). *Reasoning about uncertainty*. Cambridge, MA: MIT Press.
Hawthorne, J. (2004). *Knowledge and lotteries*. Oxford: Oxford university Press.
Kennedy, C. (2007). Vagueness and gradability: The semantics of relative and absolute gradable predicates. *Linguistics and Philosophy*, *30*(1), 1–45.
Kratzer, A. (1981). The notional category of modality. In H.-J. Eikmeyer & H. Rieser (Eds.), *Words, worlds, and contexts: new approaches in word semantics, no. 6 in Research in text theory* (pp. 38–74). Berlin: de Gruyter.
Krifka, M. (1992). Thematic relations as links between nominal reference and event domains. In I. Sag & A. Szabolcsi (Eds.), *Lexical matters* (pp. 29–53). Stanford, CA: CSLI Publications.
Kruschke, J. (2010). *Doing Bayesian data analysis*. Manuscript, Indiana University.
Lasersohn, P. (2005). Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy*, *28*, 643–686.
Lewis, D. (1979). Attitudes de dicto and de se. *The Philosophical Review*, *88*, 513–543.
Lewis, D. (1980). A subjectivist's guide to objective chance. In R. Jeffrey (Ed.), *Studies in inductive logic and probability* (Vol. 2). University of California Press. (Reprinted in *Philosophical papers*, Vol. 2 by L. Davis, Ed.).
Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, *74*, 549–567.
MacFarlane, J. (2011). Epistemic modals are assessment-sensitive. In B. Weatherson & A. Egan (Eds.), *Epistemic modality*. Oxford: Oxford University Press.
Matthewson, L., Rullmann, H., & Davis, H. (2007). Evidentials as epistemic modals: Evidence from St'át'imcets. In J. van Craenenbroeck (Ed.), *Linguistic variation yearbook 2007*. Amsterdam: John Benjamins.
McCready, E. (2007). Context shifting in questions and elsewhere. In E. Puig-Waldmüller (Ed.), *Proceedings of Sinn und Bedeutung 11* (pp. 433–447). Barcelona: Universitat Pompeu Fabra.
McCready, E. (2008). Evidentials, knowledge and belief. In Y. Nakayama (Ed.), *Proceedings of LENLS 5*.
McCready, E. (2010). Evidential universals. In T. Peterson & U. Sauerland (Eds.), *Evidence from Evidentiality* (Vol. 28, pp. 105–128) of *UBC Working Papers in Linguistics*. University of British Columbia.
McCready, E. (2011). Testimony, trust, and evidentials. To appear in *Evidentials and Modals*, C.-M. Lee & J. Park, (Eds.), CRiSPI, Brill Publications.
McCready, E. (2012). Emotive equilibria. *Linguistics and Philosophy*, *35*, 243–283.
McCready, E., & Asher, N. (2006). Modal subordination in Japanese: Dynamics and evidentiality. In A. Eilam, T. Scheffler, & J. Tauberer (Eds.), Penn Working Papers in Linguistics 12.1 (pp. 237–249).
McCready, E., & Ogata, N. (2007). Evidentiality, modality, and probability. *Linguistics and Philosophy*, *30*(2), 147–206.
Millar, A. (1991). *Reasons and experience*. Oxford: Oxford University Press.
Murray, S. (2010). Evidentiality and the structure of speech acts. Ph.D. thesis, Rutgers.
Ogata, N. (2005). *A multimodal dynamic predicate logic of Japanese evidentials*. Paper presented at Language Under Uncertainty workshop: Kyoto University.
Perry, J. (1979). The problem of the essential indexical. *Noûs*, *13*, 3–21.
Pollock, J., & Cruz, J. (1999). *Contemporary theories of knowledge*. Lanham: Rowman Littlefield.
Quine, W. V. O. (1960). *Word and object*. Cambridge: The MIT Press.
Samuelson, P. (1938). A note on the pure theory of consumers' behavior. *Economica*, *5*(17), 61–71.
Sapir, E. (1922). The Takelma language of southwestern Oregon. In F. Boas (Ed.), *Handbook of American Indian languages* (Vol. 2, pp. 1–296). Washington: Government Printing Office.
Stalnaker, R. (2008). *Our knowledge of the internal world*. Oxford: Oxford University Press.
Stanley, J. (2005). *Knowledge and practical interests*. Oxford: Oxford University Press.

Stephenson, T. (2007). Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy*, *30*(4), 487–525.
van Ditmarsch, H., van der Hoek, W., & Kooi, B. (2007). *Dynamic epistemic logic*. Berlin: Springer.
Veltman, F. (1996). Defaults in update semantics. *Journal of Philosophical Logic*, *25*, 221–261.
von Fintel, K., & Gillies, T. (2008). CIA leaks. *The Philosophical Review*, *117*, 77–98.
Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
Wittgenstein, L. (1991). *On certainty*. Oxford: Wiley-Blackwell.

A Categorial Grammar Account of Information Packaging in Japanese

Hiroaki Nakamura

Abstract A particular characteristic of Japanese language is that information structure is explicitly indicated by overt morphological means, i.e., marking expressions with the topic particle WA, which has attracted significant attention in the literature mainly from a pragmatic perspective. This study sheds light on the semantic effects that this particle has on the contents of sentences and proposes a categorial grammar approach to information packaging. The purpose of this study is two-fold: First, by observing minimal pairs of sentences that differ only in the use of the subject-marking particles and the presence/absence of focal accents on the particles, we identify the influences of these particles on the truth conditions of sentences, and explore proper semantic representations for subjects marked with topic and nominative particles. Then, we examine the syntactic function of this topic particle to show a type of *concord* with sentence-final predicates, as suggested by the term *kakari-josi* or concord/coherence particle used in Japanese traditional linguistics. On the basis of the results, we argue that the topic particle induces information packaging (Vallduví 1992) as its lexical property, thereby yielding tripartite information structures following Hajičová et al. (1998), which can be automatically obtained through readings of categorial proof nets. In addition, we also focuses on sentences with sentence-internal topics, cleft-constructions and multiple topic sentences from the viewpoint of incremental processing within a categorial grammar framework.

This study has tremendously benefited from the discussions with Kei Yoshimoto, Yoshiki Mori, and Masahiko Kobayashi. I am also indebted to the reviewers for this volume for their helpful comments on the previous versions of this study. Any remaining errors are my sole responsibility.

H. Nakamura (✉)
International Education Center, Hirosaki University, 1-1, Bunkyo-cho,
Hirosaki, Aomori 036-8224, Japan
e-mail: hiroaki-nakamura@ax2.mopera.ne.jp
http://www.hirosaki-u.ac.jp

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*,
Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_9,

Keywords Topic particle WA · Topic-comment articulation · Information packaging · Categorial grammar · Syntax-semantics interface

1 Introduction

This study examines the truth-conditional effects that the topic-particle has on the interpretation of Japanese sentences, and attempts to explain how information packaging is realized by the lexical property of the topic particle in terms of the tight syntax-semantics relationship built into the theory of grammar. This topic particle, from the start, has attracted significant attention in the study of Japanese (often referred to as a topic-prominent language) generative grammar. However, the main concern is syntactic analysis, such as which position topic phrases occupy in a tree diagram, or a pragmatic difference in discourse status (old/new) between subjects and topics. Although it is generally agreed that topics tend to occur sentence-initially, convey given information, and establish some entities as a locus of information update, there are obvious exceptions to these generalizations.

First, we consider some examples in which the choices of topic and subject markers have a direct influence on the truth conditions of sentences, and re-examine the semantic characterization of topicalized and non-topicalized sentences, including those with contrastive and exhaustive listing readings. As seen in (1), the topic particle is taken as a topic operator, which forms a tripartite semantic structure for a topicalized sentence, taking as arguments a domain restrictor (a set of entities denoted by a topic-marked phrase) and a nuclear scope (main predication denoted by the remaining parts of the sentence and applied to the former):

(1) $\text{Top}\, x\, (Px \rightarrow Qx)$

Next, we explore derivations of various topicalized sentences including those in which topics appear sentence-internally and even in embedded clauses. Assuming the semantic characterization of topicalized sentences in (1), sentence-internal topics give rise to discontinuity of the constituents to be mapped into the nuclear scope, which seems to be very difficult to deal with for any theory with a strict compositional characterization of the syntax-semantics interface.

The main goal of this paper is to show that a semantic structure like example (1) can be automatically derived from the syntactic analysis of a topicalized sentence. Adopting Combinatory Categorial Grammar, we assign a higher-order category over predicates to a topic-marked expression. This is associated with a matrix predicate through a lexical property of the topic particle to derive an appropriate nuclear scope. Then, we focus on sentences with sentence-internal topics, which form discontinuity in the main predication. We will argue that topicalized sentences with different word orders and discontinuity can be parsed by adopting the discontinuity categorial operators proposed by Morrill (2011) and that proper interpretation of topicalized sentences can be automatically derived from their syntactic analyses represented as categorial proof nets. While discontinuous constituency usually presents difficulties in processing sentences, on account of the increasing unresolved dependencies at intermediate stages, it might be related to marked linguistic phenomena. Therefore, this study attempts to associate the complexity profiles with focus related readings of topics and subjects (contrastive and exhaustive readings).

2 Use of WA and GA in the Literature

Before presenting our analysis of information packaging induced by the topic marker WA, let us review some properties of this particle in comparison with the nominative case marker GA. Recently, following the classification by Diesing (1992), it has been argued that subjects of kind-/individual-level predicates are marked with WA by default, whereas those of stage-level predicates are marked with GA, even though the latter may be followed by WA when their referents are evoked from previous discourse. Assuming that indefinite subjects in generic sentences are required to have existential presuppositions, even if their referents are *new* in discourse, all WA-marked phrases should be treated as presupposition triggers because the existence of the referents of such expressions cannot be canceled under negation.

Along similar lines, Kuroda argued that sentences expressing categorical judgment (mostly individual or kind-level sentences) have a strong tendency to have their subjects marked with WA (see papers in Kuroda 1992, 2005 among others), as illustrated in (2):

(2) Saikin-no nihonjin-wa/$^{*?}$nihonjin-ga se-ga takai.
Recent-Gen Japanese-Top /Japanese-Nom height-Nom high.
'Recent Japanese are tall.'

As often revealed in the literature, the use of WA is usually permitted only in matrix clauses (called the root phenomena, see Heycock (2008) for a summary) that express enduring properties of the subjects. However, question words and corresponding phrases in answers can never be marked with WA[1],as shown in (3):

---

[1] The complex question phrase of form *dono* 'which' + N can be followed by WA, with contrastive connotations, as observed in (i):

(*i*) a. Kimi-wa dono hon-o/-WA yonda-no?
you-Top which book-Acc/-Top yonda-no?
'Which book(s) did you read?'

b. Chomusukii-wa moo yomi-masita.
Chomsky-Top already read-Past
'I already read Chomsky's book(s).'

The object marked with topic indicates that the set of books from which the answer was chosen is familiar from context among the interlocutors.

(3)

a	Dare-ga/*Dare-wa	se-ga	takai-no?
	Who-Nom/Who-Top	height-Nom	tall-Q
	'Who is tall?'		
b.	Tanaka-san-ga/*Tanaka-san-wa	takai-desu.	
	Tanaka-Mr.-Nom/Takana-Mr.-Top	tall-Pres.	

(3b) may be acceptable if the topic-marked subject is interpreted as a contrastive topic, such as ‘at least Mr. Tanaka is tall.’

Since Kuno (1973), it is generally agreed that WA marks constituents whose referents are already present in the common ground (in other words, given, familiar, predictable/recoverable, etc.), but this generalization appears to be greatly exaggerated. As seen in example (4), new referents can (and sometimes must) be marked with WA when introduced into discourse, and the topic marker cannot be replaced with the nominative case marker GA:

(4)

a.	Syaku Esyo-wa/*ga	Enkou-ji-no	syamon	nari-ki.
	Priest Esyo-Top/Nom	Enko-temple-Gen	priest	be-Past
	Houshi,	ikeri-si toki, ...		
	Priest	alive-Be when ...		
	'Priest Esyo was a monk of Enko temple. When he was alive, he ...'			
		Nihon-Reiiki: Dai-19, p. 136		
b.	Sou Esyo-wa/*ga	Enko-ji-no	sou	-de-atta.
	Priest-Esho-Top/*Nom	Enko-temple-Gen	priest-	be-Past
	Kono-sou-wa,	seizen, ...		
	this-Priest-Top	in life ...		
		(Translation of (4a) in modern Japanese)		

It has been assumed that the choice between the topic and nominative subject is relevant only in regard to pragmatics, and the semantic effects of information packaging have received limited attention in the literature. However, the choice between the two particles often brings about serious effects regarding the truth conditions of sentences.

3 Semantic Effects of Information Packaging

First, let us see an English example introduced by Rooth (1992), which shows that focus is decisive for the interpretation of the sentences like (5). Accented words are underlined here and henceforth.

(5) a. In English orthography, a ‘**U**’ always follows a ‘Q’. (true)
b. In English orthography, a ‘U’ always follows a ‘**Q**’. (false)

On the basis of English orthography, (5a) is true, saying that, whenever there is a 'Q', it is followed by a 'U', whereas (5b) states that a 'U' only appear after a 'Q' in English, which is incorrect. We can observe a similar contrast in Japanese sentences, as seen in (6a) and (6b), where the difference in the use of topic and subject markers is not merely pragmatic:

(6) a. Eigo-no seisyo-hou-ni-oite, U-ga tuneni Q-no atoni
English-Gen orthography-In U-Nom always Q-Gen after
arawarer-u. (true)
appear-Pres

b. Eigo-no seisyo-hou-ni-oite, U-wa tuneni Q-no atoni
English-Gen orthography-In U-Top always Q-Gen after
arawarer-u. (false)
appear-Pres

As shown above, (6a) with the GA-marked subject is true, whereas most Japanese would judge (6b) as false if the topic is interpreted as a thematic one. Even if the nominative subject in (6a) is stressed and it yields an exhaustive listing reading, the truth of this statement does not change (and it even becomes closer to the meaning of the English sentence in (5a))(to which we will return later). Accounting for the intuitions that yield different truth conditions for sentences in (6), the indefinite nouns include some quantificational force and the difference in truth conditions must be (at least partly) reflected in the lexical semantics of these particles because there is no other difference between the two sentences. In addition, (6b) can be taken as a characterizing sentence, in the sense of Krifka et al. (1995), thus representing the predication of the enduring property of 'U's. However, it is important to note that, when the topic is focalized and the contrastive reading is forced, the judgment becomes subtle. Many speakers, including this author, still find (6b) with the focalized topic to be true, which means that at least 'U's may follow 'Q's without mentioning anything regarding the falsehood of other vowels showing up after 'Q's.

At this point, let us consider the proper semantic interpretations of the four cases, beginning with sentence (6a), which includes the unstressed nominative subject. If *tsuneni* 'always' does not appear in (6a), then it is reasonable to assume that this sentence means there is at least one occurrence of a 'U' following a 'Q' at a given time. It is also possible to regard this sentence as a generic one (by universally quantifying over worlds[2]) even if *tsuneni* does not occur because generic statements are expressed by simple present predicates. Therefore, we can assume that the nominative subject in (6a) induces existential quantification, which can be represented as in (7):

(7) $\exists x[U(\mathrm{x}) \wedge Follow(x, Q)]$

[2] Endriss represents the interpretations of sentences with quantificational adverbs like *tsuneni* 'always' as in (i):

(i) $\forall s[In(s, Q) \rightarrow Follow(s, U, Q)]$

See Endriss (2009) and references cited therein regarding the motivations and problems for representation (i).

Suppose that $U(x)$ roughly means 'there is an occurrence of 'U'/some letter x is a 'U'.' Let us ignore the details of the internal structure of the right conjunct *Follow* (x, Q) for the time being. If *tsuneni* appears to make its generic reading vivid, we can introduce a universal quantifier that binds an event variable shared by the two conjuncts, to denote that (7) holds in every situation. It should be noted that the interpretation for (6a) is not true in a strict sense without a universal quantifier like *tsuneni* because it does not exclude the possibilities of 'Q's being followed by other vowels, which is incorrect in English orthography.

Next let us turn to (6b) with the subject marked as topical by WA. This sentence is clearly false and this judgment is unanimously agreed upon by my informants. Because the only difference between (6a) and (6b) is the choice between the particles WA and GA, the falsehood of (6b) must be accounted for in terms of the semantics of these particles. Let us assume that the relationship between the set of occurrences of 'U' and the set of relevant entities (i.e., vowels) that occur after 'Q's can be expressed with material implication (i.e., the subset relation). Because the subjects of kind-/individual-level predicates have a strong tendency to be followed by WA, it is natural to assume that WA-marked phrases induce universal quantification. In other words, a generic/universal quantifier is inherent in WA. Note that the passive sentence in (8) is true because it states an inherent property of 'Q', not 'U.'

(8)	Eigo-no seisyo-hou-ni-oite,	Q-wa	U-ni	kouzoku-sarer-u.
		Q-Top	U-By	follow-Pass-Pres
	'In English orthography, Qs are followed by 'U's.'			

As shown above, (8) means that the set of all instances of 'Q' are included by the set of characters followed by 'U'. Accordingly, the meaning of (6b) should be represented as in (9a), while that of (8) should be represented as in (9b).

(9) a. $\forall x/\text{Gen}x[U(x) \rightarrow Follow\,(Q, x)]$
b. $\forall x/\text{Gen}x[Q(x) \rightarrow Follow\,(x, U)]$

Of course, because 'U's appear after a variety of letters other than 'Q's, the set of 'U's should never be a subset of the set of letters following 'Q's whereas the set of instances of 'Q' is included in the set of all alphabets preceding 'U's. (9a) and (9b) correctly reflect our intuition concerning the truth conditions of (6b) and (8), thus indicating that the topic particle actually encodes the implication relation.[3]

Now, let us consider the remaining two possibilities, i.e., the focalized versions of GA and WA. First, observe example (10) with the stressed nominative subject.

(10)	Eigo-no	seisyo-hou-ni-oite,	U-**GA**	tuneni	Q-no	atoni
	English-Gen	orthography-In	U-Nom	always	Q-Gen	after
	arawarer-u.	(true)				
	appear-Pres					

[3] Here we do not consider a model in which there are no 'U's and (6b) is judged to be true because an empty set is a subset of every set.

Even if adverbs like *tsuneni* do not occur, sentence (10) is still true, and exactly conveys the intended meaning. Sentences with the exhaustive-listing GA are usually translated into English like "'U's and (only 'U's) follow 'Q's/It is 'U's that follow 'Q's" in Japanese linguistics. Let us assume that, following Rooth (1992), focus induces a set of alternatives like Existential Introduction in natural deduction. In logic, Existential Introduction is freely available based on entailment. For instance, "John loves Mary" entails that "someone loves someone" ($Loves(m)(j)) \vDash \exists x \exists y.Loves(y)(x)$). However, in terms of question-answer congruence, the application of the rule needs to be linguistically constrained in order to generate relevant alternative sets for cases that imply exhaustiveness (in (10), some occurrences of 'U' follow 'Q's, and no other vowels do not show up after 'Q's). Furthermore, following Rooth and other researchers, let us assume that for any focused expression X, there is a class ALT(X) of its alternatives. Note here that ALT(X) does not refer to a set of occurrences of 'U's that do not follow 'Q's, but to the set of vowels other than 'U's. We will not pursue this complicated notion of 'alternatives' here, and simply represent the meaning of (10) as in (11):

(11) $\exists x[U(x) \wedge \forall y[Follow(Q, y) \rightarrow y = x]$

If there is no fear of confusion, we will represent the meaning of (10) with the exhaustive listing GA simply as $\forall x[Follow\,[(Q, x) \rightarrow U\,(x)]]$, where the set associated with the antecedent and the one associated with the consequent are reversed. In other words, it means that the set of occurrences of any vowel that appear after 'Q's is included in the set of all occurrences of 'U's. However, this does not say anything about the possibilities of 'U's following characters other than 'Q.'

Finally, let us briefly consider the semantics of the contrastive topic WA although we do not present any all-encompassing account of contrastive topics. Observe sentence (12):

(12)					
Eigo-no	seisyo-hou-ni-oite,	U-**<u>WA</u>**	tuneni	Q-no	atoni
English-Gen	orthography-In	U-Cont.	always	Q-Gen	after
arawarer-u					
appear-Pres					

This sentence means that at least "'U's can follow 'Q's." Tomioka (2010) suggests that the contrastive topic conveys a sense of incompleteness, non-finality, and/or uncertainty. In order to understand what incompleteness/non-finality/uncertainty actually means, let us examine a dialogue cited by Tomioka 2010:(5)

(13) A: Who passed?
B: **<u>KEN</u>**-wa/Ken-**<u>WA</u>** uka-ta.
KEN-TOP/Ken-TOP pass-PAST
'(At least) Ken passed.'

To make the meaning of sentence (13B) with the contrastive topic clearer, he created a scenario: Speaker B is an examiner, and Speaker A assumes that B has complete knowledge regarding the results of the exam. Let us suppose that three students (let's say, Ken, Mari, and Erika) took the exam. He suggests that A would conclude that, "based on the assumption that B knows the outcome of the exam, coupled with the general Gricean principle that requires B to be as informative as possible, would lead to the conclusion that Mari and Erika did not pass." According to Tomioka, sentence (13B) implies that the speaker does not wish to share the results of the other students' exams to the hearer.

However, we can easily come up with many contexts in which B's answer, which sounds incomplete or under-informative, can become natural even if he does not have complete knowledge of the outcome. For example, it is clearly possible that B does not know the results of the other students' exams, and he wants to say that he is sure that Ken passed, but does not want to commit himself to revealing the results of Mari and Erika.[4] Kuroda (2005) gives a similar account to this property of the contrastive WA, which he called *anti-exhaustivity*, an implicature carried by the contrastive WA, "to the effect that there is an entity other than what the WA phrases designates that does not (or less strongly, that might not) satisfy the predicate." We represent the meaning of the focalized WA, with recourse to the notion of alternative sets, again. Sentence (12) suggests that there are alternatives to 'U' relevant in discourse but the speaker is not committed to the truth values of implicit propositions in which alternatives (in this case, vowels other than 'U') are substituted for 'U.'

There is another important point that appears to be directly involved with the truth-conditional content of (12). Even if the focalized WA phrase evokes the set of alternatives (set of other vowels), (12) cannot be true if it conveys a generic reading (i.e., totality of things belonging to the set). In order for (12) to be judged true, as many informants including me judged, with the meaning that at least some instances of U follow Qs', this totality connotation must be weakened (although some speakers I consulted with have judged (12) to be false).

The non-totality brought about by the contrastive WA is illustrated by the following pair of sentences. It is widely assumed that contrastive topics can take a narrow scope with respect to negation, as illustrated in (14):

(14) a. Rijikai-ni kyoju-wa zenin-(ga) syusseki-sina-katta.
board-meeting-At professors-Top all-(Nom) attend-Neg-Past
'All of the professors did not attend the board meeting.'

b. Rijikai-ni kyoju-wa zenin-**WA** syusseki-sina-katta.
board-meeting-At professors-Top all-CT attend-Neg-Past
'Not all the professors attend the board meeting.'

While (14a) simply negates the proposition that all professors attended the board meeting, and should be represented as $\neg\exists x[\mathit{Professor}(x) \wedge \mathit{Attended\text{-}the\text{-}board\text{-}meeting}(x)]$, (14b) is true only if there is at least one professor who did not attend the board meeting, as shown in logical form (15):

(15) $$\neg\forall x[Professor(x) \rightarrow Attend\,(Meeting)(x)]$$
$$= \exists x[Professor(x) \wedge \neg Attend\,(Meeting)(x)]$$

[4] For instance, following B's answer in (13), we can add a sentence like *Hoka-no hito-no koto-wa sir-anai-kedo* "I don't know the results of the others, though" which sounds quite natural.

Although there must be more to be said about entailment, implicature and presupposition conveyed by the contrastive WA here, a detailed analysis about these concepts goes beyond the scope of this present study. It would suffice to give a semantic representation for (12) with the contrastive WA here.

(16) $\exists x[(U(x) \wedge Follow(Q, x)) \wedge \exists y[(\neg U(y) \wedge (Follow(Q, y) \vee \neg Follow(Q, y))]$

Taking minimal pairs of sentences containing GA-marked and WA-marked subjects, this section has examined four possibilities resulting from the choice of these particles and the presence/absence of focal accents, and has provided logical representations for the sentences. All the differences in logical form must be ascribed to the semantics of these particles. The following sections will focus on the packaging functions of WA and explore the derivations of sentences with WA-marked phrases, which appear in sentence-initial and sentence-internal positions.

4 Deriving Topicalized Sentences: A First Approximation

Let us take a more syntactic approach to information packaging in order to automatically obtain interpretations considered in the previous section from syntactic parsing developed in a categorial framework. We will show that the topic particle WA performs the topic-comment articulation in two ways: (i) it picks up an expression denoting an entity/entities already familiar from previous discourse as a locus of information update; and (ii) it assembles the remaining parts of the sentence into an information unit which updates the information content of the entity/entities. In the previous section, we have represented the meaning of topicalized sentences with material implication. Let us adopt a tripartite structure like (17) for a unified semantic analysis of sentences with topics, following Hajičová et al. (1998), and write the topic-comment segmentation to save space as in (18)(see Nakamura 2006):

(17) TOPIC(x)
Restriction: Predicate1(...x...)
Nuclear Scope: Predicate2(...x...)

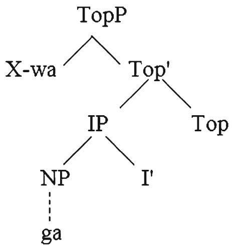
(18) TOP x[Px → Qx] (P ≠ ∅)

From this point on, let us consider how the parsing of sentences containing topics proceeds on a left-to-right, word-by-word basis, which corresponds to the process of information update. Semantic interpretations like (17) and (18) are derived from the resulting parse trees based on the principle of compositionality. In order to capture the syntactic and semantic properties of the topic marker WA from a dynamic perspective, it can be stated that the topic marker WA includes the context change potential (CCP) as in (19) in the sense of Heim (1983), where CCPs are defined as instructions specifying certain operations of context change. Suppose that a topic-marked phrase carries the so-called existence presupposition usually correlated with definite descriptions and that Heim's description of the CCP of "if" can be applied to the CCP of the topic marker WA.

(19) $c\ + \text{A-WA B} = \text{c}\backslash(\text{c} + \text{A}\backslash\text{c} + \text{A}\ + \text{B})$

The instruction of WA is to conjoin the current context with the presupposition A (denoted by a WA-marked expression, meaning that entities denoted by it are present in the CG), and the resulting context is updated by the new information B (denoted by the remaining parts of the sentence). We will implement such change in information state in the incremental parsing of topic sentences later. This description of the functions of WA appears to conform to the configuration in (20) assumed in the generative grammar tradition, where the differences between subjects and topics have been captured in the following tree structure:

(20)



The tree structure in (20) suggests that topic-marked elements show up only in the left-peripheral position of sentences, but this assumption seems dubious. As shown later, almost any element in any position can be marked with a WA-particle although topics occurring sentence-internally tend to be interpreted as conveying contrastive readings. Furthermore, the structure in (20) does not offer a mechanism that maps syntactic structures to intended semantic interpretations or information structures.

As a first approximation to syntactic analyses of topicalized sentences, following a brief introduction to Combinatory Categorial Grammar (CCG), we define the syntactic category for topics in order to incorporate the semantic properties into the lexical semantics of the particle WA, and show a sample derivation of a topic sentence that reflects its information structure with recourse to a strict characterization of the syntax-semantics interface. Let us limit ourselves here to reviewing a few necessary rules for this analysis. CCG is a mildly context-sensitive formalism, defining a set of combinatory rules to flexibly deal with standard and nonstandard surface constituency (including unbounded dependency) while maintaining direct compositionality. Only the following three rules of concatenation in CCG are relevant for our purposes:

(21)

$$\begin{array}{llllll}
\text{a.} & \text{X/Y}:f & \text{Y}:a & \rightarrow & \text{X}:fa & \\
 & \text{Y:a} & \text{X}\backslash\text{Y}:f & \rightarrow & \text{X:fa} & \\
\text{b.} & \text{X/Y}:g & \text{Y/Z}:f & \rightarrow_B & \text{X/Z}:gf & \\
 & \text{Y}\backslash\text{Z}:f & \text{X}\backslash\text{Y}:g & \rightarrow_B & \text{X}\backslash\text{Z}:gf & \\
\text{c.} & \text{X}:a \ \rightarrow_T & \text{T}\backslash(\text{T/X}):\lambda f.fa & \text{or} & \text{T}/(\text{T}\backslash\text{X}):\lambda f.fa &
\end{array}$$

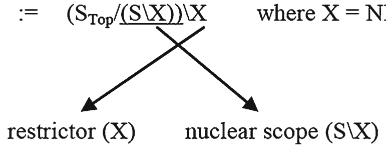
Following Steedman (1996, 2000), the ‘result leftmost’ notation is used here in which a rightward-combining functor over a domain Y into a range X is written as X/Y, whereas the corresponding leftward-combining functor is written as X\Y. (21a) is the rule of function application to concatenate expressions in the canonical order. An expression of functional category X/Y combines with an adjacent argument of category Y to yield a result of category X and interpretation $f a$, the result of applying f to a. This rule, for example, first combines a transitive verb with an object to yield a verb phrase, and then combines the verb phrase with a subject to produce a sentence. In this case, order-preserving associativity (re-bracketing) is assumed in grammar. The rule of function composition in (21b) allows a main function of category X/Y to combine with a subordinate function of category Y/Z to yield a new function of category X/Z. (21c) is the rule of type-raising, which we need to deal with a wide range of topicalized expressions in subsequent sections. The rule turns an argument category X into a functor over X. For instance, this operation converts a subject NP (which would normally be an argument to a verb phrase of category S\NP) into a function looking forward for a verb phrase, S/(S\NP), to produce a sentence.

Let us define the category for the topic marker WA as in (22), which serves as a (lexical) type shifter:

(22) Category for topicalized phrases

X-WA := S_{Top}/(S\X): TOP(x) [P(x) → Q(x)]

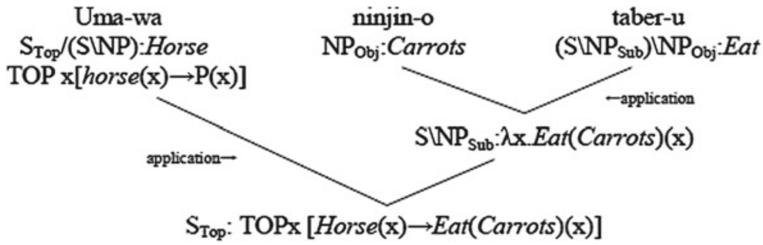
-WA := (S_{Top}/(S\X))\X where X = NP, PP, or S



The category of the topic particle WA in (22) lifts the category of a topicalized expression to a functor taking an open proposition, to yield a (matrix) sentence with the topic-comment structure. A topicalized expression of category X corresponds to a restrictor (i.e., the locus of information update) and the remaining part of the sentence of category S\X (which includes a gap corresponding to the topicalized expression) represents the nuclear scope, which updates or adds information to the context specified by the former. The category defined in (22) for WA eventually results in a tripartite information structure including the TOPIC operator. It is assumed that this information packaging is induced by another important property of WA as a concord particle called *kakari-joshi* in Japanese traditional grammar. Following this definition, an expression of any category can be extracted and marked with the topic marker WA. In terms of semantics, the category for WA indicates that a higher functor category is assigned to a topicalized expression. The meaning denoted by this raised category is compatible with the semantics of topics seen in Sect. 3. The derivation of sentence (23) can be shown as in (24):

(23) Uma-wa ninjin-o taber-u.
horse-Top carrots-Acc eat-Pres
'Horses eats carrots.'

(24)



As the category for the topic defined in (22) shows, the category that the topic marker WA first combines with is underspecified and WA can mark any constituent. As shown later, the remaining parts of a sentence can be packaged even if this operation results in a nonstandard constituent. Though representations like (24) properly show the derivations of sentences with topic-marked elements appearing left-peripherally in a compositional manner, this definition will have considerable difficulties dealing with sentences with sentence-internal topics because it gives rise to a proliferation of categories for topics, depending on the contexts in which topics appear. In addition, they do not reflect the dynamic process of information update suggested by the context change potential of WA in (19). Because Japanese is a head-final language in which functors usually apply to arguments backwards, the derivations require all of the elements to be present from the beginning, although nonincrementality is partially improved by raising categories of topicalized expressions. Assuming that information update should be performed in a time-linear manner in principle, this study presents the syntactic parsing of sentences with topics through a different approach (that proceeds incrementally from left to right), and shows that interpretations of sentences with topics (i.e., tripartite structures, as shown in (17) and (18) can be obtained as an independent process by reading the resulting syntactic representations called proof nets.

5 Proof Net Approach to the Topic-Comment Articulation

Recently, there have been a significant number of proof net approaches proposed in theoretical/computational linguistics. We adopt the categorial proof net analysis advocated by Morrill (2000, 2004, 2011). The theory of proof nets is based on Lambek categorial grammar, the derivations of which are usually presented via sequent calculus for concatenation, as in (25) (Morrill 2004):

(25)

a. $A \Rightarrow A$ id. $\qquad \dfrac{\Gamma \Rightarrow A \qquad \Delta(A) \Rightarrow B}{\Gamma(\Delta) \Rightarrow B}\,\text{Cut}$

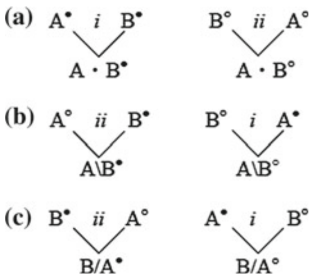
b. $\dfrac{\Gamma \Rightarrow A \qquad \Delta(B) \Rightarrow C}{(\Gamma, A\backslash B) \Rightarrow C}\,\backslash L \qquad \dfrac{A, \Gamma \Rightarrow B}{\Gamma \Rightarrow A\backslash B}\,\backslash R$

c. $\dfrac{\Gamma \Rightarrow A \qquad \Delta(B) \Rightarrow C}{\Delta(B/A, \Gamma) \Rightarrow C}\,/L \qquad \dfrac{\Gamma, A \Rightarrow B}{\Gamma \Rightarrow B/A}\,/R$

d. $\dfrac{\Gamma(A,B) \Rightarrow C}{\Gamma(A\bullet B) \Rightarrow C}\,\bullet L \qquad \dfrac{\Gamma \Rightarrow A \qquad \Delta \Rightarrow B}{\Gamma, \Delta \Rightarrow A\bullet B}\,\bullet R$

Morrill develops an incremental parsing algorithm under the slogan 'syntactic structures as proof nets,' which generates proof nets to analyze a wide range of constructions including discontinuous constituency. The concatenation rules can be reproduced for the construction of proof nets as follows. A polar category formula is shown as a Lambek categorial type labelled with input ($^\bullet$) or output ($^\circ$) polarity. A category formula derives a binary ordered tree (a logical link) in which the leaves are labeled with polar atoms. Some examples of logical links relevant to our analysis can be shown in (26). For the complete definition of categorial proof nets, see Morrill (2004, 2011)

(26)



The polarities indicate sequent sidedness, input for antecedent and output for succedent. The polarity propagation follows the sidedness of subformulas in the sequent rules, as illustrated in (26); in the anteedent (input) rule for A\B the subformula A goes in a succedent (output) and, in the succedent (output) rule, the subformula B goes in an antecedent (input). The labels *i* and *ii* indicate whether a rule with *i* or *ii* is unary or binary. In (26), note that, in the output links, the order of subformulas is switched and the category with output polarity is adjacent to the first label with input polarity. A frame is a list comprising a unique output polar type tree followed by input polar type trees. A proof net is the result of connecting by an identity link every leaf in a frame with a complementary leaf. According to Morrill (2004), proof nets must satisfy the correctness criteria in (27):

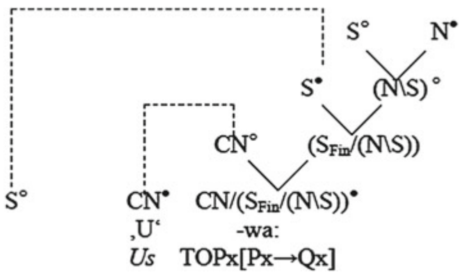
(27) Planarity The identity links are planar in the list of ordering.
Acyclicity Every cycle crosses both edges of some i-link.
No subtending No identity link connects the leftmost and rightmost descendent leaves of an output division node.

Building proof nets are performed on a word-by-word basis. Let us see how the parsing of sentence (6b) with a thematic topic (repeated here as (28) proceeds.

(28) Eigo-no seisyo-hou-ni-oite, U-wa Q-no atoni arawarer-u. (false)
English-Gen orthography-In U-Top Q-Gen after appear-Pres
'In English orthography, a 'U' always follows a 'Q'.'

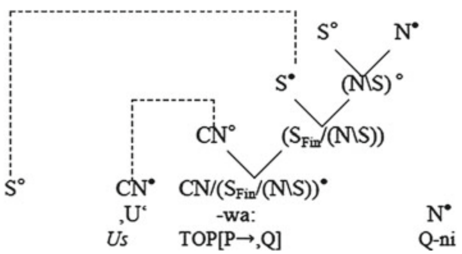
Initially, an S (with output polarity) is expected, and then when the first word, *U-wa* is perceived (let us ignore the sentence-initial adverbial in the derivation), there is no identity link in the structure. After the topic particle is processed, we have a partial proof net in (29):

(29)



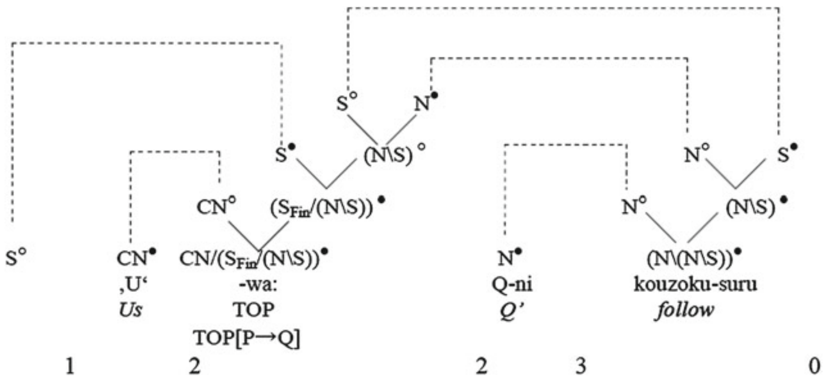
At this point, the identity links connect four complementary leaves–with two unmatched valencies (unresolved dependencies) remaining in the net. The number of unresolved dependencies indicates a measure of the course of memory load in optimal incremental processing. After processing the dative object *Q-ni*, we have the intermediate proof net (30) (We simplified its internal structure):

(30)



Because no leaf of the topic category can be linked with leaf $N^{\bullet}$ of the dative NP, we have three unmatched literals at this stage. After the verb, which the topic is in concord with, is processed, the unmatched leaves can be connected with proper complementary leaves to form the complete proof net analysis for sentence (28), as shown in (31). The number of unresolved dependencies (overarching identity links), referred to as a 'cut,' are written below the proof net:

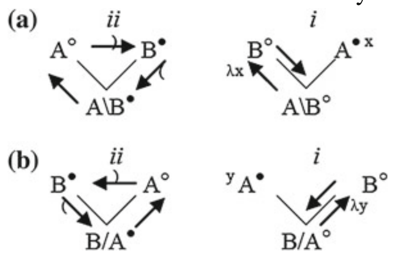
(31)



Within the categorial proof net approach, the sentence with the topic phrase can be successfully parsed in a time-linear manner, thus allowing underspecified dependencies at each intermediate stage of the parsing process from which the complexity profile of a sentence indicated by the numbers of unbounded dependencies (overarching identity links) can be obtained at each word boundary. Morrill (2004) suggests that maximal cuts and average cuts are relevant for evaluation of complexity of derivations. Notice the numbers of cuts written at the bottom of the proof net (16). Here the average cut is 1.6 and the maximal cut is 3. Section 7 will introduce the discontinuity operators, examine sentences in which topic-marked elements appear sentence-internally, and compare the complexity profiles of sentences with left-peripheral and sentence-internal topics.

Next, let us see how we can obtain the intended interpretation of sentence (28) from the semantic trip of the proof net (31) at no cost, according the core assumption regarding the tight syntax-semantics correspondence. Morrill (2004) defines the semantic trip travel instructions in (32) to yield the semantics of sentences from proof nets of which only four are relevant to our semantic analysis:

(32)



A semantic trip is the trip starting upwards from the unique output root S°, traveling on the net, and yielding the associated λ-terms as it proceeds, according to the instructions in (32). The trip bounces with the associated semantic form at the input roots and ends when it returns to the origin S°. From the proof net in (31), its meaning can be produced by a step-by-step interpretation process, which is omitted here due to space limitations. The normalized semantics resulting from the trip on proof net (31) should be something like (33), which is incorrect in terms of English orthography:

(33) a. $\lambda p \lambda q((\text{TOP}[p \to q])(\lambda x(U(x))(\lambda y(Follow(Q, y))$
b. $\text{TOP}[U(x) \to Follow(Q, x)]$

Following the semantic trip travel instructions in (32), this interpretation is automatically derived from the proof net analysis in (31). (33) represents the tripartite structure resulting from the topic-comment articulation executed by the topic marker WA, which is composed of the topic operator, the restrictor (locus of information update), and the nuclear scope (update of context). Although the same semantic representation can be derived by the combinatory rules of CCG, the categorial proof net is incrementally generated, thus matching the cognitive process in communication. The resulting proof net, in turn, provides complexity profiles of sentences indicated by the number of underspecified dependencies in the course of parsing. In sect. 7, we will suggest that complexity profiles obtained from parsed sentences offer a partial account to contrastiveness of sentence-internal topics, even though they result in the same truth conditions (as we have noted in Sect. 3).

6 Parsing of Cleft Constructions and Sentence-Internal Topics

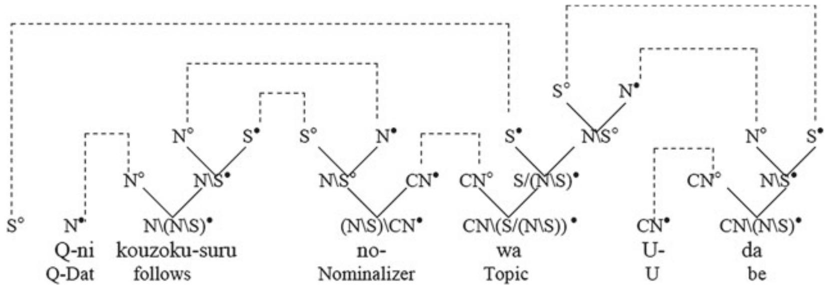
Let us now consider more complicated topicalization phenomena by focusing on sentences of higher complexities that include cleft constructions and multiple topic constructions. A typical cleft construction in Japanese is comprised of two parts; an open sentence denoting a presupposed property and a focused element followed by a copula. A presupposed open sentence is usually followed by the nominalizer *no* and the topic particle WA. A focused expression corresponding to the gap/variable in the former conveys new information. Case markers other than the nominative and accusative markers may intervene between a focus phrase and a copula. In sentence (34), the presupposition and focus parts of (28) are reversed and the existence of vowels that follow 'Q's is presupposed in (34):

(34)	Q-ni	kouzokusuru-	no-	WA	U-desu.
	Q-Dat	follow-Pres-	Nomizalizer-	TOP	U-Be-Pres

'It is 'U's that follow 'Q's.'

The proof net for (34) is shown in (35):

(35)



The category of the presupposed open proposition is converted to a common noun (CN) by the genitive case marker *no* in (35), which should be taken as a nominalization operator that maps property-denoting expressions onto property-correlates in the domain of entities (see Chierchia 1985; Partee 1986). Following Chierchia (1985), we indicate this operator as '$^\cap$', but let us tentatively assume that its type is still $\langle e, t\rangle$, not a simple individual (namely the semantic type of common nouns). The nominalized open proposition in (34) is translated as $^\cap Follow(Q)(x)$. Our semantic trip proceeds from the output polar formula S° to the input polar leaf S• of the particle WA, and then goes down to WA to translate it as usual, so the derived interpretation at this stage is $\lambda p \lambda q TOP[p \rightarrow q]$. It further proceeds to the nominalized open proposition marked with the nominalizer (which corresponds to the restrictor) and then turns to the focus part. Finally, the semantics as in (36) is obtained, thereby conveying the meaning that the set of individuals (vowels, in this context) appearing after 'Q's is included in the set of all occurrences of 'U,' which becomes true (assuming that the meaning of the copula *desu* is the identity function, $\lambda x.x$):

(36) TOP[$^\cap Follow(Q)(x) \rightarrow U(x)$]

Thus far we have only dealt with constructions in which topic phrases appear sentence-initially, which is naturally in harmony with the standard process of information update induced by the topic particle. However, we find numerous sentences with the topic phrases showing up in sentence-internal (and sentence-final) positions in Japanese texts or daily conversations, which can be be problematic for a simple tree structure analysis illustrated in (20). As an example, let us consider sentence (37):

(37) Context: A minister is questioned at the congress about the bribes that he alledgedly received. An article in a newspaper reports:

Daijin-wa	hisyo-ga	sono ken-wa	yoku	sitteiru-to	itta.
minister-TOP	secretary	that issue-TOP	well	know	said

'The minister said that his secretary knows that issue better.'

The elements conveying old and new information are intermixed in this sentence, where the two WA phrases indicate the familiarity with the lawmaker and the issue (bribery) among the interlocutors. The remaining (discontinuous) parts provide new information about the former, which appears to perform information packaging at the surface structure level. Notice that the sentence-internal WA phrase, the object in the embedded clause, does not necessarily evoke a contrastive meaning, which implies that there must be other aspects the secretary did not know about. Morrill (2004, 2011), who focused on a wide range of discontinuous constituency phenomena, proposes the discontinuity connectives in addition to familiar leftward (under) and rightward (over) slash operators, as shown in the following type map from syntactic to semantic types (Morrill 2004, p. 5, (1.16))

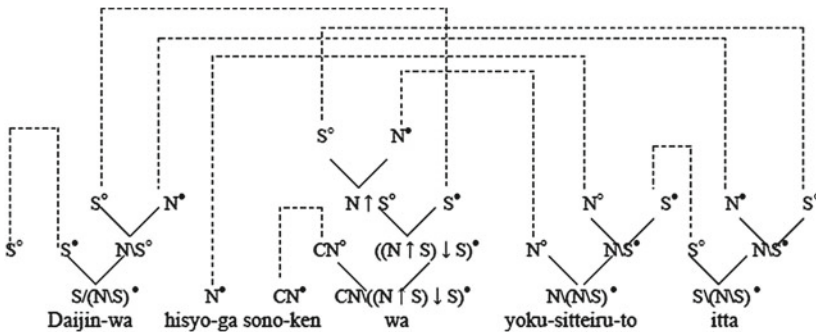
(38) a. $T(P) \quad = t(P)$ for atomic syntactic type P
b. $T(A\backslash C) \quad = T(A) \rightarrow T(C)$
c. $T(C/B) \quad = T(B) \rightarrow T(C)$
d. $T(A \downarrow C) = T(A) \rightarrow T(C)$ infix
e. $T(C \uparrow B) = T(B) \rightarrow T(C)$ extract

Morrill shows derivations of a wide range of discontinuous constituents in sequent and natural deduction presentations. Concerning the newly introduced connectives (infix and extract) that deal with discontinuity, the most important factor is to determine exact positions in order to insert separators into an expression of a discontinuous type. He strictly defines detailed procedures to derive discontinuous categories with one and more separators and to concatenate expressions of discontinuous types in sequent and labeled natural deduction. Although the present study omits these definitions here and refers readers to Morrill (2004, 2011) it would suffice to present the proof net for (37), where the second topic marked constituent combines with the complex discontinuous constituents. Assuming the discontinuous connectives in (38), the category of the topic particle WA appearing sentence-internally should be modified as in (39):

(39) WA $:= \quad X\backslash((S \uparrow N) \downarrow S) \quad (X = N, PP, S_{Comp}, \text{Conjunctive}, ...)$

The proof net for (37) should be something like (40) in which the WA-marked object of category $((S{\uparrow}N){\downarrow}S)$ combines the discontinuous complex expressions *daijin-wa hisyo-ga ... yoku sitteiru-to itta* of category $(S{\uparrow}N)$:

(40)



Note that the output category S° of the topic in the embedded clause is connected by the identity link with the input category $S^\bullet$ of the matrix verb, thereby indicating the coherence of the topic and the sentence-final verb, which is required by the lexical property of the topic marker as *kakari-joshi* or concord/coherence particle.[5] Because the modified category for the topic particle in (38) is allowed to combine with discontinuous constituents surrounding it, expressions of any category and location can be topicalized. It is important to note that we need to posit only one category for topic phrases wherever they occur, so we do not have lexical ambiguities for all occurrences of (thematic/referential) WA-phrases. After the semantic trip, we obtain the intended interpretation from the proof net, as seen in (41):

(41) TOP $x[Minister(x) \rightarrow$ (TOP $y[Issue(y) \rightarrow Said(Knew-well(y)(Secre.))(x)]]$
Topic segment Comment segment

(41) shows the result of information packaging invoked by the topic particle in which the sentence meaning is divided into two segments. The elements conveying old/predictable information are placed on the left-hand side and those conveying new/unpredictable information are located on the right-hand side. Actually, (41) merely represents the layers of topic-comment structures, and it is not exactly what this study wanted as topic-comment articulation. An appropriate information structure should be something like TOP x TOP $y[(Minister(x) \wedge Issue(y)) \rightarrow Said(Knew\text{-}well(y)(Secretary))(x)]$. However, because an additional device is necessary to map sentence (37) to a complete information structure, this particular issue remains unexplored here.

7 Contrastive Reading and Complexity Profiling

In the previous section, we saw that even deeply embedded sentence-internal topics can be given proper semantics by discontinuity connectives (and wrapping operations that insert an element of higher-order category between segments of an expression of discontinuous category), assuming the lexical property of WA as a concord particle. Even in Japanese, where topics are morphologically marked by the overt particle, there is a pronounced syntactic tendency for WA-marked constituents to be left-dislocated. In addition, it is widely assumed that sentence-initial topics can be thematic (anaphoric) or contrastive, whereas sentence-internal (or sentence-final) topics overwhelmingly convey contrastive readings (although it is sometimes difficult to distinguish the referential/anaphoric use and the contrastive use of topics). Sentence-internal topics appear to be incompatible with the natural process of information update and a speaker may use a nonstandard order of topic-comment constituency to convey additional meaning, thus implying the presence of alternative elements (Büring 1999). The categorial proof-net approach adopted in this study can account for preferences of time-linear information update process and implications such as contrastiveness carried by sentence-internal topics in terms of incremental

[5] This requirement on the particle WA provides an partial account to the so-called root phenomena of thematic WA.

processing of topicalized sentences and their complexity profiles obtained from the semantic trip of proof nets.[6]

How then can our approach associate nonstandard ordering of topics with contrastiveness? Let us compare the sentence-internal topic in (42a), where the underlined topic phrase conveys a contrastive reading, with the sentence-initial thematic topic in (42b):

(42) a.

Takusan-no	kankoo-kyaku-ga	<u>Kyoto-ni-WA</u>	maitoshi	yatte-kuru.
a lot of	tourists-NOM	Kyoto-To-Top	every-year	come-Pres.

‘A lot of tourists come to Kyoto every year.’

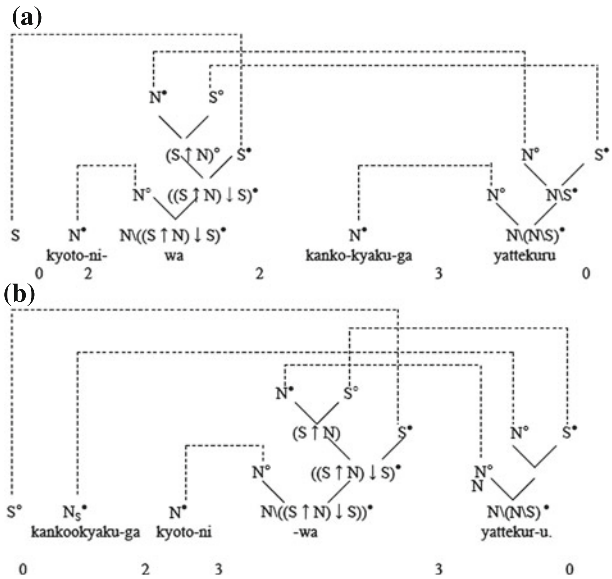
b.

Kyoto-ni-wa	takusan-no	kankoo-kyaku-ga	maitoshi	yattekur-u.
Kyoto-Loc-Top	a lot of	tourists-NOM	every-year	come-Pres

‘To Kyoto, a lot of tourists come every year.’

It should be noticed that even if WA-marked phrases are used contrastively, speakers assume hearers’ familiarity with their referents. As we have seen in Sect. 3, focalized topics implicitly evoke alternative elements (anti-exhaustive effect, in the sense of Kuroda 2005). Sentence (42a) contains the sentence-internal topic that forces a contrastive reading, thus implying that there are other cities which do not attract as many tourists, or at least that there are other cities to be compared with. (43a) and (43b) are the simplified proof nets for (42a) and (42b), respectively, from which their complexity profiles can be calculated, as shown below the nets.

(43)



As seen in (43a) and (43b), the maximal cuts are the same, 3. Note that sentence (43a) has undergone scrambling because the default order is $N_{Nom} + N_{Dat} + V$ in

[6] Here we ignore the effect of focal accents that usually fall on contrastive topics.

this case. The average cut, however, is 1.4 in the former, whereas it is 1.6 in the latter. Accordingly, the sentence-initial and sentence-final topics make a difference in the complexity profiles indicated by the numbers of dependencies between the polar leaves. Note that (all complexity profiles aside) the truth-conditional meanings of the sentences in (43) are almost identical, which implies the speaker's assumption of the hearer's familiarity with the referents of the topic. This difference in complexity profile seems to give a partial explanation regarding the general tendency of sentence-internal topics to induce contrastive interpretations, especially if we assume that the higher complexity of proof nets can be associated with canonical indications of markedness. Although more empirical evidence and formal devices are necessary to link the complexity of proof nets to the notion of contrastiveness as well as to incorporate some role of focal accents into grammar formalism, it seems natural to consider that markedness effects of sentence-internal topics can be related to the complexity of proof nets with more unresolved dependencies during incremental processing.

Conversely, all WA marked phrases share one important syntactic and semantic feature. WA must be associated with matrix predicates owing to its lexical property as a concord particle (*kakari-josi*), which requires the referents of WA marked expressions to be referentially or anaphorically interpreted. This property also explains the fact that WA marked phrases are generally reluctant to take place in referentially opaque positions. This study has shown that these syntactic and semantic properties realize information packaging, which yields tripartite information structures following Hajičová et al. (1998), and that information packaging and the resulting semantic readings can be shown in the proof net presentations, regardless of whether topics occur in sentence-initial or sentence-internal positions.

8 Conclusion

By noticing an important syntactic characteristic of the topic marker as a concord/coherence particle assumed in Japanese traditional grammar, this study argued that expressions marked with WA are taken to show a type of concord with sentence final verbs. The functions of the topic marker are two-fold: presenting a WA-marked expression as the locus of information update, and packaging the remaining parts of the sentence as an update of a hearer's information state. Therefore, the proper meaning/information structure of topicalized sentences is encoded as a lexical property of the topic particle.

In addition, this study shows how to process sentences with topics appearing in various positions incrementally, especially in terms of the categorial proof net approach. The incremental parsing strategy adopted here reflects the natural flow of information, implements topic-comment articulation, and provides intended interpretations that are automatically obtained from the semantic travels on the resulting nets. Unlike the parsing process of Dynamic Syntax approach (Kempson et al. 2001; Cann et al. 2005), which also realizes time-linear processing of an input sentence, the categorial proof net can retain the history of processing in which the complexity in understanding a sentence in question is explicitly shown. Such complexity can be indicated by the two types of cuts, a maximal cut and average cut. Morrill suggests that phenomena like garden pathing or scope preferences among multiple quantifiers can be given a new account based on the complexity profiles of processed sentences. In Japanese, topic-marked phrases can show up sentence-internally or sentence-finally, which appear to impose certain difficulties on the rigid configurational analyses of topics, as in generative grammar. Our flexible categorial grammar can assign a proper category defined by the discontinuity operators to topic-marked expressions and analyze a wide range of sentences including topics of various categories that appear in various positions.

Finally, categorial proof nets resulting from incremental parsing can yield multiple interpretations for sentences through different ways of connecting complementary leaves, the possibilities of which are constrained by the principles in (27) and measured by evaluating complexity profiles. This study showed that the tripartite information structures for sentences containing topics in various positions can be derived for free from syntactic proof nets. Furthermore, it suggested that contrastive interpretations, which non-sentence-initial topics usually evoke, can be associated with the relative complexities of the proof nets of such sentences.

References

Büring, D. (1999). Topic. In P. Bosch & R. van der Sandt (Eds.), *Focus—linguistic, cognitive, and computational perspectives* (pp. 142–165). Cambridge: Cambridge University Press.

Cann, R., Kempson, R., & Marten, L. (2005). *The dynamics of language: An introduction*. San Diego: Elsevier Inc.

Carlson, G. (1977). A unified analysis of the english bare plural. *Linguistics and Philosophy*, *1*, 413–458.

Chierchia, G. (1985). Formal semantics and the grammar of predication. *Linguistic Inquiry*, *16*, 417–443.

Diesing, M. (1992). *Indefinites*. Cambridge, MA: MIT Press.

Endriss, C. (2009). *Quantificational topics: A scopal treatment of exceptional wide scope phenomena*. Heidelberg: Springer.

Hajičová, E., Partee, B. H., & Sgall, P. (1998). *Topic-focus articulation, tripartite structures, and semantic content*. Dordrecht: Kluwer Academic Publishers.

Hasegawa, N. (1999). *Seisei-nihongogaku nyumon*. Tokyo: Taishukan.

Heim, I. (1983). File change semantics and the familiarity theory of definiteness. In Bauerle et al. (Eds.), *Meaning, use and interpretation of language* (pp. 164–189). Berlin: de Gruyter.

Hendriks, H. (2002). Information packaging: From cards to boxes. In K. van Deemster & R. Kibble (Eds.), *Information sharing* (pp. 1–33). Stanford: CSLI Publications.

Heycock, C. (2008). Japanese -WA, -GA, and information structure. In S. Miyagawa & M. Saito (Eds.), *The Oxford handbook of japanese linguistics* (pp. 54–83). New York: Oxford University Press.

Ippolito, M. (1997). On the meaning of only. *Journal of Semantics*, *25*, 45–91.

Kempson, R., Meyer-Vior, W., & Gabbay, D. (2001). *Dynamic syntax*. Oxford: Blackwell.

Krifka, M., Pelletier, F. J., Carlson, G. N., ter Meulen, A., Chierchia, G., & Link, G. (1995). Genericity: An introduction. In G. Carson & F. J. Pelletier (Eds.), *Generic book* (pp. 1–124). Chicago: University of Chicago Press.

Kuno, S. (1973). *The structure of the japanese language*. Cambridge, MA: MIT Press.

Kuroda, S.-Y. (1992). *Japanese syntax and semantics: Collected papers*. Dordrecht: Kluwer Academic Publishers.

Kuroda, S.-Y. (2005). Focusing on the matter of topic: A study of WA and GA in Japanese. *Journal of East Asian Linguistics*, *14*, 1–58.

Masuoka, T. (1987). *Meidai-no bunpoo-nihon-bunpoo-tosetsu*. Tokyo: Kuroshio Publishers.

Miyagawa, S. (1987). Wa and the WH phrase. In J. Hinds, S. K. Maynard, & S. Iwasaki (Eds.), *Perspectives on topicalization: The case of Japanese WA* (pp. 185–217). Amsterdam: John Benjamins.

Morrill, G. (2000). Incremental processing and acceptability. *Computational Linguistics*, *26*(3), 319–338.

Morrill, G. (2004). *Type logical grammar: The course reader distributed at ESSLLI 2004*. Nancy: The ESSLLI organizing committee.

Morrill, G. (2011). *Categorial grammar: Logical syntax, semantics, and processing*. New York: Oxford University Press.

Nakamura, H. (2006). Topic-comment articulation in Japanese: A categorial approach. In *Proceedings of the 20th PACLIC* (pp. 198–205). Beijing; Tsinghua University Press.

Noda, H. (1996). *WA to GA*. Tokyo, Kuroshio-Publishers.

Partee, B. H. (1986). Noun phrase interpretation and type-shifting principles. In J. Groenendijk, D. de Jongh, & M. Stokhof (Eds.), *Studies in discourse representation theory and the theory of generalized quantifiers* (pp. 93–123). Dordrecht: Foris Publications.

Portner, P., & Yabushita, K. (1998). The semantics and pragmatics of topic phrases. *Linguistics and Philosophy*, *21*, 117–157.

Portner, P., & Yabushita, K. (2001). Specific indefinites and the information structure theory of topics. *Journal of Semantics*, *18*, 271–297.

Roorda, D. (1992). Proof nets for lambek calculus. *Journal of Logic and Computation*, *2*, 211–231.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*, 75–116.

Shirai, K. (1985). *Keishiki-imiron nyuumon*. Tokyo: Sangyoo-Tosyo.

Steedman, M. (1996). *Surface structure and interpretation*. Cambridge, MA: MIT Press.

Steedman, M. (2000). *The syntactic process*. Cambridge, MA: MIT Press.

Takami, K., & Kamio, A. (1996). Topicalization and subjectivization in Japanese: Characterization and identificational information. *Lingua*, *99*, 207–235.

Tomioka, S. (2010). Contrastive topics operate on speech acts. In M. Zimmermann & C. Fery (Eds.), *Information structure: Theoretical, typological, and experimental perspectives* (pp. 115–138). Oxford: Oxford UP.

Uechi, A. (1996). Toward syntax-information mapping. In *Japanese/Korean linguistics 5* (pp. 387–403). Stanford: CSLI Publications.

Vallduví, E. (1992). The information component. Garland.

Vallduví, E. (1996). The linguistic realization of information packaging. *Linguistics*, *34*, 459–519.

A Note on the Projection of Appositives

Rick Nouwen

Abstract This article offers a thorough examination of the scopal properties of (mainly nominal) appositives. It is often descriptively noted that apposition is scopeless in the sense that its content escapes the scope of any operators that occur in the sentence the appositive is anchored in. I focus on exceptions to that characterisation and compare to what extent existing formal semantic analyses of apposition offer a handle on such exceptions. I then propose an analysis that predicts–rightly it turns out–that the exceptional cases, where appositives occur in the scope of a matrix operator, are part of a general pattern. Unfortunately, this analysis also over-generates severely. This issue, however, offers a new insight in the interaction between the scope of the appositive and the scope of its anchor. A final set of observations ultimately suggests that for a full understanding of appositive semantics it may be necessary to acknowledge the heterogeneity of the class of appositive constructions.

Keywords Appositives · Scope · Multidimensional semantics · Indefinites · Discourse anaphora

1 A Projection Problem for Appositives

It is often descriptively noted that appositives are *scopeless* in the sense that they escape the scope of any operator that occurs in the sentence the appositive is anchored

For valuable discussion, thanks are due to Jakub Dotlacil, Mitcho Erlewine, Katja Jasinskaja, Ora Matushansky, Claudia Poschmann, Hotze Rullmann, Philippe Schlenker, Yasutada Sudo, Ede Zimmermann as well as multiple reviewers. This work was supported by a grant from the Netherlands organisation for scientific research (NWO), which I hereby gratefully acknowledge. This note is in some ways a revision of my views in Nouwen 2007, a chapter I originally presented at LENS in Tokyo in 2006

R. Nouwen (✉)
Department of Language, Literature and Communication, Utrecht University, TLC/UiL-OTS
Trans 10, NL-3512JK, Utrecht, The Netherlands
e-mail: R.W.F.Nouwen@uu.nl

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*,
Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_10,

in (e.g. Potts 2005). For instance, the nominal appositive in (1) is interpreted outside the scope of the negation. That is, it entails that Jake is a famous boxer.

(1) It is not the case that Jake, a famous boxer, lives in Utrecht.

Nominal appositives (NAs) share this property with non-restrictive, or appositive, relative clauses (ARCs). The example in (2) also entails that Jake is a famous boxer.

(2) It is not the case that Jake, who is a famous boxer, lives in Utrecht.

It was noted by Wang et al. (2005), however, that there are cases where the scopal properties of appositive relative clauses (ARCs) and nominal appositives (NAs) diverge in interesting ways. Consider for instance the pair (3)/(4).

(3) If a professor, a famous one, publishes a book, he will make a lot of money.

(4) If a professor, who is famous, publishes a book, he will make a lot of money.

Both conditionals have a reading where *a professor* is interpreted as a specific indefinite. On that construal, the conditionals convey two things: (i) there is this professor such that if s/he publishes a book, s/he will make a lot of money; and (ii) the professor in question is famous. For (3), however, the most salient reading is one in which the NA is interpreted restrictively, yielding an interpretion synonymous to *if a famous professor publishes a book, he will make a lot of money*. This reading is absent from (4). In fact, the specific indefinite interpretation seems to be the only interpretation for (4).

There are two questions that need to be answered. The hard one is what exactly is responsible for the contrast between (3) and (4). I will have some speculations on this, but will not embark on a full-fledged attempt at answering this question in this squib. The second question, the one I want to focus on here in particular, is what accounts for the contrast between the wide-scope interpretation of the appositive in (1) and the narrow-scope interpretation in (3). In this chapter, I will evaluate the recent literature on the semantics and pragmatics of apposition, applying the proposals to the puzzling case of (3). Although I will not be able to offer a definitive analysis, I will draw some conclusions on the empirical reach of some proposed mechanisms for apposition in the literature. Ultimately, I will argue that the scopal behaviour observed in (3) is to be seen part of the general heterogeneity of appositives, in particular those with indefinite anchors.

The structure of the chapter is as follows. Section 2 gives a general introduction of the relation between the anchor and the appositive, focusing in particular on the alleged unavailability of quantified anchors. I discuss two types of theories of apposition: an account that treats appositives as predicates of the anchor due to Potts (2005) and an account that treats appositives as open propositions anaphoric to the anchor (Del Gobbo 2007; Nouwen 2007). Such theories have no immediate explanation for the restrictive interpretation of (3). In Sect. 3, I show that with the assumption of flexible attachment (Schlenker 2010a), an explanation does become available. Furthermore, the account I sketch in that section correctly predicts that restrictive readings of appositives occur more often than is usually assumed. At the same time, however, the account overgenerates in that it wrongly predicts restrictive readings to occur with definite or specific indefinites. In Sect. 4, I sketch some conclusions to be drawn from the overgeneration problem of the flexible attachment account. I moreover discuss some further issues to be considered.

2 Quantificational Anchors

It has been suggested in various parts of the literature on apposition that the anchors of appositive constructions are always referring expressions.[1] In particular, the suggestion is that *quantificational DPs* cannot anchor an appositive. For example, while the proper name is a felicitous anchor for the NA in (5-a), the quantificational expression *every boxer* in (5-b) appears not to be able to form an appositive construction.

(5) a. Jake, a famous boxer, took part in the event.
b. #Every boxer, a famous one, took part in the event.

There have been two kinds of approaches to account for this contrast. According to the theory in Potts (2005), there are compositional reasons why (5-b) is unacceptable: referential expressions and quantificational expressions differ in type, and (5-b) contains a type clash. A range of other analyses, e.g. del Gobbo (2003); Del Gobbo (2007) and Nouwen (2007), claim that (5-b) is out for purely semantic reasons, that concern a referential relation that needs to be established between anchor and appositive. I will now briefly discuss Potts's analysis and argue that the semantic approach more accurately covers the data.

2.1 Potts 2005

Potts (2005) treats apposition as an example of a phenomenon of two-dimensional content. That is, the fact that the appositive in (6) is scopeless is captured by assuming this sentence gives rise to two levels of content, containing two *independent* propositions, namely (6-a) and (6-b).

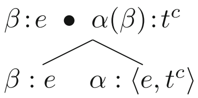
(6) It's not the case that Jake, a famous boxer, lives in Utrecht.
a. It's not the case that Jake lives in Utrecht.
b. Jake is a famous boxer.

[1] See Potts (2007) and Del Gobbo (2007) for discussion and references to such suggestions, which include Ross (1967), Rodman (1976), McCawley (1981), McCawley (1988) and Huddleston and Pullum (2002).

The content in (6-a) is the *at issue* content, which (in this case) ends up being asserted. In contrast, (6-b) is not at issue (and not asserted) but instead has the status of a *conventional implicature*. This means, for instance, that it is content that is not up for discussion in the subsequent discourse, whereas the at issue content is.

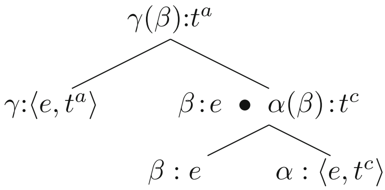
Potts proposes to distinguish the two levels of content in the type-system. He distinguishes two types t: t^a for the at issue level and t^c for the secondary, conventionally implicated level. An expression of type $\langle e, t^a\rangle$ is a predicate that when combined with an appropriate subject results in an at issue proposition, while an $\langle e, t^c\rangle$ predicate has special properties due to the following composition rule.

(7)



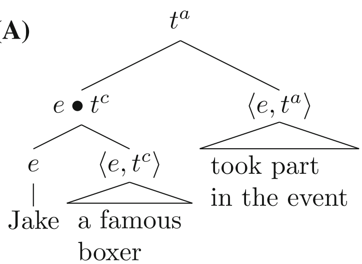
What this rule says is that when a type $\langle e, t^c\rangle$ predicate combines with its subject, it forms a complex object consisting of this subject and a propositional conventional implicature. Combinatorily, this complex $\bullet$-object behaves as if it were a regular type e. That is, the conventional implicature plays no role in the remainder of the derivation, as illustrated by the tree in (8).

(8)

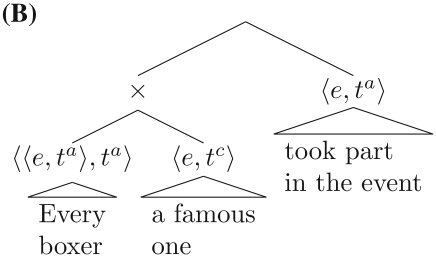


These combinatorics have consequences for the anchors of appositives, if, as Potts assumes, nominal appositives are $\langle e, t^c\rangle$ predicates. While tree A is fine, B presents a type clash.[2]

(A)



(B)



A crucial element in Potts's approach is that the anchor of the apposition is used twice in the compositional process, once as an argument in the matrix sentence and once

[2] The possibility of quantifier raising complicates this contrast. See Sect. 3.2.

as an argument of the predicate denoted by the appositive. I will now discuss some arguments that indicate that the relation between anchor and appositive is anaphoric, rather than compositional.

2.2 Appositives Have an e-Type Anaphoric Subject

It has been observed on several occasions that there is a striking similarity between nominal appositives and certain e-type anaphoric phenomena (Sells 1985; Demirdache 1991; del Gobbo 2003; Del Gobbo 2007; Nouwen 2007). One way to see this is to have a closer look at the parallel between the anaphoric potential of a quantificational noun phrase and its capacity to anchor an appositive.

While strong quantifiers can only bind singular variables in their scope, they can antecede semantically plural pronouns outside their scope (Kamp and Reyle 1993; van den Berg 1993; Nouwen 2003). To illustrate, (9-a) does not have an interpretation where for the majority of groups of students, each of the students in the group thinks the group is a good team. Beyond the sentence level, however, distributive quantifiers do license plural anaphora, as in (9-b).

(9) a. #Most students think they are a good team.
b. Most students came to the party. They had a good time.

Within the scope of the distributive quantifier singular variables may be bound, as in (10-a), where, despite its plural form '*they*', the pronoun ranges over single students. Singular anaphora is not possible outside the quantifier's scope.[3]

(10) a. Most students think they are smart.
b. Most Dutch men are arrogant. #He thinks he is very knowledgable.

Exactly these anaphoric possibilities are paralleled in the data on nominal appositives. Consider, for instance, the similarities between (11) and (12). Just like distributive quantifiers license plural but not singular discourse anaphora, they license plural but not singular nominal appositives.

---

[3] There are exceptions to this generalisation, namely cases of *telescoping* (Roberts 1987), such as Roberts' example (via Barbara Partee) in (i).

(i) Each degree candidate walked to the stage. He took his diploma from the Dean and returned to his seat.

It remains to be seen whether the parallel that I sketch below between discourse anaphora with quantifiers and the anchoring of an appositive extends to such exceptional cases.

(11) a. Jake lives in Utrecht. He is a famous boxer.
b. Every boxer took part in the event. #He is famous.
c. Every climber made it to the summit. They were all experienced adventurers.

(12) a. Jake, a famous boxer, lives in Utrecht.
b. Every Dutch boxer, #a famous one, took part in the event.
c. Every climber, all experienced adventurers, made it to the summit.

The observation that non-referential anchors may form appositive constructions under certain constraints (constraints that mirror those on discourse anaphora) forms a problem for the approach of Potts, where it is assumed that the anchor itself is the subject of the appositive.

2.3 Appositives as Propositions in Discourse

The data in the previous subsection suggests that the subject of an appositive is not the anchor, but rather a pronoun anaphoric to this anchor. Here, the relevant anaphoric relation is akin to *discourse anaphora*, which subsumes both coreference and e-type anaphora (to the exclusion of binding). A consequence of this view is that appositives are not predicates as in Potts (2005), but rather (open) propositions (Del Gobbo 2007; Nouwen 2007; Heringa 2012).

Del Gobbo (2007) assumes that appositives move at logical form to adjoin to a discourse node that dominates the matrix CP.[4] In other words, the interpretation of an appositive is exactly like that of a clause separated from the matrix clause in discourse. At logical form, (11) and (12) are indistinguishable.

The propositional account of apposition has no immediate explanation for the puzzling data of Wang et al. (2005). Clearly, (13-a) and (13-b) are not equivalent in interpretation. In fact, the pronoun in (13-b) cannot be anaphorically linked to the indefinite *a professor* unless it is interpreted as a specific indefinite.

(13) a. If a professor, a famous one, writes a book, he will make a lot of money.
b. If a professor writes a book, he will make a lot of money. He is famous.

[4] In Nouwen (2007), I also assumed that appositives are propositional and are linked to their anchor via anaphora. However, in that article I attempted to construct a framework that is faithful to the work of Potts in the sense that appositives are interpreted in situ. The result is what one could call a one-and-a-half-dimensional semantics: whilst the propositional content of an appositive is separated from the propositional content of the matrix sentence, appositive and host sentence are interpreted with respect to the same assignment function. One problem with the framework of Nouwen (2007) is that it is very difficult to define a semantics for negation without encoding in that semantics that its scope should ignore any appositive material. As far as I can see, similar problems extend to simpler systems based on similar ideas, as, for instance, the logic used in AnderBois et al. (2010).

Nevertheless, as we will see next, the propositional account opens up a way to account for the data. For if (13-a) is an exceptional case where the appositive proposition is interpreted not in discourse but *in situ*, the derivation of the desired interpretation falls out naturally.

3 Appositives as Propositions with Flexible Attachment

The analysis I have sketched in the previous section differs in one important respect from the approach of Potts. Whilst in Potts' system appositives are interpreted *in situ*, in the approach of Del Gobbo their representation is somehow syntactically removed from their surface scope domain.

Potts (2005) argues in detail for an *in situ* approach, pointing out for instance that in several languages the case marking of the appositive coincides with that of the anchor. Yet, the literature also offers some equally persuasive arguments against an *in situ* approach, like the observation that the anchor and the appositive do not appear to form a constituent (McCawley 1981; Schlenker 2010a).[5]

Schlenker (2010a, b) collects evidence for a mobile view of appositives: the attachment of appositives is flexible in nature. Schlenker's key argument is based on data from French. For instance, the absence of a condition C violation in (14) suggests that the relative clause is not interpreted in situ. Conversely, the subjunctive form in the appositive in (15) can only be accounted for if the appositive is interpreted with a low attachment (i.e. inside the scope of *conceivable*). The corresponding discourse version of (15), in (16), for instance, is unacceptable. The resulting reading for (15) is moreover one in which the relative clause is interpreted *within* the scope of *conceivable*.

(14) [Le President]$_i$ est si compliqué qu' il$_i$ a donné au ministre
the president is so complicated that he has given to-the minister
de la Justice, qui n'aime pas Sarkozy$_i$, une tache impossible.
of the justice, who NEG-like NEG S a task impossible

(15) Il est concevable que Jean ait appelé sa mére, qui ait
it is conceivable that J has-sub called his mother, who has-sub
appelé son avocat
called her lawyer

(16) Il est concevable que Jean ait appelé sa mére. *Elle ait
it is conceivable that J has-sub called his mother. She has-sub
appeé son avocat.
called her lawyer.

[5] That is, at the very least such data indicate that an *in situ* approach cannot maintain that the appositive is composed to the matrix sentence using the standard mode of composition.

Semantically, Schlenker (2010a) follows the propositional account presented in the previous section: appositives are propositions that are anaphorically linked to the anchor. Syntactically, however, Schlenker (2010a) interprets the data as pointing to the possibility of what he calls *flexible attachment*: appositions can be attached to any node of propositional type dominating the anchor (Schlenker 2010a).

The data in (13-a) could be seen as further suggestive evidence for the flexible nature of apposition, given that it shows that apart from the usual widest scope interpretation of appositive material, there are cases where a narrow scope interpretation surfaces. I now explore a way of using flexible attachment to account for such cases.

3.1 Flexible Attachment Predicts Restrictive Interpretations for Appositives

For simple examples like (17-a), Schlenker's approach does not differ much from del Gobbo's. That is, (17-a) is interpreted as (17-b) at logical form. (I am assuming here, with Schlenker, that the attached appositive proposition is interpreted conjunctively. I am hoping the logical form notation is otherwise self-explanatory, despite its informal presentation.)

(17) a. Jake, a famous boxer, lives in Utrecht.
  b. [*Jake*$_i$ *lives in Utrecht*] AND [*he*$_i$ *is famous*]

Flexible attachment provides a handle on the examples discussed by Wang et al. (2005). That is, (18-a) can be analysed as (18-b), which yields the desired interpretation.

(18) a. If a professor, a famous one, publishes a book, he will make a lot of money.
  b. *If* [*a professor*$_i$ *publishes a book* AND *he*$_i$ *is a famous professor*], *he*$_i$ *will make a lot of money*.

Further options for interpretation are limited. Like quantifiers, conditionals cannot bind singular variables in discourse.[6]

(19) If a professor publishes a book, he will often make a lot of money.
  #He is famous.

[6] Unlike quantifiers, they do not set up plural discourse referents either. The example in (i) cannot be interpreted as saying that the professor who publish a book (and make a lot of money) are famous.

(i) If a professor publishes a book, he will make a lot of money. #They are famous.

As a consequence of (19), an interpretation of (18-a) with high attachment of the appositive, outside the conditional is unacceptable, for this would give the following logical form[7]:

(20) [*If* [*a professor publishes a book*], [*he will make a lot of money*]] AND [$he_{\#}$ *is a famous professor*]

Here, and in what follows, I indicate the unacceptability of such logical forms by indexing the pronoun with a #. This indicates that no suitable reference resolution is possible.

What *is* an option for (18-a) is to interpret the indefinite as a specific one. In that case, high attachment is possible again.[8]

(21) [*There is this* $professor_i$] AND [*If* [he_i *publishes a book*], [he_i *will make a lot of money*]] AND [he_i *is a famous professor*]

Now consider the following contrast:

(22) a. If a professor, a famous one, publishes a book, he will make a lot of money.
b. Every professor, #a famous one, published a book.

Flexible attachment gives a way to explain this contrast too. In (22-a) there are two nodes of type *t* dominating the noun phrase that comes with the NA: the if-clause and the matrix sentence. We thus have the two possible logical forms (18-b) and (21). For (22-b) only one analysis is available, namely (23), since the matrix sentence itself is the only propositional node dominating anchor site.

(23) [*Every professor publishes a book*] AND [$he_{\#}$ *is a famous professor*]

As explained above, the form in (23) is infelicitous because strong quantifiers do not license singular discourse anaphora (cf. (11-b)). Since there is no other logical form available, the example is uninterpretable. If we add more structure to the quantificational subject, more logical forms are derived, since there are more propositional nodes. For instance, (24) does receive an interpretation, and it is indeed one in which the appositive is interpreted as part of the quantifier restrictor.

(24) Every professor who wrote a book, one on linguistics, is eligible for a sabbatical.

---

[7] Note that apart from high and low attachment, the flexible attachment approach allows, in principle, also for intermediate attachments, once there are enough suitable attachment nodes.

[8] This LF is intended as a theory-neutral representation of specific indefinites which captures both their wide-scope behaviour and their accessibility in discourse. Technically, it would resemble the referential account of specific indefinites of Fodor and Sag (1982), but nothing hinges on this. Any theory that accounts for the scopal and referential behaviour of specific indefinites will do.

So far, we have seen that flexible attachment can account for cases where appositives end up having a restrictive interpretation. Given this analysis, we now come to expect non-wide scope appositives in several other configurations, and, in fact, this is what we observe. For instance, NAs anchored to an indefinite NP in the scope of negation receive an interpretation *within* the scope of negation.[9]

(25) It is not the case that a boxer, a famous one, lives in this street.

This is explained readily by our assumptions. The two logical forms we derive are in (26).

(26) a. *It is not the case that* [*a boxer*$_i$ *lives in this street* AND *he*$_i$ *is famous*]
b. [*It is not the case that a boxer*$_i$ *lives in this street*] AND [*he*$_\#$ *is famous*]

In (26-b) the indefinite is not accessible to the pronoun and so (26-b) does not yield a felicitous interpretation. In contrast, (26-a) is interpretable and yields the observed interpretation.

3.2 Problems

Note that what is crucial in the examples under discussion here is that the NA is associated to noun phrases whose accessibility for pronominal anaphora is subject to scopal constraints. That is, the lack of a reading corresponding to the form in (26-b) is due to the fact that the referential reach of indefinites is limited to the scope of the negation. If we change the examples to include an appositive anchored to, say, a proper name, then the predictions and indeed the data change.

(27) It is not the case that Jake, a famous boxer, lives in this street.

Here, the wide scope interpretation, as in (28-b), *is* available. This is predicted too, since the referent of a proper name is globally accessible.

---

[9] All native speakers I consulted verified that this example indeed has a local interpretation. Nevertheless, an anonymous reviewer notes that the example improves a lot with a concessive *at least* in the appositive: *It is not the case that a boxer, at least a famous one, lives in this street.* I do not believe this observation in any way puts my main point into question: nominal appositives can end up being interpreted in the scope of negation. Nevertheless, and as the same reviewer notes, the intuition that concessive markers influence the way we interpret nominal appositives suggests that the relation between anchor and appositive is something that can be mediated. Below, I present more dramatic examples of how adverbial markers like "*in particular*" influence the interpretation of NAs. As the reviewer speculates, it could be we should see such markers as discourse markers, and thus we should see the projection of NAs as a phenomenon linked to discourse coherence. I concur that this is a valuable option worth investigating in the future. In fact, between the lines I do suggest below that a discourse perspective is probably helpful.

(28) a. *It is not the case that* [*Jake*$_i$ *lives in this street* AND *he*$_i$ *is famous*]
b. [*It is not the case that Jake*$_i$ *lives in this street*] AND [*he*$_i$ *is famous*]

Problematically, however, we now predict (27) to be ambiguous between the attested reading in (28-b) and the non-attested one in (28-a), which says that either Jake doesn't live in this street or he is not famous. This problem pops up in other cases too. For instance, for the case of (18-b), repeated here as (29), there is the logical form in (30), which yields an unavailable interpretation.

(29) If a professor, a famous one, publishes a book, he will make a lot of money.

(30) [*There is this professor*$_i$] AND [*If* [[*he*$_i$ *publishes a book*] AND [*he*$_i$ *is a famous professor*]] [*he*$_i$ *will make a lot of money*]]

Consistently, whenever a wide-scope interpretation is available, it blocks a possible competing narrow-scope one. Here is one way of summarising the observation:

(†) **The scope of the appositive is always at least as wide as that of its anchor, never narrower.**

This is an admittedly crude way of formulating the observation in terms of scope, which makes sense only if one thinks of referential expressions as expressions having wide(st) scope. (Thinking of a discourse representation theory-style framework is perhaps helpful here.) However, independent of the theoretical framework, as far as I can see, there is no obvious way of excluding restrictive readings of nominal appositives anchored by referential definites and indefinites. Perhaps (†) should be seen as part of a bigger observation, namely that like other assertorically inert information (where notions like presupposition or conventional implicature have been applied), the information in an appositive tends to be interpreted as scopally independent. That is, if the anchor is scopally independent, then the appositive should be too. I see only limited explanatory value in this.

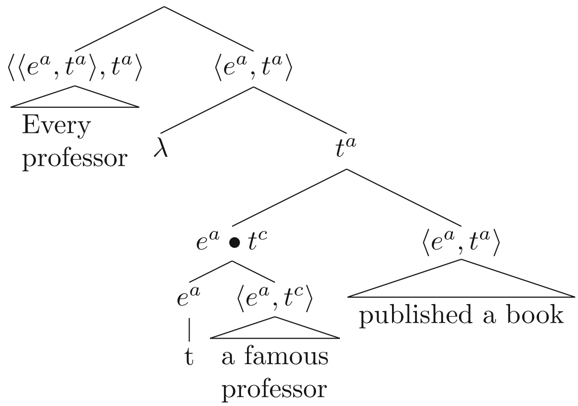
A related problem occurs with quantificational anchors. Above, I proposed that appositives in if-clauses can be read restrictively, because there is a propositional node they can attach to inside the conditional. In contrast, appositions attached to quantificational DPs have to outscope the quantifier in order to reach a propositional node. This picture ignores quantifier raising. That is, given flexible attachment, for (31-a) we could have the logical form (31-b), which would wrongly predict (31-a) has a restrictive reading.

(31) a. Every professor, #a famous one, published a book.
b. *Every professor*$_i$ [λ [[*t published a book*] AND *he*$_i$ *is famous*]]

I do not see any way of excluding quantifier raising in a configuration like (31-b). Moreover, we will need quantifier raising to deal with examples like (32). (Thanks to Katja Jasinskaja p.c. for making me aware of such examples.)

(32) Every boxer has a coach, a famous one.
a. [[*Every boxer*$_i$ [*has a coach*$_j$]] AND [*he*$_\#$ *is a famous coach*]]
b. [*Every boxer*$_i$ [λ [[*t has a coach*$_j$] AND [*he*$_j$ *is famous*]]]]

Interestingly, Potts's (2005) analysis of quantified appositives could be seen having an issue that is related to the over-generalisation problem of the flexible attachment approach. The type clash that accounts for the absence of quantified appositives in Potts' system (see sect. 2.1), disappears once quantifier raising is allowed for. For (31-a), the following derivation would be an option.

(33)


However, it is not immediately clear what this ends up meaning. Given Potts' interpretation of *parsetrees* (trees such as this one), (31-a) would denote the conjunction of *every professor* λ [*t published a book*] and the open proposition *t is a famous professor*. That is, Potts' system allows for no interaction between at issue and non-at issue content, and so the trace left behind by the raised quantifier, whilst bound on the at issue level, is left unbound—and thereby uninterpretable—on the secondary level. While this would correctly rule (31-a) out, it is questionable whether such rigid separation is maintainable, given the pleas for anaphoric accounts of appositives such as Del Gobbo (2007), but also given data that show more generally that cross-dimensional anaphoric links exist (e.g. Amaral et al. 2007 and AnderBois et al. 2010.)

4 Discussion

The flexible attachment approach I drafted above is successful in correctly predicting more restrictive readings with indefinites than just the conditionals cases observed by Wang et al. (2005). However, it fails to limit such readings to (non-specific) indefinites only. The obvious next step is to investigate what is so special about indefinites. One clue is that the nominal appositives we have been looking at have an indefinite form just like their anchor.

4.1 Correction

On the basis of the parallelism of form between anchor and appositive, Jasinskaja (p.c.) suggests that restrictive appositives such as that in (34-a) are part of an independent phenomenon, namely that of correction. That is, the idea would be that (34-a) is paraphrasable by (34-b).

(34) a. If a professor, a famous one, writes a book, he will make a lot of money.
b. If a professor—correction: a FAMOUS professor—writes a book, he will make a lot of money.

If (34-a), on its restrictive reading, is a case of correction, then it immediately follows why such readings are absent with other anchors: correction necessarily involves two formally parallel DPs.

(35) #I know every student—correction: a PROFESSOR.

An analysis of the data will now go as follows. A propositional account of apposition without flexible attachment (as in del Gobbo 2007) will correctly predict that no restrictive readings exist for nominal appositives. The restrictive reading that we do observe, i.e. the data in Wang et al. and the restrictive examples I discussed in the previous section, are exactly those cases where we confuse corrections with regular apposition.

It is not easy to evaluate the idea that (34-a) is a case of correction. I do believe, however, there is some suggestive evidence against such a claim. First of all, the restrictive readings do not always display the neat parallelism we find in (34-a). Consider (36):

(36) If two professors, both famous academics, write a book together, they will make a lot of money.

This example has the same kind of restrictive reading as the one Wang et al. identified for (34-a). However, it is not a likely case of correction. For instance, (37) is infelicitous as a correction.

(37) Today, Mary met two professors—correction: both FAMOUS (academics).

Heringa (2012) discusses some differences between apposition and correction that provide some further insights. Nominal appositives are infelicitous if they are not followed by comma intonation. In contrast, corrections have an optional pause. The data is as follows. (I indicate the comma intonation here with '—'.)

(38) a. John gave a student—his favourite one#(—) a book.
b. John gave a student—I mean a professor (—) a book.

As (39) shows, comma intonation cannot be removed from the Wang et al. data, suggesting that we are not dealing with a correction phenomenon here.

(39) If a professor—a famous one#(—) writes a book—he will make a lot of money.

For a full rebuttal of the reduction of (39) to correction, a more in-depth comparison to correction phenomena is in order. This falls squarely outside the scope of this short paper. The suggestion that the Wang et al. examples are not your regular nominal appositives, but some separate phenomenon that resembles it, is one worth pursuing, though.[10] As I will suggest in the remainder of this chapter, it looks like apposition is a heterogeneous phenomenon and that the scopal (or attachment) possibilities differ from subclass to subclass. Let me first go back to the beginning of this chapter and return to the contrast Wang et al. observed between nominal appositives and appositive relative clauses.

4.2 The Case of Appositive Relative Clauses

The flexible attachment approach to the projection behaviour of nominal appositives that I described above came in two parts: (i) semantically, nominal appositives are interpreted as conjuncts with a discourse anaphoric subject; (ii) syntactically, they may attach at any propositional node that dominates the anchor. These two ingredients, which I borrowed from Schlenker (2010a), account for the restrictive use of appositives anchored by indefinites, but they will not do for appositive relative clauses (ARCs). As I mentioned in the opening section of this chapter, ARCs resist the kind of restrictive readings of if-clauses that are observed with nominal appositives (Wang et al. 2005).

(40) If a professor, a famous one, publishes a book, he will make a lot of money.

(41) If a professor, who is famous, publishes a book, he will make a lot of money.

[10] An anonymous reviewer moreover suggests that the observation that the relevant examples of embedded nominal appositives may be marked in particular ways (e.g. *at least*, *correction*!, see footnote 9) could further support a view that such examples form a separate class. Such markers could for instance be seen to identify a particular coherence relation that goes hand in hand with a local interpretation of the appositive.

Schlenker's approach to ARCs differs from my adaptation of that approach for (40) only with respect to the pragmatic component. According to Schlenker, ARCs are locally *non-trivial*, while at the same time they are *translucent*.

Translucency (Schlenker 2010a)—content is *translucent* if and only if it can be made locally trivial by adding uncontroversial assumptions to the context
Pragmatic condition on ARCs (Schlenker 2010a)—Appositive relative clauses are translucent

Translucency means that we should be able to add unsurprising assumptions to the context that make the ARC locally trivial: that make the local context entail the ARC. Note that such a pragmatic principle cannot account for (40). If given some additional assumption it is locally entailed that the referent for *a professor* is famous, then we would not get a restrictive reading, but rather one in which all professors are famous. It is moreover difficult to see how in this case this involves an *unsurprising* assumption (since professors are not generally famous), and so it is that Schlenker's ARC proposal makes a prediction with respect to (41). Since it *is* possible to add an unsurprising assumption to the context about some specific professor, it is predicted that *a professor* in (41) is interpreted as a wide-scope indefinite. It appears that this is indeed the only available reading. In other words, an account involving flexible attachment could account for the difference between (40) and (41) by only imposing translucency on appositive relative clauses, but not on nominal appositives. Such an approach could perhaps also prove useful for contrasts such as the one in (42), from Klein (1977) (as cited in Heringa 2012).[11]

(42) a. There was a bird, a wild swan, in the air.
b. #There was a bird, which was a wild swan, in the air.

Note, however, that translucency does not generally enforce a specific reading on an embedded indefinite. Compare, for instance, (41–43). Here, a narrow scope reading is available for the indefinite *and* the appositive relative clause. This is because here it is perfectly possible to add to the context an assumption like *all students are required to fill in form B35*. Of course, given the universal nature of this assumption, the ARC does not *restrict* the conditional.

(43) If a student, who by the way is required to comply with all Statutory Policies, asks for legal advice, it is best practice to contact the school lawyer.

A similar example is (44).

(44) I wonder whether a presupposition, which by definition is part of the common ground, can ever be forcefully denied.

[11] An anonymous reviewer notes however that (i-b) improves if the ARC is not a predicate, as in *There was a bird, which we later learned was a wild swan, in the air.*

The indefinite *a presupposition* has narrow scope in the example. The appositive relative clause is locally interpreted, but not restrictive. That is, the ARC is interpreted to hold universally of presuppositions.

4.3 Conclusion

The general picture emerging from the discussion so far is that indefinites more easily resist a scopeless interpretation for appositives they anchor than non-indefinites do. While nominal appositives anchored by an indefinite appear to be unconstrained in the ability to be interpreted in situ, appositive relative clauses can only do so when they are translucent in the sense of Schlenker.

There exist further cases where appositives anchored by indefinites break with run of the mill properties of apposition. Take appositives marked by *in particular*:

(45) A reptile, in particular a snake, is a dangerous animal.

Importantly, (45) does not express *that reptiles are in particular snakes*. Instead, the appositive seems to combine with the VP: *a snake is a particularly dangerous animal*.[12] Note that such appositives are typically involved in generic statements, including conditionals with indefinite-containing if-clauses. Here is a variation on the Wang et al. example with *in particular*:

(46) If a professor, in particular a famous one, writes a book, he will make a lot of money.

Obviously, the low attachment strategy we used above will not work here. It is not at all clear what to do with *in particular* in the appositive conjunct:

(47) *If* [[*a professor*$_i$ *writes a book*] AND [*he*$_i$ *is in particular famous*]] [*he*$_i$ *will make a lot of money*]

An analysis needs to do justice to the fact that *in particular* has a degree function in such examples. In (46), *in particular* seems to express that famous professors make (even) more money after writing a book than other professors, just like (45) states that snakes are (even) more dangerous than other reptiles. For the case of (46), there is no attachment site that gives the correct interpretation. What is needed instead is to reconstruct the whole conditional, as in (48).

(48) [*If a professor writes a book, he will make a lot ofmoney*] *AND* [*particularly* [*If a famous*$_{focus}$ *professor writes a book, he will make a lot of money*]]

[12] Similar observations move Heringa (2012) to exclude such cases from his discussion of appositive constructions.

Obviously, no similar interpretation strategy is available for the Wang et al. conditional (49-a). This is because (49-a) does not entail (49-b).

(49) a. If a professor, a famous one, publishes a book, he will make a lot of money.
b. If a professor publishes a book, he will make a lot of money.

The upshot is that appositives can express various relations to indefinite anchors. These can for instance be inclusive, as in (46) or restrictive as in (49-a). It appears then that there is no *one* analysis for appositives, but that we should try to explain the relation between certain kinds of appositives and certain kinds of relations with the anchor. Whilst restrictive patterns of projection can be explained quite straightforwardly using the semantic and syntactic component of the proposal in Schlenker (2010a), the flexible attachment inherent in that proposal is limited to only a subclass of anchors. In other words, the puzzle put forward by Wang et al. (2005) remains, but as this final section has shown it turns out to be part of a much more general puzzle involving the various ways in which indefinites may relate appositively.

References

Amaral, P., Roberts, C., & Smith, E. A. (2007). Review of the logic of conventional implicatures by chris potts. *Linguistics and Philosophy*, *30*, 707–749.
AnderBois, S., Brasoveanu, A., & Henderson, R. (2010). Crossing the appositive / at issue meaning boundary. In *Proceedings of SALT 20* (pp. 328–346)
van den Berg, M. (1993). Full dynamic plural logic.
del Gobbo, F. (2003). Appositives at the Interface. Ph.D. thesis, University of California, Irvine.
del Gobbo, F. (2007). On the syntax and semantics of appositive relative clauses. In N. Dehé & Y. Kavalova (Eds.), *Parentheticals*. Amsterdam: John Benjamins.
Demirdache, H. (1991). Resumptive chains in restrictive relatives, appositives and dislocation. Ph.D. thesis, Massachusetts Institute of Techology.
Fodor, J. D., & Sag, I. (1982). Referential and quantificational indefinites. *Linguistics and Philosophy*, *5*, 355–398.
Heringa, H. (2012). Appositional constructions. Ph.D. thesis, University of Groningen.
Huddleston, R., & Pullum, G. (2002). *The Cambridge grammar of the english language*. Cambridge: Cambridge University Press.
Kamp, H., & Reyle, U. (1993). *From Discourse to Logic*. Dordrecht: D. Reidel.
Klein, M. (1977). Appositionele constructies in het Nederlands (appositive constructions in Dutch). Ph. D. thesis, Universiteit van Nijmegen.
McCawley, J. (1981). The syntax and semantics of english relative clauses. *Lingua*, *53*, 99–149.
McCawley, J. (1988). *The syntaxtic phenomena of english*. Chicago: Chicago University Press.
Nouwen, R. (2003). Complement anaphora and interpretation. *Journal of Semantics*, *20*(1), 73–113.
Nouwen, R. (2007). On appositives and dynamic binding. *Journal of Language and Computation*, *5*(1), 87–102.
Potts, C. (2005). The logic of conventional implicatures, volume 7 of Oxford studies in theoretical linguistics. Oxford University Press.

Potts, C. (2007). Conventional implicatures, a distinguished class of meanings. In G. Ramchand & C. Reiss (Eds.), *The Oxford Handbook of Linguistic Interfaces, Studies in Theoretical Linguistics* (pp. 475–501). Oxford: Oxford University Press.

Roberts, C. (1987). Modal Subordination, anaphora and distributivity. Ph. D. thesis, University of Massachussets, Amherst.

Rodman, R. (1976). Scope phenomena, movement transformations, and relative clauses. In B. Partee (Ed.), *Montague Grammar* (pp. 165–176). New York: Academic Press.

Ross, J. R. (1967). Constraints on Variables in Syntax. Ph. D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Schlenker, P. (2010a). Supplements within a unidimensional semantics i: scope. In *Proceedings of the 2009 Amsterdam Colloquium*.

Schlenker, P. (2009). Supplements within a unidimensional semantics ii: Epistemic status and projection. In *Proceedings of NELS 2010b*.

Sells, P. (1985). Restrictive and non-restrictive modification. Technical report, CSLI, Report 84–28, Stanford.

Wang, L., Reese, B., & McCready, E. (2005). The projection problem of nominal appositives. *Snippets*, *10*, 13–14.

Towards Computational Non-associative Lambek Lambda-Calculi for Formal Pragmatics

Norihiro Ogata

Abstract This paper will propose a new "mathematical foundation" for formal pragmatics, based on *Non-associative Lambek Lambda Calculi* (Wansing 1993; Buszkowski 1987, 1997) which are enhanced by substructural modalities $\underset{\{s\}}{!}$ for each substructurality s (Jacobs 1994; Morrill 1994), computational monads $\mathcal{T}$ as in Computational Lambda Calculi (Moggi 1991; Benton 1995; Benton and Wadler 1996; Goubault-Larrecq et al. 2008), and new type constructor $\overset{\alpha}{\multimap}$ for each α-*position*. I will show that the resulting system, called *the Computational Lambek* $\alpha\lambda$-*Calculus* ($\lambda_{c\alpha\odot!}$), is enough to treat formal pragmatics including information structures, underspecification, and communicative interactions.

Keywords Non-associative Lambek Lambda Calculi, Computational Monads, Formal Pragmatics

1 Introduction

The re-discovery of Lambek calculus (Lambek 1958) as a *resource logic* has led us to find the relation among Lambek calculus, *Lambek* λ-*calculus* (Buszkowski 1987; Wansing 1993; Polakow and Pfenning 1999; Buszkowski 1997; Restall 2000) or *ordered* λ-*calculus* (Gabbay 1996; Walker 2005), and *Monoidal Closed Categories* (Blute and Scott 2004) or *Residuated Categories* (Lambek 1968, 2004) which amounts to the relation among *linear* λ-*calculus* (Benton and Wadler 1996; Bierman 1995), *intuitionistic multiplicative linear logic* (Girard 1987), and *Symmet-*

This is the last paper written by Norry Ogata before his untimely death. As such, the editors have decided to leave it mostly unchanged, despite some insightful comments from reviewers (whom we would like to thank). We have only made occasional minor edits for clarity.

N. Ogata (✉)
Department of Language and Information Science, School of Language and Culture,
Osaka University, Osaka, Japan

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*, Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_11,

ric Monoidal Closed Categories (Blute and Scott 2004), and the classic relation among *intuitionistic logic*, the simply-typed λ-calculus, and *Cartesian Closed Categories* and by the *Curry-Howard Isomorphism*. In particular, the pure Non-Associative Lambek (*NLC*) λ-calculus generated by the Curry-Howard Isomorphism of the pure *NLC* λ-terms reflects data structures such as *binary trees*, or non-commutative non-associative groupoids. This fact makes us expect the modelling of operations of tree structures which are proposed in linguistics by pure Non-associative Lambek λ-terms (*NLC* λ-terms), where *trees can be regarded as bits of information which trees express*, i.e., contra Montagovian semantic types, type t is replaced by type $\mathsf{prop(osition)}$.

However, the *NLC* λ-terms are too impoverished to treat such linguistic operations of trees and other linguistic phenomena treated by natural language semantics and pragmatics such as information structures or presupposition generations/consumptions, since the λ-operators of the *NLC* λ-terms only bind the free variables in the leftmost or rightmost highest edges. Especially, the pragmatics of natural language including presupposition generation, consumption in phenomena such as anaphora, information structure such as focus, topic, and word order, underspecified interpretations with preference, and others, requires a variety of operations of trees. Therefore, we will expand the *NLC* as the following three steps:

1. I will generalize the *NLC* λ-calculus to Lambek $\alpha\lambda$-calculi, which are the class between Linear λ-calculi and the *NLC* λ-calculi,
2. The Lambek $\alpha\lambda$-calculus is extended by the addition of *computational polynomial functors* proposed by Moggi (1991) and others (Benton 1995; Benton and Wadler 1996; Goubault-Larrecq et al. 2008), called *the computational Lambek $\alpha\lambda$-calculus* $\lambda_{c\alpha\odot!}$,
3. Then I will apply it to problems on formal pragmatics such as information structure, presupposition generation and resolution, underspecification with preference, interaction, and so on.

2 Preliminaries

2.1 The Linear Lambda Calculus $\boldsymbol{\lambda_{\multimap\otimes!}}$ Generated by the Curry-Howard Isomorphism with Multiplicative Exponential Linear Logic

Definition 1 (*PTerm and PType of* $\lambda_{\multimap\otimes!}$) Let Var ($\ni x$) a set of variables, $*$ the unit, c a constant, $\frown$ the application in $\lambda_{\multimap\otimes!}$, then PTerm ($\ni t$) a set of pseudo-terms of $\lambda_{\multimap\otimes!}$ is defined as follows:

$$t ::= * \mid x \mid c \mid (\hat{\lambda} x : A.t) \mid (t_1 {}^\frown t_2) \mid (t_1 \otimes t_2) \mid (\mathit{let}\ t_1 = t_2\ \mathit{in}\ t_3) \mid (\mathit{derelict}\ t)$$
$$\mid (\mathit{copy}\ t\ \mathit{as}\ x, y\ \mathit{in}\ u) \mid (\mathit{discard}\ t_1\ \mathit{in}\ t_2) \mid$$
$$(\mathit{promote}\ t_1, \ldots, t_n\ \mathit{for}\ x_1, \ldots, x_n\ \mathit{in}\ t_3)$$

Let BType ($\ni \mathsf{a}$) be a set of basic types, and PType ($\ni A, B, C, \ldots$) a set of pseudo-types. Then A is defined as follows:

$$A ::= I \mid \mathsf{a} \mid (A_1 \multimap A_2) \mid (A_1 \otimes A_2) \mid !A$$

For each $t \in$ PTerm and $A \in$ PType, $t : A$ is a typing formula which means t is of type of A.

Definition 2 ($\vdash$ *of* $\lambda_{\multimap\otimes!}$) Let $\Gamma, \Delta, \ldots$, i.e., $\Gamma = (t_1 : A_1, \ldots, t_n : A_n)$, be sequences of typing formulas. In particular, $\Gamma; \Delta$ is the *concatenation* of Γ and Δ, i.e., if $\Gamma = (t_1 : A_1, \ldots, t_n : A_n)$ and $\Delta = (u_1 : B_1, \ldots, u_m : B_m)$, then $\Gamma; \Delta = (t_1 : A_1, \ldots, t_n : A_n, u_1 : B_1, \ldots, u_m : B_m)$.

Then $\Gamma \vdash t : A$ is a *type assignment* in $\lambda_{\multimap\otimes!}$, which means that in context Γ, pseudo-term t is of type of A in $\lambda_{\multimap\otimes!}$.

Then the type assignment of rules of $\lambda_{\multimap\otimes!}$ are defined as follows:

$$(Ax)\ \frac{}{x : A \vdash x : A} \qquad (Con)\ \frac{c \text{ is a constant of type of } A}{c : A \vdash c : A}$$

$$(I_{Unit})\ \frac{}{() \vdash * : I} \qquad (E_{Unit})\ \frac{\Gamma \vdash t : I \quad \Delta \vdash u : B}{\Delta[\Gamma/(* : I)] \vdash (\mathsf{let}\ t = *\ \mathsf{in}\ u) : B}$$

$$(I_{\multimap})\ \frac{\Gamma, x : A \vdash t : B}{\Gamma \vdash (\hat{\lambda} x : A.t) : A \multimap B} \qquad (E_{\multimap})\ \frac{\Gamma \vdash t : A \multimap B \quad \Delta \vdash u : A}{\Gamma; \Delta \vdash (t {}^\frown u) : B}$$

$$(I_{\otimes})\ \frac{\Gamma \vdash t : A \quad \Delta \vdash u : B}{\Gamma; \Delta \vdash (t \otimes u) : A \otimes B} \qquad (E_{\otimes})\ \frac{\Gamma \vdash t : A \otimes B \quad \Delta, x : A, y : B \vdash u : C}{\Gamma; \Delta \vdash (\mathsf{let}\ t = x \otimes y\ \mathsf{in}\ u) : C}$$

Furthermore, the structural rules of *structural modality* ! of $\lambda_{\multimap\otimes!}$ are defined as follows:

$$(P(ermutation))\ \frac{\Gamma, x : A, y : C, \Delta \vdash t : B}{\Gamma, y : C, x : A, \Delta \vdash t : B}$$

$$(L(eft\text{-}association))\ \frac{(\Gamma_1; (\Gamma_2; \Gamma_3)) \vdash t : B}{((\Gamma_1; \Gamma_2); \Gamma_3) \vdash t : B}$$

$$(R(ight\text{-}association))\ \frac{((\Gamma_1; \Gamma_2); \Gamma_3) \vdash t : B}{(\Gamma_1; (\Gamma_2; \Gamma_3)) \vdash t : B}$$

$$(C(ontraction))\ \frac{\Gamma \vdash t : !A \quad \Delta, x : !A, y : !A \vdash u : B}{\Gamma; \Delta \vdash (\mathsf{copy}\ t\ \mathsf{as}\ x, y\ \mathsf{in}\ u) : B}$$

$$(W(eakening))\ \frac{\Gamma \vdash t : !A \quad \Delta \vdash u : B}{\Gamma; \Delta \vdash (\mathsf{discard}\ t\ \mathsf{in}\ u) : B}$$

$$(Dereliction)\ \frac{\Gamma \vdash t : !B}{\Gamma \vdash (\mathsf{derelict}\ t) : B}$$

$$(Promotion)\ \frac{\{\Gamma_i \vdash t_i : !A_i\}_{i<m+1} \quad (y_0 : A_0, \ldots, y_m : A_m) \vdash u : B}{(x_0 : A_0, \ldots, x_m : A_m) \vdash (\mathsf{promote}\ t_0, \ldots, t_m\ \mathsf{for}\ y_0, \ldots, y_m\ \mathsf{in}\ u) : !B}$$

If there is Γ such that $\Gamma \vdash t : A$, then t is called a $\lambda_{\multimap\otimes!}$-term or simply, a term, and A a $\lambda_{\multimap\otimes!}$-type, or simply, a type.
The logic of types of $\lambda_{\multimap\otimes!}$ is called *intuitionistic multiplicative exponential linear logic* (*MELL*) and the correspondence between $\lambda_{\multimap\otimes!}$-terms and *MELL* is a form of *the Curry-Howard Isomorphism*.

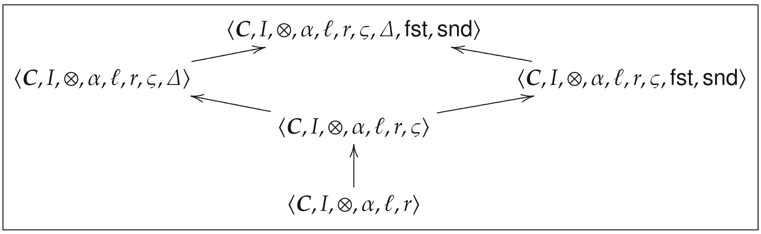
Definition 3 (β-*reductions of* $\lambda_{\multimap\otimes!}$) β-reduction rules of $\lambda_{\multimap\otimes!}$ are defined as follows:

- $\hat{\lambda} x$ properly binds x in t iff x occurs in t exactly once freely.
- $((\hat{\lambda} x : A.t)^\frown u) \rightarrow_\beta t[u/x]$ if $\hat{\lambda} x$ properly binds x in t.
- $(let * = * \; in \; t) \rightarrow_\beta t$
- $(let \; t \otimes u = x \otimes y \; in \; s) \rightarrow_\beta s[t/x, u/y]$
- $(derelict \; (promote \; \boldsymbol{u} \; for \; \boldsymbol{x} \; in \; t)) \rightarrow_\beta t[\boldsymbol{u}/\boldsymbol{x}]$
- $(discard \; (promote \; \boldsymbol{u} \; for \; \boldsymbol{x} \; in \; t) \; in \; s) \rightarrow_\beta (discard \; \boldsymbol{u} \; in \; s)$
- $(copy \; (promote \; \boldsymbol{t} \; for \; \boldsymbol{x} \; in \; u) \; as \; y, z \; in \; s)$
  $\rightarrow_\beta (copy \; \boldsymbol{t} \; as \; v, w \; in$ (s[($promote \; v \; for \; \boldsymbol{x} \; in \; u$)/y, $promote \; w \; for \; \boldsymbol{x} \; in \; u$/z])), where $\boldsymbol{t} = (t_0, \ldots, t_m)$, $\boldsymbol{x} = (x_0, \ldots, x_m)$, v, $w = \boldsymbol{x}$
- $(t^\frown(u^\frown s)) \rightarrow_\beta ((t^\frown u)^\frown s)$
- $((t^\frown u)^\frown s) \rightarrow_\beta (t^\frown(u^\frown s))$
- $(t \otimes (u \otimes s)) \rightarrow_\beta ((t \otimes u) \otimes s)$
- $((t \otimes u) \otimes s) \rightarrow_\beta (t \otimes (u \otimes s))$

2.2 The Generalized Jacobs Separation of ! and Introduction of ¡

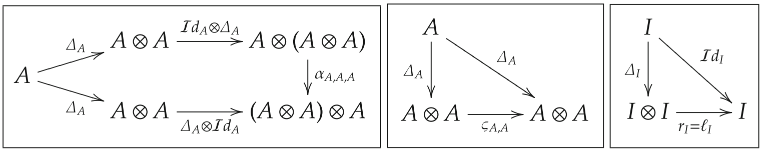
Jacobs (1994) separated ! in *MELL*, which is the structural modality and promotes *MELL* to the $\langle\rightarrow, \wedge\rangle$-fragment of Intuitionistic Logic, to two structural modalities: $\underset{w}{!}$ and $\underset{c}{!}$, which mean weakening and contraction, respectively. The type logic of $\lambda^{A}_{\multimap\otimes\underset{c\,w}{!}}$ amounts to the intuitionistic affine logic (*IAL*) and the type logic of $\lambda^{R}_{\multimap\otimes\underset{w\,c}{!}}$ to the intuitionistic relevant logic (*IRL*).

Jacobs (1994) proved that $! = \underset{w}{!}\underset{c}{!} = \underset{c}{!}\underset{w}{!}$ and

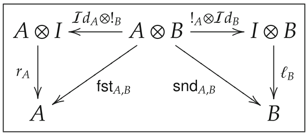


where $\langle \boldsymbol{C}, I, \otimes, \alpha, \ell, r\rangle$ is a *monoidal closed category* (see Appendix A), which is a categorical semantics of non-commutative linear logic and *LC*, $\langle \boldsymbol{C}, I, \otimes, \alpha, \ell, r, \varsigma\rangle$ a *symmetric monoidal closed category (smcc)* (see Appendix A), which is a categorical semantics of multiplicative linear logic, $\langle \boldsymbol{C}, \boldsymbol{I}, \otimes, \alpha, \ell, r, \varsigma, \Delta \rangle$ a smcc with diagonal Δ, which is a categorical semantics of relevant logic, $\langle \boldsymbol{C}, \boldsymbol{I}, \otimes, \alpha, \ell, r, \varsigma, \mathit{fst}, \mathit{snd} \rangle$ a smcc with projections $\mathit{fst}, \mathit{sndsnd}$, which is a categorical semantics of affine logic and called an *affine smcc*, and $\langle \boldsymbol{C}, \boldsymbol{I}, \otimes, \alpha, \ell, r, \varsigma, \Delta, \mathit{fst}, \mathit{snd} \rangle$ a smcc with diagonal and projections, which is a categorical semantics of intuitionistic logic, i.e., it is also a Cartesian closed category, satisfying the following conditions:

- $\Delta : \mathcal{I}d \to (-) \otimes (-)$ commutes the following three diagrams:



- $\mathit{fst}, \mathit{snd}$ are projections, which commute the following diagram:



and $\mathit{fst} \circ \Delta = \mathit{fst}$, $\mathit{snd} \circ \Delta = \mathit{snd}$, and $\Delta \circ (\mathit{fst} \otimes \mathit{snd}) = \mathcal{I}d$.

Similarly, ! in *MELL* can be separated to

$$! = \underset{wceplr}{!!!!!!} = \cdots = \underset{rlpecw}{!!!!!!},$$

where

1. the left-association modality $\underset{l}{!}$
2. the right-association modality $\underset{r}{!}$
3. the permutation modality $\underset{p}{!}$

$\underset{lr}{!!} = \underset{rl}{!!} = \bigcirc$ and $\underset{p}{!} = \triangle$ of Morrill's (1994) associativity structural modality and permutation substructural modality, respectively. I will write those composites of structural modalities as $\underset{s}{!}$ for any $s \in \{l, r, p, c, w, e\}$ and if $s = \emptyset$ then $[\![\underset{s}{!}]\!] = \mathcal{I}d$.

Semantically, as Jacobs (1994) notes, these structural modalities change $\otimes$ and "," as well, since $[\![\otimes]\!] = [\![,]\!]$. For example, in relevant logic $\otimes$ allows Contraction, so $\otimes$ and "," in relevant logic are written by $\underset{c}{\otimes}$ and $\overset{c}{,}$, respectively. Similarly, in affine logic $\otimes$ allows Weakening, so $\otimes$ and "," in affine logic are written by $\underset{w}{\otimes}$ and $\overset{w}{,}$, respectively. I will apply such a change of notation to the cases of l, r, p.

The Non-associative Lambek λ-calculi is promoted to the Lambek λ-calculus by $\underset{l}{!} \circ \underset{r}{!}$. Similarly, the Lambek λ-calculus is promoted to the Linear λ-calculus by $\underset{p}{!}$.

Furthermore, I introduce the deletion modalities of substructural modalities, written by $\underset{s}{¡}$ for each $s \subseteq \{l, r, p, e, c, w\}$. These modalities satisfy the following conditions:

1. $\underset{s}{¡}\underset{s'}{!} = \underset{s'-s}{!}$ if $s \subseteq s'$
2. $\underset{s}{¡}(a : A) = (a : A)$
3. $\underset{s}{¡}\Gamma \overset{s'}{,} \Delta = \underset{s}{¡}\Gamma \overset{s'-s}{,} \underset{s}{¡}\Delta$ if $s \subseteq s'$

In the categorical semantics of these type systems, α corresponds to *association rules*, ς the *permutation* rule. Therefore, the base categories of Lambek λ-calculus are monoidal closed categories such as

$$\langle \boldsymbol{C}, \boldsymbol{I}, \otimes, \alpha, \ell, r \rangle$$

and the base categories of Non-associative Lambek λ-calculus are *groupoid closed categories* such as

$$\langle \boldsymbol{C}, \boldsymbol{I}, \otimes, \ell, r \rangle.$$

Strictly speaking, structural modalities $\underset{s}{!}$ are modelled by *comonads* (See Appendix A) and *closedness* is modelled by *exponent functors* and their *evaluation functors*, *coevaluation functors* or *currification functors*.

These issues remain unsolved, although I will partially try to formalize them in Appendices A and B.

3 From Lambek Lambda Calculus to Non-associative Lambek Lambda Calculus to Lambek α Lambda Calculi

This section will pursue the framework "tree (with extra data structures) as (pragmatic) meaning" by exploiting Lambek λ-calculi.

3.1 Lambek Lambda Calculus $\boldsymbol{\lambda}_{/\bullet\backslash!}$

First, I introduce Lambek Lambda Calculus which is Curry-Howard-isomorphic with *LC* (Kanazawa 1992, 1999; Lambek 1958) or noncommutative linear logic (Yetter 1990; Abrusci 1991, 2002, 2003; Blute et al. 2002), which has already been proposed by Wansing (1993), Buszkowski (1987, 1997), and others, as follows:

Definition 4 (*PTerm and Ptype of* $\lambda_{/\bullet\backslash !}$) Let Var ($\ni x$) be a set of variables, $*$ the unit, c a constant, $\lambda^{>}$ the rightward lambda-operator, $\lambda^{<}$ the leftward lambda-operator, $^{>}$ the rightward-application, $^{<}$ the leftward-application, $\bullet$ the associative Lambek product in $\lambda_{/\bullet\backslash !}$, then PTerm ($\ni t$) a set of pseudo-terms of $\lambda_{/\bullet\backslash !}$ is defined as follows:

$$t ::= * \mid x \mid c \mid (\lambda^{>}x : A.t) \mid (\lambda^{<}x : A.t) \mid (t_1{}^{>}t_2) \mid (t_1{}^{<}t_2) \mid (t_1 \bullet t_2)$$
$$\mid (let\ t_1 = t_2\ in\ t_3) \mid (\underset{s}{derelict}\ t) \mid (copy\ t\ as\ x, y\ in\ s) \mid (\underset{s}{discard}\ t_1\ in\ t_2)$$
$$\mid (\underset{s}{promote}\ t_1, \ldots, t_n\ for\ x_1, \ldots, x_n\ in\ t_3)$$

Let BType ($\ni \mathsf{a}$) be a set of basic types, and PType ($\ni A, B, C, \ldots$) a set of pseudo-types. Then A is defined as follows:

$$A ::= I \mid \mathsf{a} \mid (B/A) \mid (A\backslash B) \mid (A \bullet B) \mid \underset{s}{!}A$$

For each $t \in$ PTerm and $A \in$ PType, $t : A$ is a typing formula which means t is the type of A.

Definition 5 ($\vdash$ *of* $\lambda_{/\bullet\backslash !}$) The type assignment rules of $\lambda_{/\bullet\backslash !}$ are defined by (Ax), (Con), (I_{Unit}), (E_{Unit}), plus following rules:

$$(I_/)\ \frac{\Gamma, x : A \vdash t : B}{\Gamma \vdash (\lambda^{>}x : A.t) : B/A} \qquad (I_\backslash)\ \frac{x : A, \Gamma \vdash t : B}{\Gamma \vdash (\lambda^{<}x : A.t) : A\backslash B}$$

$$(E_/)\ \frac{\Gamma \vdash t : B/A \quad \Delta \vdash u : A}{\Gamma; \Delta \vdash (t^{>}u) : B} \qquad (E_\backslash)\ \frac{\Delta \vdash u : A \quad \Gamma \vdash t : A/B}{\Delta; \Gamma \vdash (u^{<}t) : B}$$

$$(I_\bullet)\ \frac{\Gamma \vdash t : A \quad \Delta \vdash u : B}{\Gamma; \Delta \vdash (t \bullet u) : A \bullet B} \qquad (E_\bullet)\ \frac{\Gamma \vdash t : A \bullet B \quad \Delta, x : A, y : B \vdash u : C}{\Gamma; \Delta \vdash (\mathsf{let}\ t = x \bullet y\ \mathsf{in}\ u) : C}$$

As for substructural rules, the rules on p, c, w, e, i.e., (P), (E), (C) and (W) in Sect. 2.1 are required.

Definition 6 (*β-reductions of* $\lambda_{/\bullet\backslash !}$) β-reduction rules of $\lambda_{/\bullet\backslash !}$ are defined as follows:

- $\lambda^{>}x$ properly binds x in t iff x occurs in the rightmost place of t exactly once freely.
- $\lambda^{<}x$ properly binds x in t iff x occurs in the leftmost place of t exactly once freely.
- $((\lambda^{>}x : A.t)^{>}u) \rightarrow_\beta t[u/x]$ if $\lambda^{>}x$ properly binds x in t.
- $(u^{<}(\lambda^{<}x : A.t)) \rightarrow_\beta t[u/x]$ if $\lambda^{<}x$ properly binds x in t.
- $(let * = *\ in\ t) \rightarrow_\beta t$
- $(let\ t \bullet u = x \bullet y\ in\ s) \rightarrow_\beta s[t/x, u/y]$
- $(\underset{s}{derelict}\ \underset{s}{promote}\ \boldsymbol{u}\ for\ \boldsymbol{x}\ in\ t) \rightarrow_\beta t[\boldsymbol{u}/\boldsymbol{x}]$
- $(\underset{s}{discard}\ \underset{s}{promote}\ \boldsymbol{u}\ for\ \boldsymbol{x}\ in\ t\ in\ s) \rightarrow_\beta (\underset{s}{discard}\ \boldsymbol{u}\ in\ s)$

- $(copy\ \underset{s}{promote}\ \boldsymbol{t}\ for\ \boldsymbol{x}\ in\ u\ as\ y, z\ in\ s)$
  $\rightarrow_\beta (copy\ \boldsymbol{t}\ as\ v, w\ in\ (s[\underset{s}{promote}\ \boldsymbol{v}\ for\ \boldsymbol{x}\ in\ u/y,$
  $\underset{s}{promote}\ \boldsymbol{w}\ for\ \boldsymbol{x}\ in\ u/z])),$

  where $\boldsymbol{t} = (t_0, \ldots, t_m), \boldsymbol{x} = (x_0, \ldots, x_m), v, w = \boldsymbol{x}$
- $(t^\circ(u^\circ s)) \rightarrow_\beta ((t^\circ u)^\circ s)$, where $\circ \in \{^>, ^<\}$
- $((t^\circ u)^\circ s) \rightarrow_\beta (t^\circ(u^\circ s))$, where $\circ \in \{^>, ^<\}$
- $(t \bullet (u \bullet s)) \rightarrow_\beta ((t \bullet u) \bullet s)$
- $((t \bullet u) \bullet s) \rightarrow_\beta (t \bullet (u \bullet s))$

The associativity of $\lambda_{/\bullet\backslash!}$ makes $\lambda_{/\bullet\backslash!}$ inappropriate for a generic model of natural languages (trees). For example, "John loves Mary" is analyzed as a $\lambda_{/\bullet\backslash!}$-term (John$^<$(loves$^>$Mary)) which has its derivation proof:

$$(E_\backslash)\ \frac{\dfrac{}{\text{John} : \mathsf{e} \vdash \text{John} : \mathsf{e}} \qquad (E_/)\ \dfrac{\dfrac{}{\text{loves} : (\mathsf{e}\backslash\mathsf{prop}_E)/\mathsf{e} \vdash \text{loves} : (\mathsf{e}\backslash\mathsf{prop}_E)/\mathsf{e}} \qquad \dfrac{}{\text{Mary} : \mathsf{e} \vdash \text{Mary} : \mathsf{e}}}{\text{loves} : (\mathsf{e}\backslash\mathsf{prop}_E)/\mathsf{e}, \text{Mary} : \mathsf{e} \vdash (\text{loves}^>\text{Mary}) : \mathsf{e}\backslash\mathsf{prop}_E}}{\text{John} : \mathsf{e}, \text{loves} : (\mathsf{e}\backslash\mathsf{prop}_E)/\mathsf{e}, \text{Mary} : \mathsf{e} \vdash (\text{John}^<(\text{loves}^>\text{Mary})) : \mathsf{prop}_E}$$

but by the associativity, (John$^<$(loves$^>$Mary)) is β-contracted to ((John$^<$loves)$^>$ Mary), which is not a modelling which is not generally accepted, although sometimes such a contraction is convenient to treat some linguistic phenomena. Therefore, the associativity must be controlled.

3.2 Non-associative Lambek Lambda Calculus $\lambda_{-\bullet\bullet-\odot!}$

Next, I introduce Non-associative Lambek Lambda Calculus which is Curry-Howard-isomorphic with *NLC* (de Groote 1999; de Groote and Lamarche 2002; Wansing 2002). The similar idea has already been proposed as *type logic grammar* by Morrill (1994) and Moortgat (1997), but there is no strict Curry-Howard Isomorphism, since the terms which are typed in type logic grammar denote functions or constants in a Cartesian Closed Category, whereas the terms of $\boldsymbol{\lambda}_{-\bullet\bullet-\odot!}$ denote functions or constants in a Monoidal Closed Category or a Residuated Category.

Definition 7 (*PTerm and PType of* $\lambda_{-\bullet\bullet-\odot!}$) Let Var ($\ni x$) be a set of variables, $*$ the unit, c a constant, $\lambda^\triangleright$ the rightward lambda-operator, $\lambda^\triangleleft$ the leftward lambda-operator, $^\triangleright$ the rightward-application, $^\triangleleft$ the leftward-application, $\odot$ the non-associative Lambek product in $\lambda_{-\bullet\bullet-\odot!}$, then PTerm ($\ni t$) a set of pseudo-terms of $\lambda_{-\bullet\bullet-\odot!}$ is defined as follows:

$$t ::= * \mid x \mid c \mid (\lambda^{\triangleright}x : A.t) \mid (\lambda^{\triangleleft}x : A.t) \mid (t_1{}^{\triangleright}t_2) \mid (t_1{}^{\triangleleft}t_2) \mid (t_1 \odot t_2)$$
$$\mid (let\ t_1 = t_2\ in\ t_3) \mid (\underset{s}{derelict}\ t) \mid (copy\ t\ as\ x, y\ in\ s)$$
$$\mid (\underset{s}{discard}\ t_1\ in\ t_2) \mid (\underset{s}{promote}\ t_1, \ldots, t_n\ for\ x_1, \ldots, x_n\ in\ t_3)$$

Let BType ($\ni \mathsf{a}$) be a set of basic types, and PType ($\ni A, B, C, \ldots$) a set of pseudo-types. Then A is defined as follows:

$$A ::= I \mid \mathsf{a} \mid (B \bullet\!\!- A) \mid (A -\!\!\bullet B) \mid (A \odot B) \mid \underset{s}{!}A$$

For each $t \in$ PTerm and $A \in$ PType, $t : A$ is a typing formula which means t is of type of A.

Definition 8 (*The Contexts of* $\lambda_{-\bullet\bullet-\odot!}$) The contexts CONT ($\ni \Gamma, \Delta, \ldots$) of $\lambda_{-\bullet\bullet-\odot!}$ is defined as follows:

$$\Gamma ::= () \mid (a : A, \Gamma) \mid (\Gamma, a : A) \mid (\Gamma, \Delta)$$

We must notice that (Γ, Δ) is not the fusion or concatenation of Γ and Δ but a tree which has its left branch Γ and its right branch Δ.

The type assignment rules consist of (Ax), (Con), (I_{Unit}), (E_{Unit}) and the following rules:

Definition 9 ($\vdash$ *of* $\lambda_{-\bullet\bullet-\odot!}$)

$$(I_{\bullet-})\ \frac{\Gamma, x : A \vdash t : B}{\Gamma \vdash (\lambda^{\triangleright}x : A.t) : B \bullet\!\!- A} \qquad (I_{-\bullet})\ \frac{x : A, \Gamma \vdash t : B}{\Gamma \vdash (\lambda^{\triangleleft}x : A.t) : A -\!\!\bullet B}$$

$$(E_{\bullet-})\ \frac{\Gamma \vdash t : B \bullet\!\!- A \quad \Delta \vdash u : A}{(\Gamma, \Delta) \vdash (t^{\triangleright}u) : B} \qquad (E_{-\bullet})\ \frac{\Delta \vdash u : A \quad \Gamma \vdash t : A -\!\!\bullet B}{(\Delta, \Gamma) \vdash (u^{\triangleleft}t) : B}$$

$$(I_{\odot})\ \frac{\Gamma \vdash t : A \quad \Delta \vdash u : B}{(\Gamma, \Delta) \vdash (t \odot u) : A \odot B} \qquad (E_{\odot})\ \frac{\Gamma \vdash t : A \odot B \quad \Delta, (x : A, y : B) \vdash u : C}{(\Gamma, \Delta) \vdash (\mathsf{let}\ t = x \odot y\ \mathsf{in}\ u) : C}$$

As for substructural rules, the rules on l, r, p, c, w, e, i.e., (L), (R), (P), (E), (C) and (W) in Sect. 2.1 are required.

Definition 10 (*β-reductions of* $\lambda_{-\bullet\bullet-\odot!}$) β-reduction rules of $\lambda_{-\bullet\bullet-\odot!}$ are defined as follows:

- $RE(x) = x$, $RE(c) = c$, $RE(t^{\triangleright}u) = u$, $RE(t \odot u) = u$;
- $LE(x) = x$, $LE(c) = c$, $LE(t^{\triangleleft}u) = t$, $RE(t \odot u) = t$;
- $\lambda^{\triangleright}x$ properly binds x in t iff x occurs in $RE(t)$ exactly once freely.
- $\lambda^{\triangleleft}x$ properly binds x in t iff x occurs in $LE(t)$ exactly once freely.
- $((\lambda^{\triangleright}x : A.t)^{\triangleright}u) \rightarrow_{\beta} t[u/x]$ if $\lambda^{\triangleright}x$ properly binds x in t.
- $(u^{\triangleleft}(\lambda^{\triangleleft}x : A.t)) \rightarrow_{\beta} t[u/x]$ if $\lambda^{\triangleleft}x$ properly binds x in t.
- $(let * = *\ in\ t) \rightarrow_{\beta} t$
- $(let\ t \odot u = x \odot y\ in\ s) \rightarrow_{\beta} s[t/x, u/y]$
- $(\underset{s}{derelict}\ (\underset{s}{promote}\ \boldsymbol{u}\ for\ \boldsymbol{x}\ in\ t)) \rightarrow_{\beta} t[\boldsymbol{u}/\boldsymbol{x}]$

- $(\underset{s}{discard}\ (\underset{s}{promote}\ \boldsymbol{u}\ for\ \boldsymbol{x}\ in\ t)\ in\ s) \rightarrow_\beta (\underset{s}{discard}\ \boldsymbol{u}\ in\ s)$
- $(copy\ (\underset{s}{promote}\ \boldsymbol{t}\ for\ \boldsymbol{x}\ in\ u)\ as\ y, z\ in\ s)$
  $\rightarrow_\beta (copy\ \boldsymbol{t}\ as\ v, w\ in\ (s[(\underset{s}{promote}\ \boldsymbol{v}\ for\ \boldsymbol{x}\ in\ u)/y, (\underset{s}{promote}\ \boldsymbol{w}\ for\ \boldsymbol{x}\ in\ u/z])),$
  where $\boldsymbol{t} = (t_0, \ldots, t_m), \boldsymbol{x} = (x_0, \ldots, x_m), v, w = \boldsymbol{x}$

$\lambda_{\multimap \bullet \bullet \multimap \odot !}$-terms are highly restricted even though they are enhanced by structural modalities such as $\underset{r}{!}$ and $\underset{l}{!}$. For example, suppose "What ...?" is an operator which binds the so-called *trace* of the operator in sentence "What did John give to Mary?", i.e., the "What ...?"-operator must bind x in:

$$t = (\text{John}^{\triangleleft}((\text{give}^{\triangleright} x)^{\triangleright}(\text{to}^{\triangleright}\text{Mary})))$$

The operator cannot bind x in t regardless of what structural modalities are used, since $\lambda^{\triangleright} x$ and $\lambda^{\triangleleft} x$ can only bind variables on "the right edge" and "the left edge", respectively. In this case, x must be bound by operators exactly once. In this sense, the λ-calculus required for modelling natural language has a flavor of the linear λ-calculus, and the condition that the bound position is decided has a flavor of the Lambek λ-calculus. Therefore, what we need for modelling natural language is located between the linear λ-calculus and the Lambek λ-calculus in the sense of sensitivities of resource consumptions and the bound positions. In the Sect. 3.3, I propose a λ-calculus, called Non-associative Lambek $\alpha\lambda$-calculus, which is resource-sensitive and allows binding into a variety of positions. Such positions will be called α-positions.

3.3 The Lambek $\alpha\lambda$-Calculus $\lambda_{\alpha \odot !}$

First, I define α in the term "α-position", as follows:

Definition 11 α is an arbitrary position in a tree, defined as follows:

$$\alpha\ (\in \mathcal{A}) ::= \circ \ |\ \triangleleft\ |\ \triangleright\ |\ \alpha_1\alpha_2\ |\ *\,\alpha\ |\ op\ \alpha$$

where $*$ means an arbitrary finite iteration, and $op\ \alpha$ is the opposite of α, which is defined by recursion on α, as follows:

1. $op \circ = \circ$
2. $op \triangleleft = \triangleright$
3. $op \triangleright = \triangleleft$
4. $(op\ \alpha_1\alpha_2) = (op\ \alpha_1)(op\ \alpha_2)$
5. $op\ (*\alpha) = *(op\ \alpha)$
6. $op\ (op\ \alpha) = \alpha$

Definition 12 (*PTerm and PType of* $\lambda_{\alpha\odot!}$). Let Var ($\ni x$) a set of variables, $*$ the unit, c a constant, λ^{α} the α-position binding lambda-operator, $^{\alpha}$ the α-position-application, $\odot$ the non-associative Lambek product in $\lambda_{\alpha\odot!}$, then PTerm ($\ni t$) a set of pseudo-terms of $\lambda_{\alpha\odot!}$ is defined as follows:

$$t ::= * \mid x \mid c \mid (\lambda^{\alpha} x : A.t) \mid (t_1{}^{\alpha} t_2) \mid \mid (t_1 \overset{\alpha}{\odot} t_2) \mid (let\ t_1 = t_2\ in_{\alpha}\ t_3) \mid (\underset{s}{derelict}\ t)$$
$$\mid (copy\ t\ as\ x, y\ in\ s) \mid (\underset{s}{discard}\ t_1\ in\ t_2) \mid$$
$$(\underset{s}{promote}\ t_1, \ldots, t_n\ for\ x_1, \ldots, x_n\ in\ t_3)$$

Let BType ($\ni \mathsf{a}$) be a set of basic types, and PType ($\ni A, B, C, \ldots$) a set of pseudo-types. Then A is defined as follows:

$$A ::= I \mid \mathsf{a} \mid (A \overset{\alpha}{\multimap} B) \mid (A \overset{\alpha}{\odot} B) \mid \underset{s}{!}A$$

For each $t \in$ PTerm and $A \in$ PType, $t : A$ is a typing formula which means t is of the type A.

Definition 13 (*The Contexts of* $\lambda_{\alpha\odot!}$) The contexts CONT ($\ni \Gamma, \Delta, \ldots$) of $\lambda_{\alpha\odot!}$ is defined as follows:

$$\Gamma ::= () \mid (a : A, \Gamma) \mid (\Gamma, a : A) \mid (\Gamma, \Delta) \mid (\Gamma \overset{\alpha}{;} \Delta)$$

where $\Gamma = \Gamma_1 \overset{\alpha}{;} \Gamma_2$ means the α-position fusion, i.e., Γ_2 is at the α-position in Γ and *the α-position in Γ* is defined by recursion on α, as follows:

- if $\alpha = \circ$ and $\Gamma = t : A$, then $t : A$ is at the α-position in Γ;
- if $\alpha = \triangleleft$ and $\Gamma = (t : A, \Delta)$, then $t : A$ is at the α-position in Γ;
- if $\alpha = \triangleright$ and $\Gamma = (\Delta, t : A)$, then $t : A$ is at the α-position in Γ;
- if $\alpha = \alpha_1\alpha_2\triangleleft$, $\Gamma = (\Gamma_1, (t : A, \Gamma_2))$, Γ_1 is the α_1-position, and Γ_2 is the α_2-position, then $t : A$ is at the α-position in Γ;
- if $\alpha = \alpha_1\alpha_2\triangleright$, $\Gamma = (\Gamma_1, (\Gamma_2, t : A))$, Γ_1 is the α_1-position, and Γ_2 is the α_2-position, then $t : A$ is at the α-position in Γ;
- if for some $n < \omega$, $\alpha = *\alpha'$ and $\Gamma = ({}_0\Gamma_0, ({}_1\Gamma_1, ({}_2\Gamma_2, \ldots ({}_{n-1}\Gamma_{n-1}, \Gamma_n)_{n-1} \ldots)_2)_1)_0$ and for each $i < n+1$, Γ_i is at the α'-position in $({}_i\Delta)$ then all Γ_i $(i < n+1)$ are at the α-positions in Γ;
- if for some $n < \omega$, $\alpha = *\alpha'$ and $\Gamma = ({}_0({}_1({}_2 \ldots ({}_{n-1}\Gamma_n, \Gamma_{n-1})_{n-1} \ldots, \Gamma_2)_2, \Gamma_1)_1, \Gamma_0)_0$ and for each $i < n+1$, Γ_i is at the α'-position in $({}_i\Delta)$ then all Γ_i $(i < n+1)$ are at the α-positions in Γ;

and $\Gamma(x : A)_{\alpha}$ means that $x : A$ occurs once and only once at the α-position in Γ.

Definition 14 ($\vdash$ *of* $\lambda_{\alpha\odot!}$) The type assignment rules consists of (Ax), (Con), (I_{Unit}), (E_{Unit}), $(I_{\odot})$, $(E_{\odot})$ and the following rules:

$$(E_{\overset{\alpha}{\multimap}})\ \frac{\Gamma \vdash t : A \quad \Delta \vdash u : A \overset{\alpha}{\multimap} B}{(\Gamma \overset{\alpha}{,} \Delta) \vdash (t^{\alpha} u) : B} \qquad (I_{\overset{\alpha}{\multimap}})\ \frac{\Gamma(x : A)_{\alpha} \vdash t : B}{\Gamma \vdash (\lambda^{\alpha} x : A.t) : A \overset{\alpha}{\multimap} B}$$

$$(E_{\overset{\alpha}{\odot}})\ \frac{\Gamma \vdash t : A \overset{\alpha}{\odot} B \quad \Delta \overset{\alpha}{,} (x : A, y : B) \vdash u : C}{(\Gamma \overset{\alpha}{,} \Delta) \vdash (\mathit{let}\ t = x \overset{\alpha}{\odot} y\ \mathit{in}_{\alpha}\ u) : C} \qquad (I_{\overset{\alpha}{\odot}})\ \frac{\Gamma \vdash t : A \quad \Delta \vdash u : B}{(\Gamma \overset{\alpha}{,} \Delta) \vdash (t \overset{\alpha}{\odot} u) : A \overset{\alpha}{\odot} B}$$

Definition 15 (*β-reductions of* $\lambda_{\alpha\odot!}$) β-reduction rules of $\lambda_{\alpha\odot!}$ are defined as follows:

- t occurs at the α-position in u iff:
  - if $\alpha = \circ$ and $t \in \mathrm{Var} \cup \mathrm{Con}$, then t is at the α-position in t;
  - if $\alpha = \triangleleft$ and $t = (s^{\triangleleft} u)$ then s is at the α-position in t;
  - if $\alpha = \triangleright$ and $t = (s^{\triangleright} u)$ then u is at the α-position in t;
  - if $\alpha = \alpha_1 \alpha_2 \triangleleft$, $t = (s^{\alpha_1}(t'^{\alpha_2} u))$, then t' is at the α-position in t;
  - if $\alpha = \alpha_1 \alpha_2 \triangleright$, $t = (s^{\alpha_1}(t'^{\alpha_2} u))$, then u is at the α-position in t;
  - if for some $n < \omega$, $\alpha = *\alpha'$ and $t = (u^{\alpha}(t_0{}^{\alpha}(\ldots(t_{n-1}{}^{\alpha} t_n)\ldots)))$, then all $t_0, \ldots, t_n$ are at the α-positions in t;
- $\lambda^{\alpha} x$ properly binds x in t iff x occurs in the α-position in t exactly once freely.
- $((\lambda^{\triangleright} x : A.t)^{\triangleright} u) \to_{\beta} t[u/x]$ if $\lambda^{\triangleright} x$ properly binds x in t.
- $(u^{\triangleleft}(\lambda^{\triangleleft} x : A.t)) \to_{\beta} t[u/x]$ if $\lambda^{\triangleleft} x$ properly binds x in t.
- $(\mathit{let} * = *\ \mathit{in}_{\alpha}\ t) \to_{\beta} t$
- $(\mathit{let}\ t \overset{\alpha}{\odot} u = x \overset{\alpha}{\odot} y\ \mathit{in}_{\alpha}\ s) \to_{\beta} s[t/x, u/y]$ if $x \overset{\alpha}{\odot} y$ is in the α-position in s;
- $(\underset{s}{\mathit{derelict}}\ (\underset{s}{\mathit{promote}}\ \boldsymbol{u}\ \mathit{for}\ \boldsymbol{x}\ \mathit{in}\ t)) \to_{\beta} t[\boldsymbol{u}/\boldsymbol{x}]$
- $(\underset{s}{\mathit{discard}}\ (\underset{s}{\mathit{promote}}\ \boldsymbol{u}\ \mathit{for}\ \boldsymbol{x}\ \mathit{in}\ t)\ \mathit{in}\ s) \to_{\beta} (\underset{s}{\mathit{discard}}\ \boldsymbol{u}\ \mathit{in}\ s)$
- $(\underset{s}{\mathit{copy}}\ (\underset{s}{\mathit{promote}}\ \boldsymbol{t}\ \mathit{for}\ \boldsymbol{x}\ \mathit{in}\ u)\ \mathit{as}\ y, z\ \mathit{in}\ s)$
  $\to_{\beta} (\underset{s}{\mathit{copy}}\ \boldsymbol{t}\ \mathit{as}\ v, w\ \mathit{in}\ (s[(\underset{s}{\mathit{promote}}\ \boldsymbol{v}\ \mathit{for}\ \boldsymbol{x}\ \mathit{in}\ u)/y, (\underset{s}{\mathit{promote}}\ \boldsymbol{w}\ \mathit{for}\ \boldsymbol{x}\ \mathit{in}\ u)/z]))$,
  where $\boldsymbol{t} = (t_0, \ldots, t_m)$, $\boldsymbol{x} = (x_0, \ldots, x_m)$, $v, w = \boldsymbol{x}$

These system is categorically modelled by *α-residuated groupoidal closed categories*, as follows:

Definition 16 A α-residuated groupoidal closed category is a tuple

$$\langle \boldsymbol{C}, \{\overset{\alpha}{\odot}\}_{\alpha \in \mathcal{A}}, \{\overset{\alpha}{\multimap}\}_{\alpha \in \mathcal{A}}, I, \ell, r \rangle,$$

where

- the functors $\overset{\alpha}{\odot} : \boldsymbol{C} \times \boldsymbol{C} \to \boldsymbol{C}$ and the functors $\overset{\alpha}{\multimap} : \boldsymbol{C} \times \boldsymbol{C} \to \boldsymbol{C}$ for each $\alpha \in \mathcal{A}$, satisfying

$$\boldsymbol{C}(C \overset{\alpha}{\odot} A, B) \cong \boldsymbol{C}(C, A \overset{\alpha}{\multimap} B);$$

  In particular, there are evaluation map ev_{α} and coevaluation map $coev_{\alpha}$ for each $\alpha \in \mathcal{A}$ s.t.

$$\mathsf{ev}_\alpha : (A \overset{\alpha}{\multimap} B) \overset{\alpha}{\odot} A \to B$$
$$\mathsf{coev}_\alpha : C \to (A \overset{\alpha}{\multimap} C \overset{\alpha}{\odot} A)$$

- the unit object $I \in \boldsymbol{ObC}$;
- the isomorphism functors $\ell_A : I \overset{\alpha}{\odot} A \overset{\cong}{\to} A$, $r_A : A \overset{\alpha}{\odot} I \overset{\cong}{\to} A$ for each $A \in \boldsymbol{ObC}$ and $\alpha \in \mathcal{A}$, satisfying
$$\ell_I = r_I : I \overset{\alpha}{\odot} I \overset{\cong}{\to} I$$
- providing a semantics of the Lambek $\alpha\lambda$-calculus, as follows ((I_{Unit}), (E_{Unit}), (Ax), and (Con) are the same with other systems):

$$(E_{\overset{\alpha}{\multimap}}) \left[\!\!\left[\frac{\Gamma \vdash t : A \quad \Delta \vdash u : A \overset{\alpha}{\multimap} B}{(\Gamma \overset{\alpha}{,} \Delta) \vdash (t^\alpha u) : B} \right]\!\!\right] = \mathsf{ev}_{\alpha,[\![A]\!],[\![B]\!]} \circ ([\![\Gamma \vdash t : A]\!] \overset{\alpha}{\odot} [\![\Delta \vdash u : A \overset{\alpha}{\multimap} B]\!])$$

$$(I_{\overset{\alpha}{\multimap}}) \left[\!\!\left[\frac{\Gamma(x : A)_\alpha \vdash t : B}{\Gamma(* : I)_\alpha \vdash (\lambda^\alpha x : A.t) : A \overset{\alpha}{\multimap} B} \right]\!\!\right] = \mathsf{coev}_\alpha([\![\Gamma(x : A)_\alpha \vdash t : B]\!])$$

$$(I_{\overset{\alpha}{\odot}}) \left[\!\!\left[\frac{\Gamma \vdash t : A \quad \Delta \vdash u : B}{(\Gamma \overset{\alpha}{,} \Delta) \vdash (t \overset{\alpha}{\odot} u) : A \overset{\alpha}{\odot} B} \right]\!\!\right] = [\![\Gamma \vdash t : A]\!] \overset{\alpha}{\odot} [\![\Delta \vdash u : B]\!]$$

$$(E_{\overset{\alpha}{\odot}}) \left[\!\!\left[\frac{\Gamma \vdash t : A \overset{\alpha}{\odot} B \quad \Delta(x : A, y : B)_\alpha \vdash u : C}{\Delta(\Gamma)_\alpha \vdash (\mathsf{let}\ t = x \overset{\alpha}{\odot} y\ \mathsf{in}_\alpha\ u) : C} \right]\!\!\right]$$

$$= \mathsf{ev}_{\alpha,[\![A]\!],[\![B]\!]} \circ (\lambda^\alpha [\![z]\!] \in [\![A \odot]\!]. \left[\!\!\left[\Delta(z)_\alpha \vdash t : A \overset{\alpha}{\odot} B \right]\!\!\right] \overset{\alpha}{\odot} [\![\Gamma \vdash u : C]\!])$$

$\lambda_{\alpha\odot!}$ is sufficiently expressive to treat the data structures and related operations of natural language. However, $\lambda_{\alpha\odot!}$ is not expressive enough to treat the formal pragmatics of natural language, which requires interactions with context and discourse situations, although the formal pragmatics is deeply connected to word order, anaphora, ellipsis, and prosody, which can also be properties of syntactic structures. I propose a combination of the Lambek $\alpha\lambda$-calculus with the computational λ-calculus (Moggi 1991), which includes *monads* of *side-effects*, which can treat the interactions between contexts or discourse situations and syntactic structures. The technical difficulty of this combination is that Lambek $\alpha\lambda$-calculus corresponds to residuated groupoid categories, whereas the computational λ-calculus corresponds to Cartesian closed categories. However, Benton (1995) and others (Benton and Wadler 1996; Goubault-Larrecq et al. 2008) have already provided a combined system of Linear λ-calculus and the computational λ-calculus at the category-theoretic level. According to Benton and others (Benton 1995; Benton and Wadler 1996), if $\boldsymbol{C} \underset{\mathcal{F}}{\overset{\mathcal{G}}{\rightleftarrows}} \boldsymbol{S}$ with $\mathcal{F}$ left adjoint to $\mathcal{G}$, where $\boldsymbol{C}$ is a Cartesian closed category and $\boldsymbol{S}$ a symmetric monoidal closed category and the adjunction is symmetric monoidal, then $! = \mathcal{F} \circ \mathcal{G}$ and $\mathcal{T} = \mathcal{G} \circ \mathcal{F}$, where $\mathcal{T}$ is the functor of a computational monad in the sense of the computational λ-calculi. I call this property the *Plotkin-Benton-Wadler Property*.

Between each residuated α groupoid category $\boldsymbol{G}$ and each symmetric monoidal closed category $\mathcal{S}$, one of the Generalized Jacobs properties $\boldsymbol{G} \underset{\underset{l,r,e,p}{!}}{\overset{\overset{!}{l,r,e,p}}{\rightleftarrows}} \boldsymbol{S}$ holds. Therefore, by combining the Plotkin-Benton-Wadler Property and the Generalized Jacobs property between residuated α groupoid categories and symmetric monoidal closed categories, we can introduce computational monads and make the computational Lambek α calculus. But its detailed proof must be left for future work.

3.4 The Computational Lambek $\alpha\lambda$-Calculus $\lambda_{c\alpha\odot!}$

The computational Lambek $\alpha\lambda$-terms and the Lambek $\alpha\lambda$-types are also extended as follows: ($q \in \mathbb{Q}$)

$$t \in PTerm, \quad A, B \in PType$$

$$t ::= \cdots \mid q \mid (\mathsf{let}_{\mathcal{F}} x = t_1 \;\mathsf{in}\; t_2) \mid (t_1 t_2) \mid [t]_{\mathcal{F}} \mid \langle t_1, t_2 \rangle \mid (\mathit{fst}\; t) \mid (\mathit{snd}\; t) \mid (\mu x.t)$$

$$A ::= \cdots \mid \mathbf{x} \mid (A \to B) \mid (A \times B) \mid \mathcal{F}\, A \mid (\mu\mathbf{x}.A)$$

where $\mathbf{x}$ is a variable over types and $\mathcal{F}$ is a computational functor with its computational monad $\langle \mathcal{F}, \eta^{\mathcal{F}}, \mu^{\mathcal{F}} \rangle$ or their equivalent Kleisli triple $\langle \mathcal{F}, \eta^{\mathcal{F}}, {*}^{\mathcal{F}} \rangle$ (see Appendix A).

The type assignment rules are:

$$(let\mathcal{F})\ \frac{\Gamma \vdash t_1 : \mathcal{F}A_1 \quad x : A_1 \vdash t_2 : \mathcal{F}A_2}{\Gamma \vdash \mathsf{let}_{\mathcal{F}} x = t_1 \;\mathsf{in}\; t_2 : \mathcal{F}A_2} \qquad (unit\mathcal{F})\ \frac{\Gamma \vdash t : A}{\Gamma \vdash [t]_{\mathcal{F}} : \mathcal{F}A}$$

The categorical semantics of these rules are defined as follows:

$$\left[\!\!\left[\frac{\Gamma \vdash t_1 : \mathcal{F}A_1 \quad x : A_1 \vdash t_2 : \mathcal{F}A_2}{\Gamma \vdash \mathsf{let}_{\mathcal{F}} x = t_1 \;\mathsf{in}\; t_2 : \mathcal{F}A_2}\right]\!\!\right] = [\![\Gamma \vdash t_1 : \mathcal{F}A_1]\!]; [\![x : A_1 \vdash t_2 : \mathcal{F}A_2]\!]^{*^{\mathcal{F}}}$$

$$\left[\!\!\left[\frac{\Gamma \vdash t : A}{\Gamma \vdash [t]_{\mathcal{F}} : \mathcal{F}A}\right]\!\!\right] = [\![\Gamma \vdash t : A]\!]; {\eta^{\mathcal{F}}}_{[\![A]\!]}$$

In particular, as defined by Moggi and others (Moggi 1989, 1991; Ramsey and Pfeffer 2002; Wadler 1992), the *state transition monad*, the *continuation monad* $\mathbb{C}$, the *interactive input monad*, the *interactive output monad*, the *parser monad*, the *communicating process monad*, the *probability monad*, the *monad of computation with complexity (or preference)*, and other possible monads are useful for treating problems in formal pragmatics. Their combinations are also useful for formal pragmatics. For example, *the monad of underspecification with preference* can be defined by a combination of the state transition monad, the continuation monad, the parser monad, and the monad of computation with preference.

In particular, I use the *finite powerset monad* for word order, presupposition generation and consumption, and information structure, defined as follows:

Definition 17 ($\vdash$ *of* $\lambda_{c\alpha\odot!}$: $\mathscr{P}_\omega$, $\mathbb{C}$, ∂, *and* $\flat$)

$$(I_{\mathscr{P}_\omega}1)\ \frac{\Gamma \vdash t : B}{\Gamma \vdash \{t\} : \mathscr{P}_\omega(B)} \qquad (I_{\mathscr{P}_\omega}2)\ \frac{A \textit{ is a type}}{\Gamma \vdash \emptyset : \mathscr{P}_\omega(A)}$$

$$(E_{\mathscr{P}_\omega})\ \frac{\Gamma \vdash t : \mathscr{P}_\omega(\mathscr{P}_\omega(A))}{\Gamma \vdash \bigcup t : \mathscr{P}_\omega(A)} \qquad (I_{\mathscr{P}_\omega}3)\ \frac{\Gamma \vdash t : \mathscr{P}_\omega(A) \quad \Delta \vdash u : \mathscr{P}_\omega(A)}{\Delta;\Gamma \vdash t \cup u : \mathscr{P}_\omega(A)}$$

$$(Unit_{\mathbb{C}(\alpha,R)})\ \frac{\Gamma \vdash t : B}{\Gamma \vdash \lambda^\alpha x : (B \overset{\alpha}{\multimap} R).(x^\alpha t) : (B \overset{\alpha}{\multimap} R) \overset{\alpha}{\multimap} R}$$

$$(let_{\mathbb{C}(\alpha,R)})\ \frac{\Gamma \vdash t : (A \overset{\alpha}{\multimap} R) \overset{op\ \alpha}{\multimap} R \quad \Gamma, x : A \vdash f : A \overset{\alpha}{\multimap} ((B \overset{op\ \alpha}{\multimap} R) \overset{\alpha}{\multimap} R)}{\Gamma \vdash \lambda^\alpha x : (B \overset{\alpha}{\multimap} R).t^\alpha(\lambda^\alpha y : A.f^\alpha y^\alpha x) : (B \overset{\alpha}{\multimap} R) \overset{\alpha}{\multimap} R}$$

where type $(A \overset{\alpha}{\multimap} R) \overset{op\ \alpha}{\multimap} R$ *is also written by* $\mathbb{C}(A,\alpha,R)$;

$$(\oslash_{\mathbb{C}(\alpha,R)})\ \frac{\Gamma \vdash t : \mathbb{C}(A,\alpha,R) \quad \Delta \vdash u : \mathbb{C}(A,\alpha,R)}{\Gamma;\Delta \vdash t \odot (\mathsf{let}\ (findv_A\ u) = (push_A\ t)\ \mathsf{in}\ u) : \mathbb{C}(A,\alpha,R) \oslash_A \mathbb{C}(A,\alpha,R)}$$

$$(I_\partial 1)\ \frac{\Gamma \vdash a : \mathscr{P}_\omega(A) \quad \Delta \vdash f : A \overset{\alpha}{\multimap} B}{\Delta;\Gamma \vdash \partial^\alpha{}_a(f) : \partial^\alpha(A \overset{\alpha}{\multimap} B)} \qquad (I_\partial 2)\ \frac{\Gamma \vdash b : \mathscr{P}_\omega(A) \quad \Delta \vdash \partial^\alpha{}_a(f) : \partial^\alpha(A \overset{\alpha}{\multimap} B)}{\Gamma;\Delta \vdash \partial^\alpha{}_{a\cup b}(f) : \partial^\alpha(A \overset{\alpha}{\multimap} B)}$$

$$(I_\flat)\ \frac{\Gamma \vdash f : A}{\Gamma \vdash \flat^\alpha{}_a(f) : \flat^\alpha(A)}$$

where

$$\begin{aligned}
[\![\Delta;\Gamma]\!] &= [\![\Delta]\!] \times [\![\Gamma]\!] \\
[\![\partial^\alpha(A \overset{\alpha}{\multimap} B)]\!] &= \{\partial^\alpha{}_a(f) \mid a \in [\![A]\!].\forall b \in a.f^\alpha a \in [\![B]\!]\} \\
[\![A \oslash_C B]\!] &= \{t \overset{\alpha}{\odot} (\mathsf{let}\ (findv_C\ u) = (push_C\ t)\ \mathsf{in}\ u) \mid t \in [\![A]\!], u \in [\![B]\!]\}
\end{aligned}$$

where *findv*, *push*, and *del* are defined by recursion on *PTerm*, as follows:

$$\begin{aligned}
findv_C\ x &= \begin{cases} x & \text{if } x \text{ is a variable of type } C \\ * & \text{otherwise} \end{cases} \\
findv_C\ (t^\alpha u) &= findv_C\ t \otimes findv_C\ u \\
findv_C\ (\lambda^\alpha x.t) &= del(x, findv_C\ t) \\
del(x,x) &= * \\
del(x, y \otimes t) &= \begin{cases} del(x,t) & \text{if } x = y \\ y \otimes del(x,t) & \text{otherwise} \end{cases} \\
push_C\ t &= \begin{cases} \{t\} & \text{if } t \text{ is a term of type } C \\ \emptyset & \text{otherwise} \end{cases} \\
push_C\ (t^\alpha u) &= push_C\ t \cup push_C\ u \\
push_C\ (\lambda^\alpha x.t) &= push_C\ t
\end{aligned}$$

If $a \in \mathscr{P}(A \times \{u\})$, where $u : A \rightarrow \mathbb{Q}$, then we can introduce an ordering over A, and it is useful to treat a kind of preference order among the elements of A by defining for each $a = \langle \mathbf{a}, u \rangle$, $b = \langle \mathbf{b}, u \rangle \in \mathscr{P}(A \times \{u\})$, $a \prec b$ if and only if $u(\mathbf{a}) < u(\mathbf{b})$.

The β-conversions of ${\partial^{\alpha}}_a$ are:

1. $({\partial^{\alpha}}_a(f))^{\alpha} t \rightarrow_{\beta} f^{\alpha} t \quad$ if $t \in a$;
2. ${\partial^{\alpha}}_a({\partial^{\alpha}}_b(t)) \rightarrow_{\beta} {\partial^{\alpha}}_{a \cup b}(t)$;
3. ${\flat^{\alpha}}_a({\partial^{\alpha}}_a(t)) \rightarrow_{\beta} t$

That is, an *ordered definiteness* ${\partial^{\alpha}}_{a,u}$ is defined by a β-conversion, as follows:

$$({\partial^{\alpha}}_{a,u}(f))^{\alpha} t \rightarrow_{\beta} \; f^{\alpha} t \quad \text{if } max_u(a) = t$$

4 Applications of Lambek $\alpha\lambda$-Calculus to Some Formal Pragmatic Problems

4.1 The English?-Focus Combinator and Additive Particle Combinators

Although Steedman (2000) treats the English focus by CCG with θ (the theme marker) and ρ (the rheme marker), the denotation of θ and ρ are not given.

(1) a. Who does John love?
b. John loves [Mary$_F$].
c. Anyone else?
d. John loves Susan, too.

The exhaustiveness of $\lambda x.x$ *loves Mary* is updated by (1d). Therefore, the additive particle such as *too*, *also* must be formalized as an update function of exhaustive functions.

(2) a. John loves [Mary$_F$].
b. ${\partial^{\alpha}}_a((\lambda^{\triangleright\triangleright} x : e.(\text{John}^{\triangleleft}(\text{loves}^{\triangleright} x))))^{\triangleleft}(F_{\triangleleft, \mathsf{e} \overset{\triangleleft\triangleleft}{\multimap} \mathsf{prop}_E} \text{Mary})$

where $F_{\alpha,\alpha'} : \forall x.x \overset{\alpha'}{\multimap} \partial(x \overset{\alpha}{\multimap} \mathsf{prop}_E) \overset{\alpha}{\multimap} \mathsf{prop}_E$ s.t.

$$(F_{\alpha,\alpha'}\ t)^{\alpha} f = \begin{cases} ((\flat_a f)^{\alpha'} f)^{\alpha} t & \text{if } t \in a \\ \bot & \text{otherwise} \end{cases}$$

(3) a. John loves Susan, too.
b. ${\partial^{\alpha}}_a((\lambda^{\triangleright\triangleright} x : e.(\text{John}^{\triangleleft}(\text{loves}^{\triangleright} x))))^{\triangleleft}(\text{Susan}^{\triangleleft}\text{too})$

where $\text{too}_{\alpha,\alpha',\alpha''} : \forall x.x \overset{\alpha'}{\multimap} \partial(x \overset{\alpha}{\multimap} \mathsf{prop}_E) \overset{\alpha}{\multimap} \mathsf{prop}_E$ s.t.

$$f^{\alpha''}(t^{\alpha'}\mathrm{too}_{\alpha,\alpha',\alpha''}) = \begin{cases} \partial_{\{t\}}\partial_a(f)^{\alpha}t & \text{if } a \cup \{t\} \in \mathscr{P}_\omega(A) \text{ for some type } A \\ \bot & \text{otherwise} \end{cases}$$

4.2 A Left-Dislocation Topicalization Combinator

(4) a. Mary, John loves.
b. $(\mathrm{Mary}^{\triangleleft}L_{\triangleright\triangleright})^{\triangleright}(\lambda^{\triangleright\triangleright}x : e.(\mathrm{John}^{\triangleleft}(\mathrm{loves}^{\triangleright}x)))$

where $L_\alpha : \forall x.x \overset{\triangleleft}{\multimap} \partial_a(x \overset{\alpha}{\multimap} \mathsf{prop}_E) \overset{\triangleleft}{\multimap} \mathsf{prop}_E$ s.t.

$$(t^{\triangleleft}L)^{\triangleright} f = \begin{cases} f^{\alpha}t & \text{if } f = \partial_a(g) \text{ and } t \in a \\ \bot & \text{otherwise} \end{cases}$$

4.3 Word Orders in East Asia

Almost all Tibeto-Burman languages have the SOV word order. However, some Tibeto-Burman languages also have the OSV word order by attaching the objective marker as follows: (Naxi: Iwasa (1983))

(5) a. ngə³ t'ɯ¹ la¹.
I him hit
'I hit him.'
b. t'ɯ¹ to¹ ngə³ la¹.
him OBJ me hit
'I hit him.'

In Memba (Sun et al. 1980):

(6) a. pe¹³ ji⁵³ ci pri¹³ wo⁵³ neʔ¹³.
he letter wrote Modal.Particle
'He wrote a letter.'
b. ji⁵³ ci le³¹ pe¹³ pri¹³ wo⁵³ neʔ¹³.
letter Particle he wrote Modal.Particle
'He wrote a letter.'

As well as these Tibeto-Burman examples, Old Japanese also has related cases, a kind of leftward fronting caused by particle-attachments such as *so*, *ka*, etc., called *kakari-musubi* in the traditional descriptive Japanese grammar (cf. Yamada 1912; Ōno et al. 1974), as follows:

(7) a. umasi kuni *so* akidu-sima Yamato no kuni *Φ*a.
wonderful country Particle dragonfly-island Japanese Gen country Top
'What a wonderful country, the Japanese country is!' (Man'yōshūno. 2)
b. *Φ*ito-tu matu iku yo *ka* he-runu.
one-CLASS pine how-many generation Particle pass-Perfect.adnominal

'Solitary pine, how many generations of man have you known?' (Man'yōshū no. 1042)

For example, (5a) and (5b) are formalized, respectively, as follows:

(8)
$$\frac{\text{ngə}^3 : \mathsf{e} \qquad \dfrac{\text{pe}^{13} : \mathsf{e} \quad \text{la}^1 : \mathsf{e} \overset{\lhd}{\multimap} (\mathsf{e} \overset{\lhd}{\multimap} \mathsf{prop}_{Naxi})}{\text{pe}^{31\lhd}\text{la}^1 : \mathsf{e} \overset{\lhd}{\multimap} \mathsf{prop}_{Naxi}}}{\text{ngə}^{3\lhd}(\text{pe}^{31\lhd}\text{la}^1) : \mathsf{prop}_{Naxi}}$$

(9)
$$\frac{\dfrac{\boxed{\Xi}\;\dfrac{\dfrac{\dfrac{\text{ngə}^3 : \mathsf{e} \qquad \dfrac{x : \mathsf{e} \quad \text{la}^1 : \mathsf{e} \overset{\lhd}{\multimap} (\mathsf{e} \overset{\lhd}{\multimap} \mathsf{prop}_{Naxi})}{x^{\lhd}\text{la}^1 : \mathsf{e} \overset{\lhd}{\multimap} \mathsf{prop}_{Naxi}}}{\text{ngə}^{3\lhd}(x^{\lhd}\text{la}^1) : \mathsf{prop}_{Naxi}}}{\lambda^{\rhd\lhd}x : \mathsf{e}.(\text{ngə}^{3\lhd}(x^{\lhd}\text{la}^1)) : \mathsf{e} \overset{\rhd\lhd}{\multimap} \mathsf{prop}_{Naxi}} \qquad a : \mathscr{P}_{\omega}(\mathsf{e})}{\partial^{\rhd\lhd}{}_a(\lambda^{\rhd\lhd}x : \mathsf{e}.(\text{ngə}^{3\lhd}(x^{\lhd}\text{la}^1))) : \partial^{\rhd\lhd}(\mathsf{e} \overset{\rhd\lhd}{\multimap} \mathsf{prop}_{Naxi})}}{} \qquad \boxed{\Xi}\;\dfrac{\text{t}'\text{ɯ}^1 : \mathsf{e} \quad \text{to}^1 : \mathsf{e} \overset{\lhd}{\multimap} \partial^{\rhd\lhd}(\mathsf{e} \overset{\rhd\lhd}{\multimap} \mathsf{prop}_{Naxi}) \overset{\rhd}{\multimap} \mathsf{prop}_{Naxi}}{\text{t}'\text{ɯ}^{1\rhd}\text{to}^1 : \partial^{\rhd\lhd}(\mathsf{e} \overset{\rhd\lhd}{\multimap} \mathsf{prop}_{Naxi}) \overset{\rhd}{\multimap} \mathsf{prop}_{Naxi}}}{(\text{t}'\text{ɯ}^{1\rhd}\text{to}^1)^{\rhd}(\partial^{\rhd\lhd}{}_a(\lambda^{\rhd\lhd}x : \mathsf{e}.(\text{pe}^{13\lhd}(x^{\lhd}\text{la}^1)))) : \mathsf{prop}_{Naxi}}$$

(7a) is formalized as follows:

(10)
$$\frac{\text{akidu-sima Yamato no kuni } \Phi\text{a} : \partial(\mathsf{e} \overset{\lhd}{\multimap} \mathsf{prop}_{OJ}) \qquad \dfrac{\mathit{so} : \mathsf{e} \overset{\lhd}{\multimap} \partial(\mathsf{e} \overset{\lhd}{\multimap} \mathsf{prop}_{OJ}) \overset{\lhd}{\multimap} \mathsf{prop}_{OJ} \qquad \dfrac{\text{umasi} : \mathsf{e} \overset{\rhd}{\multimap} \mathsf{e} \quad \text{kuni} : \mathsf{e}}{\text{umasi}^{\lhd}\text{kuni} : \mathsf{e}}}{(\text{umasi}^{\lhd}\text{kuni})^{\rhd}\mathit{so} : \partial(\mathsf{e} \overset{\lhd}{\multimap} \mathsf{prop}_{OJ}) \overset{\lhd}{\multimap} \mathsf{prop}_{OJ}}}{((\text{umasi}^{\lhd}\text{kuni})^{\rhd}\mathit{so})^{\rhd}(\text{akidu-sima Yamato no kuni } \Phi a) : \mathsf{prop}_{OJ}}$$

4.4 Anaphora in Discourse Continuations

Many researchers (Barker (2002), Shan and Barker (2006), Barker and Shan (2006), de Groote (2008)) have been developing the applications of *continuations* in the sense of the *continuation passing style (CPS)* translations of λ-terms.

In this subsections, I propose a treatment of anaphora using basically continuation (monads) in Definition 17.

(11) John sleeps. He is happy.

This is formalized as follows:

(12) $\underline{(\text{John}^{\lhd}\text{sleeps}) \odot (\mathit{let}\ (\mathit{find}\, v_{\mathsf{e}}\ (x^{\lhd}\text{is happy})) = (\mathit{push}_{\mathsf{e}}\ (\text{John}^{\lhd}\text{sleeps}))\ \mathit{in}\ x^{\lhd}\text{is happy})}$

where the underlining means that the underlined term is lifted by continuation, and (12) is converted to:

(13) $\underline{(\text{John}^{\lhd}\text{sleeps}) \odot (\textit{let } x = \text{John } \textit{in } x^{\lhd} \text{is happy})}$

Furthermore, (13) is converted to:

(14) $\underline{(\text{John}^{\lhd}\text{sleeps}) \odot (\text{John}^{\lhd} \text{is happy})}$

A short discourse involving a bound presupposition within *if-then* clause can be treated as follows:

(15) a. If John has a daughter, she is pretty.
b. If John has a daughter, she is pretty. # She likes sushi.
c. $\underline{(\text{if}^{\rhd}(\text{John}^{\lhd}(\text{has}^{\rhd}(\text{a}^{\rhd}\text{daughter})))^{\rhd}(x^{\lhd}\text{is pretty})}$, where
$\underline{((\text{if}^{\rhd}\varphi)^{\rhd}\psi)} = \underline{(\varphi \Rightarrow (\varphi \odot (\textit{let } (\textit{findv}_A\ \psi) = (\textit{push}_A\ \varphi) \textit{ in } \psi)))}(\lambda e.e)$

In the standard CPS (Sørensen and Urzyczyn 2006), $\lambda e.e$ halts continuations. Therefore, the inaccessibility to terms within *if-then*-clauses of pronouns outside of *if-then*-clauses such as (15b) can be treated.

5 Conclusion

As we have seen, if we want to provide a mathematical device which can treat syntax, semantics, and pragmatics seamlessly, we need exploit not only substructural logics and their the Curry-Howard-isomorphic λ-calculi but also computational λ-calculi.

In the present paper I could not treat *underspecification with preference, interactive processes*, i.e., *dialogue*, and other formal pragmatic issues. But the computational λ-calculi continue to develop many types of computational monads and that fact will make the analysis of these phenomena possible. These are the future directions of the project I report on here.

Appendix A: Categories, Monads, Comonads and Coalgebras

Definition 18 A pair $\boldsymbol{C} = \langle Ob\boldsymbol{C}, Mor\boldsymbol{C}\rangle$ is called a *category* if $Ob\boldsymbol{C}$ is a set of *objects* and $Mor\boldsymbol{C}$ a set of *morphisms* between $Ob\boldsymbol{C}$ which satisfies the following conditions:

- For all $A \in Ob\boldsymbol{C}$, $id_A : A \to A$ s.t. $id_A(a) = a$ is a member of $Mor\boldsymbol{C}$;
- For all $A, B \in Ob\boldsymbol{C}$ and $f : A \to B \in Mor\boldsymbol{C}$, $id_B \circ f = f$ and $f \circ id_A = f$;
- For all $A, B, C, D \in Ob\boldsymbol{C}$ and $f : A \to B$, $g : B \to C$, $h : C \to D \in Mor\boldsymbol{C}$, $h \circ (g \circ f) = (h \circ g) \circ f$;

For each $f : A \to B \in Mor\boldsymbol{C}$, $dom(f) = A$ and $cod(f) = B$.

The set of morphisms from A to B in category $\boldsymbol{C}$ is written by $\boldsymbol{C}(A, B)$

Definition 19 (Mac Lane 1998, *etc*.) Let $\boldsymbol{C}$, $\boldsymbol{D}$ be categories, a *functor* $\mathcal{F}$ from $\boldsymbol{C}$ to $\boldsymbol{D}$, written $\mathcal{F} : \boldsymbol{C} \to \boldsymbol{D}$, is a pair $\langle \mathcal{F}_{ob}, \mathcal{F}_{map} \rangle$, where $\mathcal{F}_{ob} : Ob\boldsymbol{C} \to Ob\boldsymbol{D}$, $\mathcal{F}_{map} : Mor\boldsymbol{C} \to Mor\boldsymbol{D}$ such that:

1. If $f : A \to B$, then $\mathcal{F}_{map}(f) : \mathcal{F}_{ob}(A) \to \mathcal{F}_{ob}(B)$,
2. $\mathcal{F}_{map}(g \circ f) = \mathcal{F}_{map}(g) \circ \mathcal{F}_{map}(f)$, and
3. $\mathcal{F}_{map}(id_A) = id_{\mathcal{F}_{map}(A)}$.

Given two categories $\boldsymbol{C}$ and $\boldsymbol{D}$, we construct a new category $\boldsymbol{C} \times \boldsymbol{D}$, called the *product* of $\boldsymbol{C}$ and $\boldsymbol{D}$, as follows:

- $Ob(\boldsymbol{C} \times \boldsymbol{D}) = \{\langle c, d \rangle \mid c \in Ob\boldsymbol{C}, d \in Ob\boldsymbol{D}\}$
- $Mor(\boldsymbol{C} \times \boldsymbol{D}) = \{\langle f, g \rangle \mid f : C \to C', f' : C' \to C'' \in Mor\boldsymbol{C}, g : D \to D', g' : D' \to D'' \in Mor\boldsymbol{D}, \langle f', g' \rangle \circ \langle f, g \rangle = \langle f' \circ f, g' \circ g \rangle\}$
- Functors *fst*, *snd* s.t. $\boldsymbol{C} \xleftarrow{fst_{\boldsymbol{C},\boldsymbol{D}}} \boldsymbol{C} \times \boldsymbol{D} \xrightarrow{snd_{\boldsymbol{C},\boldsymbol{D}}} \boldsymbol{D}$ are called *projections* of the product, satisfying

$$fst_{\boldsymbol{C},\boldsymbol{D}}\langle f, g \rangle = f$$

and

$$snd_{\boldsymbol{C},\boldsymbol{D}}\langle f, g \rangle = g$$

and if $\boldsymbol{C} \xleftarrow{\mathcal{F}_1} \boldsymbol{E} \xrightarrow{\mathcal{F}_2} \boldsymbol{D}$, then there is a unique functor $\mathcal{F} : \boldsymbol{E} \to \boldsymbol{C} \times \boldsymbol{D}$ with $fst_{\boldsymbol{C},\boldsymbol{D}} \circ \mathcal{F} = \mathcal{F}_1$ and $snd_{\boldsymbol{C},\boldsymbol{D}} \circ \mathcal{F} = \mathcal{F}_2$.

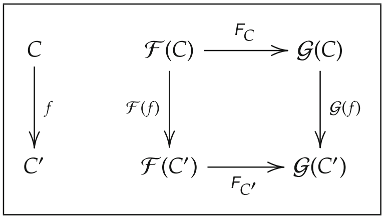
Sometimes I omit $_{ob}$ and $_{map}$ and simply write both of $\mathcal{F}_{ob}$ and $\mathcal{F}_{map}$ as $\mathcal{F}$.

Example 1

1. The *identity functor* on category $\boldsymbol{C}$: $\mathcal{I}d_{\boldsymbol{C}}(X) = X$ and $\mathcal{I}d_{\boldsymbol{C}}(f) = f$ if $X \in Ob\boldsymbol{C}$ and $f \in Mor\boldsymbol{C}$.
2. A *constant functor*: $\tilde{D}(X) = D$ for all $X \in Ob\boldsymbol{C}$ and $\tilde{D}(f) = id_D$ for all $f \in Mor\boldsymbol{C}$.
3. A *covariant powerset functor* with rank κ: $\mathscr{P}_\kappa(A) = \{X \mid X \subseteq A, |X| < \kappa\}$ and $\mathscr{P}_\kappa(f)(X) = \{f(a) \in Y \mid a \in X\}$ if $f : X \to Y$.
4. A *contravariant powerset functor* with rank κ: $\tilde{\mathscr{P}}_\kappa(A) = \mathscr{P}_\kappa(A)$ and $\tilde{\mathscr{P}}_\kappa(f)(X) = \{a \in X \mid f(a) \in Y\}$ if $f : X \to Y$.
5. The *finite probability distribution functor* (Cîrstea 2006; Moss and Viglizzo 2004): $\mathscr{D}_\omega(X) = \{\mu : X \to [0, 1] \mid \sum_{x \in X} \mu(x) = 1, |X| < \omega\}$ and $\mathscr{D}_\omega(f)(\mu)(X) = \sum_{f(a) \in Y} \mu(a)$ for $f : X \to Y$ and $\mu \in \mathscr{D}_\omega(X)$, i.e., $\mathscr{D}_\omega(f)(\mu) = \mu \circ f^{-1}$.
6. ***Set***-Polynomial Functors (Cîrstea 2006; Schröder 2008): $\mathcal{F} ::= \boldsymbol{X} \mid \tilde{C} \mid \mathcal{I}d \mid \mathscr{P}_\omega \mid \mathscr{D}_\omega \mid \mathcal{F}_1 + \mathcal{F}_2 \mid \mathcal{F}_1 \times \mathcal{F}_2 \mid \mathcal{F}_1 \circ \mathcal{F}_2 \mid \mathcal{F}^D \mid \mu\mathbf{X}.\mathcal{F}$, where for each $X \in Ob\boldsymbol{Set}$, $f : X \to Y \in Mor\boldsymbol{Set}$:

(a) functor variable: $\boldsymbol{X}$
(b) the D-depth functor: $|D| < \omega, D \in \boldsymbol{ObSet}, \mathcal{F}^D(X) = \mathcal{F}(X)^D$ and $\mathcal{F}^D(f) : \mathcal{F}^D(X) \to \mathcal{F}^D(Y)$ s.t. $\mathcal{F}^D(f)(g) = \mathcal{F}(f) \circ g$
(c) coproduct functor: $(\mathcal{F}_1 + \mathcal{F}_2)(X) = \mathcal{F}_1(X) + \mathcal{F}_2(X)$, and $(\mathcal{F}_1 + \mathcal{F}_2)(f) = \mathcal{F}_1(f) + \mathcal{F}_2(f) : \mathcal{F}_1(X) + \mathcal{F}_2(X) \to \mathcal{F}_1(Y) + \mathcal{F}_2(Y)$ s.t. $(\mathcal{F}_1 + \mathcal{F}_2)(f)(\iota_i(a)) = \iota_i(\mathcal{F}_i(f)(a))$ for $i \in \{1, 2\}$;
(d) product functor: $(\mathcal{F}_1 \times \mathcal{F}_2)\langle a, b\rangle = \langle \mathcal{F}_1(a), \mathcal{F}_2(b)\rangle$, $(\mathcal{F}_1 \times \mathcal{F}_2)\langle f, g\rangle = \langle \mathcal{F}_1(f), \mathcal{F}_2(g)\rangle$, and $(\mathcal{F}_1 \times \mathcal{F}_2)(f)(a, b) = (\mathcal{F}_1(f)(a), \mathcal{F}_2(f)(b))$;
(e) composite functor: $(\mathcal{F}_1 \circ \mathcal{F}_2)(X) = \mathcal{F}_1(\mathcal{F}_2(X))$ and $(\mathcal{F}_1 \circ \mathcal{F}_2)(f) = \mathcal{F}_1(\mathcal{F}_2(f))$;
(f) fixed-point functor (Moggi 1991): $(\mu\mathbf{X}.\mathcal{F})(X) = \bigcup_{\alpha<\omega} \mathcal{F}^\alpha(X)$ $(\mathcal{F}^0(X) = X$ and $\mathcal{F}^{\alpha+1}(X) = \mathcal{F}(\mathcal{F}^\alpha(X)))$ and $(\mu\mathbf{X}.\mathcal{F})(f)(X) = f[\mu\mathbf{X}.\mathcal{F}(X)]$

Definition 20 (Mac Lane 1998) Given two functors $\mathcal{F}, \mathcal{G} : \boldsymbol{C} \to \boldsymbol{D}$, a *natural transformation* $\mathsf{F}\colon \mathcal{F} \to \mathcal{G}$ is a function which assigns to each $C \in \mathit{Ob}\boldsymbol{C}$ an arrow $\mathsf{F}_C : \mathcal{F}(C) \to \mathcal{G}(C) \in \mathit{Mor}\boldsymbol{D}$, called a *component* of F, in such a way that every $f : C \to C' \in \mathit{Mor}\boldsymbol{C}$ yields a diagram:



which is commutative.

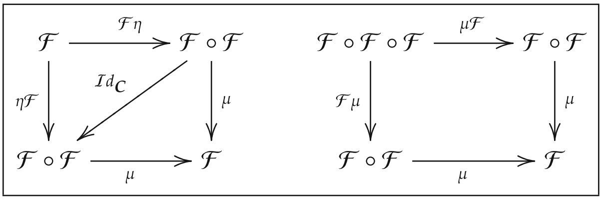
When this holds, F_C is called *natural* in C.

The natural transformation F is called a *natural equivalence* or *natural isomorphism*, written by $\mathsf{F}\colon \mathcal{F} \cong \mathcal{G}$, if every component F_C is an isomorphism.

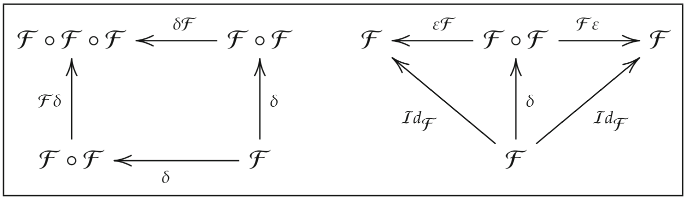
Definition 21 (Adámek et al. 1990) Let $\mathcal{F}, \mathcal{G} : \boldsymbol{A} \to \boldsymbol{B}$ be functors and $\tau : \mathcal{F} \to \mathcal{G}$ a natural transformation. Then:

- for each functor $\mathcal{H} : \boldsymbol{C} \to \boldsymbol{A}$, the natural transformation $\tau\mathcal{H} : \mathcal{F} \circ \mathcal{H} \to \mathcal{G} \circ \mathcal{H}$ is defined by
$$(\tau\mathcal{H})_C = \tau_{\mathcal{H}(C)}$$
- for each functor $\mathcal{J} : \boldsymbol{B} \to \boldsymbol{D}$, the natural transformation $\mathcal{J}\tau : \mathcal{J} \circ \mathcal{F} \to \mathcal{J} \circ \mathcal{G}$ is defined by
$$(\mathcal{J}\tau)_A = \mathcal{J}(\tau_A)$$

Definition 22 (Blute and Scott 2004) A *monad* in a category $\boldsymbol{C}$ is a triple $\langle \mathcal{F}, \eta, \mu\rangle$, where $\mathcal{F} : \boldsymbol{C} \to \boldsymbol{C}$ is an endofunctor on $\boldsymbol{C}$, $\eta : \mathcal{I}d_{\boldsymbol{C}} \to \mathcal{F}$ $(unit)$ and $\mu : \mathcal{F} \circ \mathcal{F} \to \mathcal{F}$ $(multiplication)$ are natural transformations, requiring that the following diagrams commute:



Definition 23 (Asperti and Longo 1991) A *comonad* in a category $\boldsymbol{C}$ a triple $\langle \mathcal{F}, \varepsilon, \delta \rangle$, where $\mathcal{F} : \boldsymbol{C} \to \boldsymbol{C}$ is an endofunctor on $\boldsymbol{C}$, $\varepsilon : \mathcal{F} \to \mathcal{I}d_{\boldsymbol{C}}$ (*counit*) and $\mu : \mathcal{F} \to \mathcal{F} \circ \mathcal{F}$ (*comultiplication*) are natural transformations, requiring that the following diagrams commute:



Definition 24 (Adámek 2005) Let $\mathcal{F}$ be an endofunctor on category $\boldsymbol{C}$ and $A \in Ob\boldsymbol{C}$. A pair $\mathcal{A} = \langle A, \alpha \rangle$ consisting of an object A and a morphism $\alpha : A \to \mathcal{F}(A)$, is called an $\mathcal{F}$-*coalgebra*. We call α the dynamics of the coalgebra $\mathcal{A}$.

Definition 25 (Blute and Scott 2004) A *Cartesian closed category (CCC)* $\langle \boldsymbol{C}, \times, I, (-)^{(-)}, ev, curry, fst, snd \rangle$ is a category $\boldsymbol{C}$ with a terminal object I, binary product $\times$ and exponentiation $(-)^{(-)}$, where $I \in Ob\boldsymbol{C}$ s.t for every $A \in Ob\boldsymbol{C}$, there is exactly one arrow from A to I, written by $A \xrightarrow{!_C} I$, $B^A = \{f \mid f : A \to B\}$,

$$\boldsymbol{C}(C \times A, B) \cong \boldsymbol{C}(C, B^A),$$

satisfying the following properties:

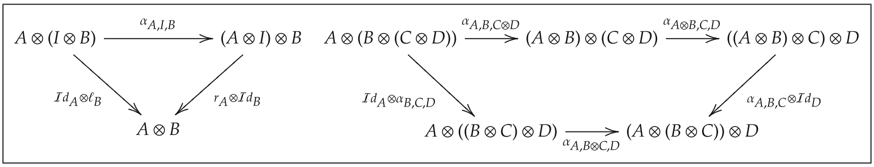
1. $ev_{A,B} : B^A \times A \to B$
2. $curry f : C \to B^A \quad \text{if } f : C \times A \to B$
3. $(Beta)\ \ ev\langle curry f, snd \rangle = f : C \times A \to B$
4. $(Eta)\ \ curry(ev(g \circ fst, snd)) = g : C \to B^A$

Definition 26 (Blute and Scott 2004) A *monoidal category* $\langle \boldsymbol{C}, I, \otimes, \alpha, \ell, r \rangle$ is a category $\boldsymbol{C}$ with:

- unit object $I \in Ob\boldsymbol{C}$,
- bifunctor $\otimes : \boldsymbol{C} \times \boldsymbol{C} \to \boldsymbol{C}$,
- isomorphic functor α s.t. $\alpha_{A,B,C} : (A \otimes B) \otimes C \overset{\cong}{\to} A \otimes (B \otimes C)$,
- isomorphic bifunctor ℓ s.t. $\ell_A : I \otimes A \overset{\cong}{\to} A$, and

- isomorphic bifunctor r s.t. $r_A : A \otimes I \xrightarrow{\cong} A$,

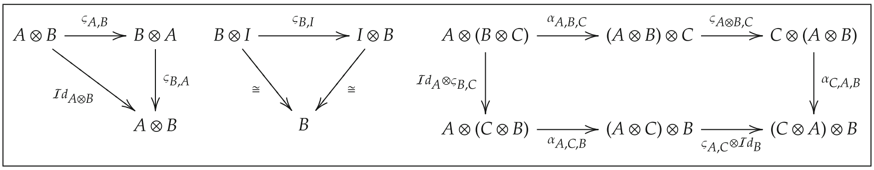
requiring that the following diagrams commute:



Definition 27 (Blute and Scott 2004) Let $\langle \boldsymbol{C}, I, \otimes, \alpha, \ell, r \rangle$ be a monoidal category. Then $\langle \boldsymbol{C}, I, \otimes, \alpha, \ell, r, \varsigma \rangle$ is a *symmetric monoidal category*, where ς is a natural isomorphism s.t.

$$\varsigma_{A,B} : A \otimes B \to B \otimes A,$$

requiring that the following three diagrams commute:



Definition 28 (Blute and Scott 2004) Let $\boldsymbol{C}$ be a (symmetric) monoidal category. Then $\langle \boldsymbol{C}, \multimap, ev, coev \rangle$ is a *(symmetric) monoidal closed category*, satisfying

$$\boldsymbol{C}(C \otimes A, B) \cong \boldsymbol{C}(C, A \multimap B),$$

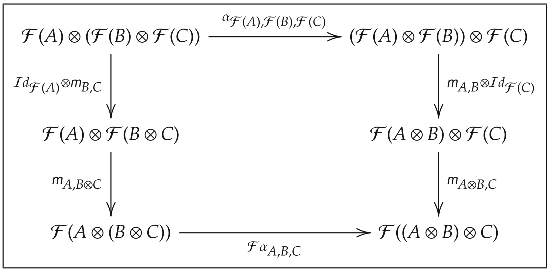
for all $A, B, C \in Ob\boldsymbol{C}$.

ev and *coev* are functors s.t.

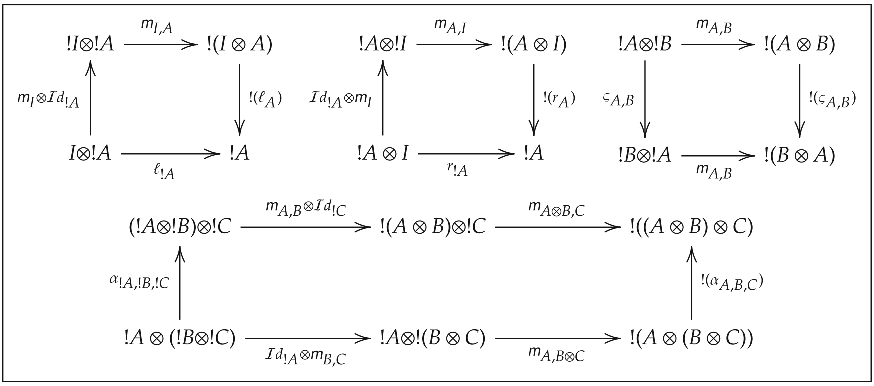
$$ev_{A\multimap B,A} : (A \multimap B) \otimes A \to B,$$
$$coev_{B,A} : B \to (A \multimap (B \otimes A)),$$

satisfying the adjoint equations.

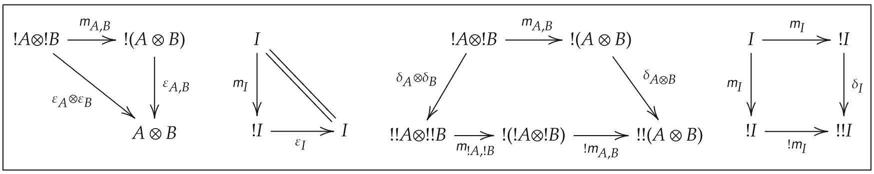
Definition 29 (Blute and Scott 2004; Mac Lane 1998) A *monoidal functor* between monoidal categories is a triple $\langle \mathcal{F}, m_{-,-}, m_I \rangle$ where $\mathcal{F} : \boldsymbol{C} \to \boldsymbol{D}$ is a functor, together with two natural transformations $m_I : I \to \mathcal{F}(I)$ and $m_{U,V} : \mathcal{F}(U) \otimes \mathcal{F}(V) \to \mathcal{F}(U \otimes V)$, requiring that the following diagram commutes in category $\boldsymbol{D}$:



Definition 30 (Bierman 1995) A triple $\langle !, m_{A,B}, m_I\rangle$ is a *symmetric monoidal functor* if and only if the following diagrams commute:



Let $\langle !, \varepsilon, \delta\rangle$ be a comonad, where ε and δ are *monoidal natural transformations*, requiring that the following diagrams commute:

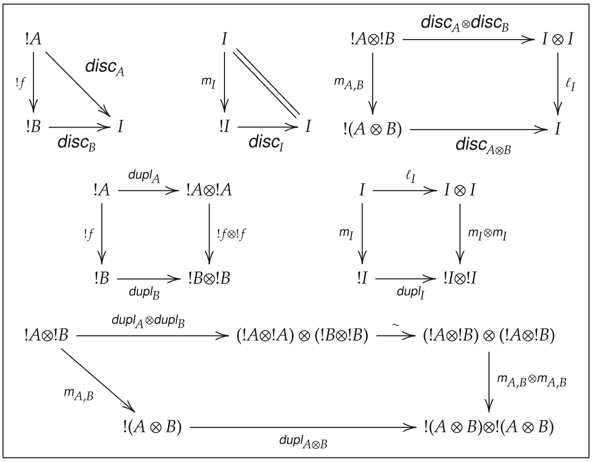


Then $\langle !, \varepsilon, \delta, m_{A,B}, m_I\rangle$ is a *symmetric monoidal comonad*.

Furthermore, if for every free !-coalgebra $\langle !A, \delta_A\rangle$ there are two monoidal natural transformations *disc* and *dupl* with components

$$disc_A : !A \longrightarrow I$$
$$dupl_A : !A \longrightarrow !A \otimes !A$$

which form a *commutative comonoid* and are coalgebra morphisms, i.e., requiring that the following diagrams commute, for any morphism $f : A \to B$:



And, if $\langle !A, dupl_A, disc_A \rangle$ satisfies all conditions of a commutative comonoid (see Bierman 1995), $dupl_A$ and $disc_A$ satisfy all conditions of coalgebra morphisms (see Bierman 1995), and whenever $f : \langle !A, \delta_A \rangle \to \langle !B, \delta_B \rangle$ is a coalgebra morphism between free coalgebras, then it is also a comonoid morphism, then a symmetric monoidal closed category $\boldsymbol{C}$ with a symmetric monoidal comonad $\langle !, \varepsilon, \delta, m_{A,B}, m_I \rangle$ is called a *linear category*, which is a categorical model of MELL.

Appendix B: Categorical Semantics of Substructural λ-Calculi

I only enumerate the categorical semantics of (Ax), (I_{Unit}), (Con), (P), (L), (R), (C), (W),[1] and $(Pro(mote))$ as follows:

Definition 31

- (Ax) $\left[\!\left[\overline{x : A \vdash x : A} \right]\!\right] = \mathcal{I}d_{[\![A]\!]}$
- (I_{Unit}) $\left[\!\left[\overline{\vdash * : I} \right]\!\right] = * \in I$
- (Con) $\left[\!\left[\overline{c : A \vdash c : A} \right]\!\right] = c \in [\![A]\!]$
- (P) $\left[\!\left[\dfrac{x : A, y : C \vdash t : B}{y : C, x : A \vdash t : B} \right]\!\right] = [\![x : A, y : C \vdash t : B]\!] \circ \varsigma_{A,C}$
- (L) $\left[\!\left[\dfrac{x : A, (y : B, z : C) \vdash t : D}{(x : A, y : B), z : C \vdash t : D} \right]\!\right] = [\![x : A, (y : B, z : C) \vdash t : D]\!] \circ \alpha_{A,B,C}$
- (R) $\left[\!\left[\dfrac{(x : A, y : B), z : C \vdash t : D}{x : A, (y : B, z : C) \vdash t : D} \right]\!\right] = [\![(x : A, y : B), z : C \vdash t : D]\!] \circ \alpha_{A,B,C}$

[1] This version of (C) and (W) are the Gentzen sequent's ones.

- (C) $\left[\!\!\left[\dfrac{\Gamma, (x :!A, y :!A) \vdash u : B}{\Gamma, t :!A \vdash (copy\ t\ as\ x, y\ in\ u) : B}\right]\!\!\right] = [\![\Gamma, (x :!A, y :!A) \vdash u : B]\!] \circ (\mathcal{I}d_{[\![\Gamma]\!]} \otimes dupl_A)$
- (W) $\left[\!\!\left[\dfrac{\Gamma \vdash t : B}{\Gamma, x :!A \vdash t : B}\right]\!\!\right] = [\![\Gamma \vdash t : B]\!] \circ (\mathcal{I}d_{[\![\Gamma]\!]} \otimes disc_A)$
- (Pro) $\left[\!\!\left[\dfrac{\{\Gamma_i \vdash t_i :!A_i\}_{i<m+1} \quad (y_0 : A_0, \ldots, y_m : A_m) \vdash u : B}{(x_0 : A_0, \ldots, x_m : A_m) \vdash (promote\ t_0, \ldots, t_m\ for\ y_0, \ldots, y_m\ in\ u) :!B}\right]\!\!\right]$ $\left(\bigotimes_{i<m+1} [\![\Gamma_i \vdash t_i :!A_i]\!]\right) ; \delta \otimes \cdots \otimes \delta; ! \left([\![(y_0 : A_0, \ldots, y_m : A_m) \vdash u : B]\!]\right)$

References

Abrusci, V. M. (1991). Phase semantics and sequent calculus for pure noncommutative classical linear propositional logic. *The Journal of Symbolic Logic*, *56*(4), 1403–1451.

Abrusci, V. M. (2002). Classical conservative extensions of Lambek calculus. *Studia Logica*, *71*(3), 277–314.

Abrusci, V. M. (2003). Towards a semantics of proofs for non-commutative logic: Multiplicatives and additives. *Theoretical Computer Science*, *294*(3), 335–351.

Adámek, J., Herrlich, H., & Strecker, G. E. (1990). *Abstract and concrete categories—The joy of cats*. New York: John Wiley and Son Inc.

Adámek, J. (2005). Introduction to coalgebra. *Theory and Applications of Categories*, *14*(8), 157–199.

Asperti, A., & Longo, G. (1991). *Categories, types, and structures: An introduction to category theory for the working computer scientist*. Cambridge: The MIT Press.

Barker, C. (2002). Continuations and the nature of quantification. *Natural Language Semantics*, *10*(3), 211–242.

Barker, C., & Shan, C.-C. (2006). Types as graphs: Continuations in type logical grammar. *Journal of Logic, Language and Information*, *15*(4), 331–370.

Benton, N. (1995). A mixed linear and non-linear logic: Proofs, terms and models (extended abstract). In Selected papers from the 8th International Workshop on Computer Science Logic (Vol. 933, pp. 121–135). Lecture Notes in Computer Science.

Benton, N., & Wadler, P. (1996). Linear logic, monads and the lambda calculus. In Proceedings of the 11th Annual IEEE Symposium on Logic in Computer, Science (pp. 420–431).

Bierman, G. M. (1995). What is a categorical model of intuitionistic linear logic? In Proceedings of the Second International Conference on Typed Lambda Calculi and Applications (Vol. 902, pp. 78–93). Lecture Notes in Computer Science, April 1995.

Blute, R., & Scott, P. (2004). Category theory for linear logicians. In Linear logic in computer science (Vol. 316, pp. 3–64). London Mathematical Society Lecture Note Series. Cambridge: Cambridge University Press.

Blute, R. F., Lamarche, F., & Ruet, P. (2002). Entropic Hopf algebras and models of non-commutative logic. *Theory and Applications of Categories*, *10*(17), 424–460.

Buszkowski, W. (1997). Mathematical linguistics and proof theory. In J. van Benthem & A. ter Meulen (Eds.), Handbook of logic and language (Chapter 12, pp. 683–736). Amsterdam: Elsevier Science B.V.

Buszkowski, W. (1987). The logic of types. In J. T. Srzednicki (Ed.), *Initiatives in logic* (Vol. 2, pp. 180–206). Dordrecht: Nijhoff.

Cîrstea, C. (2006). Modularity in coalgebra. *Electric Notes in Theoretical Computer Science*, *164*(1), 3–26.

de Groote, P. (1999). The non-associative Lambek calculus with product in polynomial time. In Proceedings of Automated Reasoning with Analytic Tableaux and Related Methods (pp. 128–139).

de Groote, P. (2008). *A type-theoretic view of dynamic logic*. Tokyo: Presented in Fifth Workshop on Lambda Calculus and Formal Grammar.

de Groote, P., & Lamarche, F. (2002). Classical non-associative Lambek calculus. *Studia Logica*, *71*(3), 355–388.

Gabbay, D. M. (1996). *Labelled deductive systems* (Vol. 1). Oxford: Clarendon Press.

Girard, J.-Y. (1987). Linear logic. *Theoretical Computer Science*, *50*, 1–102.

Goubault-Larrecq, J., Lasota, S., & Nowak, D. (2008). Logical relations for monadic types. *Mathematical Structures in Computer Science*, *18*, 1169–1217.

Iwasa, M. (1983). *Chugoku no Shosu Minzoku to Gengo (The Minority Ethnic Groups and their Languages in China)*. Tokyo: Koseikan.

Jacobs, B. (1994). Semantics of weakening and contraction. *Annals of Pure and Applied Logic*, *69*(1), 73–106.

Kanazawa, M. (1999). Lambek calculus: Recognizing power and complexity. In J. Gerbrandy, M. Marx, M. de Rijke & Y. Venema (Eds.), JFAK. Essays dedicated to Johan van Benthem on the occasion of his 50th Birthday. Amsterdam: Amsterdam University Press, Vossiuspers.

Kanazawa, M. (1992). The Lambek calculus enriched with additional connectives. *Journal of Logic, Language and Information*, *1*(2), 141–171.

Lambek, J. (1958). The mathematics of sentence structure. *The American Mathematical Monthly*, *65*, 154–170.

Lambek, J. (1968). Deductive systems and categories I. Syntactic calculus and residuated categories. *Journal of Mathematical Systems Theory*, *2*(4), 287–318.

Lambek, J. (2004). Bicategories in algebra and linguistics. In T. Ehrhard, J.-Y. Girard, P. Ruet, & P. Scott (Eds.), *Linear logic in computer science* (pp. 325–345). Cambridge: Cambridge University Press.

Mac Lane, S. (1998). *Categories for the working mathematician*. Berlin: Springer.

Moggi, E. (1989). *An abstract view of programming languages*. Technical report: Stanford University.

Moggi, E. (1991). Notions of computation and monads. *Information and Computation*, *93*(1), 55–92.

Moortgat, M. (1997). Categorial type logics. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (pp. 93–177). Amsterdam: Elsevier Science B.V.

Morrill, G. V. (1994). *Type logical grammar: Categorial logic of signs*. New York: Springer.

Moss, L. S., & Viglizzo, I. D. (2004). Harsanyi type spaces and final coalgebras constructed from satisfied theories. *Electronic Notes in Theoretical Computer Science*, *106*, 279–295.

Ōno, S., et al. (1974). *Iwanami Kogo Jiten (The Iwanami Dictionary of Old Japanese)*. Tokyo: Iwanami Shoten.

Polakow, J., & Pfenning, F. (1999). Natural deduction for intuitionistic non-commutative linear logic. In Proceedings of the 4th International Conference on Typed Lambda Calculi and Applications (pp. 295–309).

Ramsey, N., & Pfeffer, A. (2002). Stochastic lambda calculus and monads of probability distributions. In Proceedings of the 29th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (pp. 154–165).

Restall, G. (2000). *An introduction to substructural logics*. London: Routledge.

Schröder, L. (2008). Expressivity of coalgebraic modal logic: The limits and beyond. *Theoretical Computer Science*, *390*(2–3), 230–247.

Shan, C.-C., & Barker, C. (2006). Explaining crossover and superiority as left-to-right evaluation. *Linguistics and Philosophy*, *29*(1), 91–134.

Sørensen, M. H., & Urzyczyn, P. (2006). *Lectures on the Curry-Howard isomorphism*. Amsterdam, Netherlands: Elsevier Science Inc.

Steedman, M. (2000). *The syntactic process*. Cambridge: The MIT Press.

Sun, H., Lu, S., Zhang, J., & Jueya, O. (Eds.). (1980). *Menba, Luoba, Dengren de yuyan (The languages of Memba, Lakpa, and Den)*. Beijing: Zhonguo Shehui Kexue Chuanshe.

Wadler, P. (1992). Comprehending monads. *Mathematical Structures in Computer Science*, *2*(4), 461–493.

Walker, D. (2005). Substructural type systems. In B. C. Pierce (Ed.), *Advanced topics in types and programming languages* (pp. 3–43). Cambridge: The MIT Press.

Wansing, H. (1993). *The logic of information structures*. Berlin: Springer.

Wansing, H. (2002). A rule-extension of the non-associative Lambek calculus. *Studia Logica*, *71*(3), 443–451.

Yamada, Y. (1912). Narachō Bunpō-shi (The History of the Grammar of the Japanese in the Nara Era). Tokyo: Hōbunkan.

Yetter, D. N. (1990). Quantales and (noncommutative) linear logic. *The Journal of Symbolic Logic*, *55*(1), 41–64.

On the Functions of the Japanese Discourse Particle *yo* in Declaratives

David Y. Oshima

Abstract This chapter presents a novel analysis of two central uses—Davis' (2011) "guide to action" and "correction" uses—of the Japanese discourse particle *yo* occurring in declarative clauses. *Yo* accompanied by the question-rise contour has a function to add the propositional content to the modal base for priority modality relativized to the hearer, thereby indicating that the propositional content is relevant to what the hearer should and may do. *Yo* accompanied by the non-rising (flat) contour has a function to indicate that the hearer should have recognized the propositional content beforehand. Four other functions of *yo* in declaratives will also be briefly discussed. It will further be pointed out *yo* accompanied by the rise-fall contour has similar functions as *yo* accompanied by the non-rising contour, but additionally expresses the speaker's want for the hearer's sympathy and/or understanding.

Keywords Japanese · Discourse particle · Intonation · Priority modality · Conventional implicature · Dynamic semantics

1 Introduction

This chapter develops an analysis of some major functions of the Japanese discourse particle *yo*. Section 2 presents basic facts about *yo*. Section 3 briefly reviews three influential analyses of *yo*: Takubo and Kinsui (1997), McCready (2009), and Davis (2011), and discusses their limitations. Sections 4 and 5 present a novel analysis of two central uses of *yo*, which Davis (2011) calls the "guide to action" and "correc-

An earlier version of this work was presented at Logic and Engineering of Natural Language Semantics 8 in 2011. I would like to thank the audience for valuable feedback. I would also like to thank Elin McCready, Chris Davis, and the anonymous reviewers for helpful comments.

D. Y. Oshima (✉)
Department of International Communication, Nagoya University, Nagoya, Japan
e-mail: davidyo@nagoya-u.jp

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*, Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_12,

tion" uses. It will be argued (i) that *yo* accompanied by the question-rise contour has a function to add the propositional content (of the prejacent, i.e., the sentence without *yo*) to the set of propositions serving as the modal base for priority (deontic) modality relativized to the hearer, and (ii) that *yo* accompanied by the non-rising (flat) contour has a function to indicate that the hearer should have recognized the propositional content beforehand. Section 6 discusses some other functions of *yo* in declaratives. Section 7 demonstrates that *yo* accompanied by the rise-fall contour has similar functions as *yo* accompanied by the non-rising contour, but additionally expresses the speaker's want for the hearer's sympathy and/or understanding.

2 Basic Facts About *yo*

The functions of the discourse particles (also called the sentence-final particles) in Japanese, and in particular of *yo*, have attracted a great deal of attention in the literature.

Yo is one of the most frequently occurring discourse particles, and is used in a wide variety of speech styles and registers, e.g., both in male and female speech, and both in formal and informal speech. Also, it may occur in a wide range of clause types including declaratives, interrogatives, imperatives, and hortatives.

It has been recognized that *yo* exhibits rather different functions depending on the intonation accompanying it (Koyama 1997; Davis 2011).[1] *Yo* may occur with (i) the rising contour commonly referred to as the "question rise" and assigned the label "LH%" in Venditti's (2005) notational system, (ii) the non-rising contour (the "flat" contour in Kori 1997; the "falling" contour in Davis 2011) indicated by the absence of intonation label in Venditti's system, or (iii) the "rise-fall" contour assigned the label "HL%" in Venditti's system.[2] In the following, I will use ↗ to indicate the rising contour (question rise), ↘ to indicate the non-rising contour, and ↑↓ to indicate the rise-fall contour (a similar notational system is adopted in Kori 1997). Below are some examples to illustrate the usage of *yo*.[3]

---

[1] It is not immediately clear if an intonational contour is directly associated with a discourse particle like *yo*, or rather the contour is primarily an attribute of a larger utterance unit that may contain a discourse particle at its end. This issue does not have a direct bearing on the discussion in the current work.

[2] *Yo* is not compatible with the "insisting rise" contour (Kori's "↑"; Venditti's "H%"), with which some other discourse particles including *ne* are compatible (Oshima 2013).

[3] The abbreviations used in glosses are: Acc = accusative, Attr = attributive, Ben = benefactive, Cond = conditional, Cop = copula, Dat = dative, DP = discourse particle, Gen = genitive, Imp = imperative, Ipfv = imperfective, Neg = Negation, Nom = nominative, Prs = present, Pst = past, Q = question marker, Top = topic, Vol = volitional.

(1) a. A, kasa-o wasureteiru-**yo**↗
oh umbrella-Acc forget.Ipfv.Prs-*yo*
'Oh, you forgot your umbrella.'
b. Karada-ni ki-o tsukero-**yo**↗
body-Dat mind-Acc put.Imp-*yo*
'Take care of yourself.'
c. Dooiu imi-da-**yo**↘
what.kind.of meaning-Cop.Prs-*yo*
'What does that mean?'

This work primarily addresses what Davis (2011) calls the "guide to action" use and "correction" use of *yo*, but also discusses some other uses of *yo* in declaratives. *Yo* in clause types other than declaratives, as well as *yo* occurring in combination with other discourse particles such as *ne* and *wa* or the auxiliary *n*(*o*)*da* (e.g., *wasureteiru-yo-ne*, *wasureteiru-nda-yo*), will be excluded from the discussion, although it is clear that consideration of them is essential for a full understanding of *yo*.

3 Previous Discussions of *yo*

3.1 Takubo and Kinsui (1997)

Takubo and Kinsui (1997) claim, in brief, that *yo* is an inference-trigger. By uttering (2), for example, the speaker invites the hearer to make an inference such as "The hearer should take an umbrella with him" or "The picnic will be canceled"; note that the label for the question-rise contour was added by the present author, assuming that it is the intonation intended by Takubo and Kinsui.

(2) Ame-ga futteiru-**yo**↗
rain-Nom fall.Ipfv.Prs-*yo*
'It is raining.' (adapted from Takubo and Kinsui 1997:756)

"Direction to make an inference", however, is not a sufficiently specific characterization of the function of *yo* in question. Compare (3) and (4), assuming that (i) A and B are members of the same student reading club, (ii) A is in charge of buying supplies such as stationery and utensils, and (iii) A is now at a supermarket on an errand, with B accompanying him to give a hand.

(3) A: Kami koppu-mo katteokoo-ka-na.
paper cup-also buy.beforehand.Vol-Q-DP
'Perhaps I should buy some paper cups too.'
B: Kami koppu-wa mada takusan nokotteru-{**yo**↗/#∅}
paper cup-Top still many remain.Ipfv.Prs-*yo*/∅
'We've still got plenty of paper cups.'
(**Implicature**: You don't need to buy paper cups now.)

(4) B: Kami koppu-wa kawanai-no?
paper cup-Top buy.Neg.Prs-DP
'Are you not going to buy paper cups?'
A: Kami koppu-wa mada takusan nokotteru-{??**yo**↗/∅}
paper cup-Top still many remain.Ipfv.Prs-*yo*/∅
'We've still got plenty of paper cups.'
(**Implicature**: I don't need to buy paper cups now.)

(3B) and (4A) invite similar inferences and convey similar conversational implicatures, and yet the use of *yo* is compulsory in the former while it is not so, and sounds even unnatural with the rising contour, in the latter.

To give another example, (5) is more natural with *yo* accompanied by the rising contour (*yo*↗ in short) if it is uttered by B (the passenger), but is more natural without it if it is uttered by A (the driver).

(5) (**Situation**: A is driving and B is on the passenger seat. They are 100 km away from their destination.)
A, gasorin-ga moo nai-{**yo**↗/∅}
oh gasoline-Nom already absent.Prs-*yo*/∅
'Oh, we are running out of gas.'

Takubo and Kinsui's analysis does not account for the described contrasts. *Yo* with the rising contour specifically has to do with (inference regarding) what the *hearer* should do or be (see Davis 2011, p. 97 for a similar remark).

3.2 McCready (2009)

McCready (2009) suggests that *yo* is essentially a marker of *importance* (or relevance). Specifically, he argues that *yo* indicates that the informativity value—usefulness of a statement in providing an answer to the question at issue in the discourse—of the propositional content for the hearer (H) is above some contextual threshold, and also that the speaker (S) *insists* that H accepts the propositional content, even if it is not consistent with H's previous beliefs. The importance and insistence indicated by *yo* are formulated as follows:

(6) $[\![yo(\phi)]\!]=$
a. Presupposition: $\mathcal{B}_S IV_S(Q,\phi) > d_s$
(i.e.: The speaker believes that the informativity value of ϕ for the hearer with respect to the contextually specified question Q is higher than the contextually specified relevance threshold d_s.)
b. Semantics: $\sigma||sassert(\phi)||\sigma'$

where $sassert$ stands for strong assertion, i.e., the operation to update the information state with a certain proposition whether or not it is compatible with the pre-update information state; when the proposition is incompatible with the pre-update information state, *downdate* (removal of content from the information state) takes place first so that inconsistency is avoided. Formally:

(7) $\sigma||sassert(\phi)||\sigma' =$
$\sigma||\phi||\sigma'$ if $\sigma||\phi|| \neq \emptyset$
$\sigma|| \downarrow \neg\phi; \phi||\sigma'$ else.

Like Takubo and Kinsui (1997), McCready does not consider the contribution of intonation to the function of the discourse particles, although he admits that it is a crucial component for a full account (p. 467).

McCready's analysis, as it is, does not seem to account for the speaker/hearer asymmetry illustrated above. In (3) and (4), for example, the "question at issue" is presumably: "Is it necessary for A to buy paper cups?". In both scenarios, the second utterance is definitely useful in providing an answer to it.

Also, under his analysis, it is hard to explain why the use of *yo* is often felt to be superfluous in a direct answer to an explicitly asked question, as in (8), while it tends to be compulsory in a context where the speaker gives a suggestion or warning in an indirect manner, as in (9) and (10) (see Takubo and Kinsui 1997, p. 756; Inoue 1997, pp. 65–66; Davis 2011, pp. 99–100 for relevant remarks).

(8) (**Situation**: A is looking at a handwritten math formula.)
A: Kore-wa nana, soretomo ichi?
this-Top 7 or 1
'Do you have a "7" here, or is it a "1"?'
B: Nana-desu-{#**yo**↗/∅}
7-Cop.Prs.Polite-*yo*
'It's a "7".'

(9) (**Situation:** A and B are at a noodle restaurant. It is the first time for A to eat there.)

A: Soba-ni shiyoo-ka-na, soretomo udon-ni shiyoo-ka-na.
soba-Dat do.Vol-Q-DP or udon-Dat do.Vol-Q-DP
'I wonder if I should have soba (buckwheat noodles) or udon (wheat noodles).'
B: Koko-no soba-wa oishii-desu-{**yo**↗/??∅}
here-Gen soba-Top good.Prs-Polite-*yo*/∅
'The soba here is good.'
B': Koko-no soba-wa amari oishikunai-desu-{**yo**↗/??∅}
here-Gen soba-Top particularly good.Neg.Prs-Polite-*yo*/∅
'The soba here is not particularly good.'

(10) (**Situation:** A and B are at a supermarket. B takes a package of English tea from the shelf. A knows that B prefers green tea and suspects that B meant to take green tea.)

A: Sore, koocha-desu-{**yo**↗/??∅}
that English.tea-Cop.Prs.Polite-*yo*/∅
'That's English tea.'

It is counterintuitive to suppose that (8B) is less informative than (9B, B')/(10A) in their respective context.

One may suspect that McCready's analysis is suitable for *yo*↘, though not for *yo*↗. I will show in Sect. 3.3, however, that it is not adequate for *yo*↘ either.

3.3 Davis (2011)

Davis (2011) recognizes two main uses of *yo* in declaratives, which are respectively accompanied by the rising and non-rising intonation. He characterizes the function of *yo*↗, illustrated in (11) (see also (3), (5), (9) and (10)), as "guide to action", and that of *yo*↘, illustrated in (12), as "(call for) correction".

(11) A: Eiga-o miru mae-ni gohan-o tabeyoo-ka?
movie-Acc watch.Prs before meal-Acc eat.Vol-Q
'Shall we eat before watching the movie?'
B: Moo shichi-ji-sugi-deshoo? Eiga-wa
already 7-o'clock-past-Cop.Presumptive.Polite movie-Top
hachi-ji-kara-da-**yo**↗
8-o'clock-from-Cop.Prs-*yo*
'It's already past 7, right? The movie starts at 8.' (Davis 2011:19)

(12) A: Eiga-wa ku-ji-kara-da-kara gohan-o taberu
movie-Top 9-o'clock-from-Cop.Prs-because meal-Acc eat.Prs
jikan-wa juubun-ni aru-ne.
time-Top sufficiently exist.Prs-DP
'Since the movie starts at 9, there's plenty of time to eat.'
B: Chigau-**yo**↘ Eiga-wa hachi-ji-kara-da-**yo**↘
wrong.Prs-*yo* movie-Top 8-o'clock-from-Cop.Prs-*yo*
'That's wrong. The movie starts at 8.' (Davis 2011:19)

Davis develops an analysis of *yo* where the semantic contribution of the particle itself and that of the accompanying intonation are distinguished. In line with Gunlogson (2003), Davis departs from the standard Stalnakerian assumption that declaratives update the common ground (the intersection of the interlocutors' belief sets) and hypothesizes that declaratives usually have the speaker's public beliefs (those beliefs that both the speaker and the hearer acknowledge that the speaker has) as the target of update. He then argues that *yo* itself instructs to update not only the speaker's public beliefs but the hearer's public beliefs too (or more generally, *all* discourse participants' public beliefs).

The empirical consequences of this claim are not clear. Davis states that due to this contrast only a declarative with *yo* (either with the rising or non-rising contour) but not a bare declarative can be felicitously used when the hearer has to give up one or more of his previous beliefs before accepting its propositional content (pp. 112, 117). As will be shown below (with data in (15) and (16)), however, a bare declarative can naturally—and under certain circumstances, more naturally than a declarative with *yo*—be used to make a "corrective" statement. In the rest of this section, I put aside this component of Davis' account of *yo*, and focus on the others having to do with (what he calls) the "intonational morphemes" combined with *yo*.

3.3.1 The "guide to action" use

Regarding *yo*↗, Davis essentially argues that it (i) introduces a *decision problem* for the hearer (or equivalently a set of *alternative actions* from which the hearer has to choose) to the discourse, or makes reference to an existing one, and (ii) indicates that there is some alternative action *a* such that *a* cannot be determined to be optimal according to the hearer's beliefs before the update (i.e., before the propositional content is added to the hearer's beliefs), but can be determined to be optimal after the update. In the case of (9B), for example, the suggested optimal action would be to eat soba.

Davis' analysis of *yo*↗ is too restrictive in excluding its use in scenarios like (13), where the propositional content may or may not affect what the optimal action for the hearer is, and (14), where the contextual decision problem remains unsolved in the post-update context.

(13) (**Situation:** A and B are eating together. B is going to have a Buffalo wing. A knows that it is very spicy, but does not know if B likes spicy food or not.)

A: Sore, karai-**yo**↗
that spicy.Prs-*yo*
'That's spicy.'

(14) (**Situation:** A and B are at a mobile phone shop. B is considering buying a model released a while ago.)

A: Raigetsu-ni nattara atarashii moderu-ga
next.month-Dat become.Cond new.Prs model-Nom
deru-**yo**↗ Matsu kachi-ga aru-kadooka-wa
come.out.Prs-*yo* wait.Prs value-Nom exist.Prs-whether-Top
wakaranai-kedo.
know.Neg.Prs-though
'A new model will be released next month. I don't know if it is worth waiting for, though.'

In the scenario of (13), the relevant action set is presumably: {eating the Buffalo wing, not eating the Buffalo wing}. The premise that B was going to eat the Buffalo wing implies that in the pre-update context it was optimal for him to eat it. A's utterance, thus, is to be understood to make the other action (not eating the Buffalo wing) optimal. This, however, is not the intention of A here; what he means to convey is something like: "You should not eat it if you don't like spicy food" or "You should consider the fact that it is spicy before deciding whether you eat it or not". Likewise, in (14), it would be too strong to say that A tries to convince B to wait until the next month and buy the yet-to-be-released product. Rather, A merely presents a piece of information that he thinks might or might not affect B's choice.

One may argue that in cases like (13) and (14), the decision problem is whether to consider the propositional content, and the suggested optimal action is to consider it. However, if the concepts of the decision problem and the optimal action have to be interpreted in such an extended way, then it seems more reasonable to dispense with them entirely from the formulation, and suppose more simply that [ϕ *yo*↗] indicates that the speaker believes that the hearer is better off considering ϕ than not. In Sect. 4 I will present an analysis along this idea.

3.3.2 The "correction" use

Regarding *yo* accompanied by the non-rising contour, developing McCready's (2009) idea, Davis claims that it explicitly indicates that the utterance requires a non-monotonic update, i.e., an update requiring elimination of previously accepted information, on the hearer's beliefs (see also Inoue 1997, p. 63; Koyama 1997, pp. 105–106; Izuhara 2003, pp. 5–6). In the case of (12), the information to be eliminated is that the movie starts at 9, which contradicts the propositional content that the movie starts at 8.

It can be shown, however, that non-monotonicity (backed up by the speaker's willingness to explicitly correct the hearer) is not a sufficient condition for occurrence of *yo*↘. Observe the following examples:

(15) (**Situation:** Araki runs a bookstore, and Morino runs a computer store next to it. They are close friends, and often stop by each other's place during business hours for small talks. Araki comes in the computer store and asks the employee called Nomoto, assuming that Morino is there.)

A: Konchiwa. Morino-san, ima isogashii-ka-na.
hello Morino-Suffix now busy.Prs-Q-DP
'Hello. Is Morino busy now?'

a. (Morino does not work on Sundays. Araki knows it, but has forgotten that today is Sunday.)
N: Kyoo-wa nichiyoo-da-kara
today-Top sunday-Cop.Prs-because
oyasumi-desu-{**yo**↘/∅}
day.off-Cop.Prs.Polite-*yo*/∅
'He's not here because it is Sunday.'

b. (It is Monday and Morino is supposed to be there.)
N: Kyoo-wa kaze-de oyasumi-desu-{#**yo**↘/∅}
today-Top cold-by day.off-Cop.Prs.Polite-*yo*/∅
'He is taking a day off because he has a cold.'

(16) (**Situation:** Yoshio and Kazuki are friends. Yoshio is a year older than Kazuki. At Kazuki's apartment, Yoshio recalls that he had to make a phone call, but realizes that he didn't have his mobile phone with him. Yoshio sees a mobile phone on the table, and assumes that it is Kazuki's and is in a working condition.)

Y: Kore chotto tsukatte-mo ii-ka-na.
this a.little use-if good.Prs-Q-DP
'Can I use this for a while?'

a. (The phone actually is a kid's toy.)
K: A, sore, omocha-desu-{**yo**↘/∅}
oh that toy-Cop.Prs.Polite-*yo*/∅
'Oh, that's a toy.'

b. (The phone is Yoshio's.)
K: A, sore, Yoshio-san-ga kinoo wasurete-itta
oh that Y.-Suffix-Nom yesterday forget-go.Pst
yatsu-desu-{**yo**↘/∅}
one-Cop.Prs.Polite-*yo*/∅
'Oh, that's yours, Yoshio. You left it here yesterday.'

c. (The phone is Kazuki's, but it is out of battery.)
K: A, sore, batterii-ga kiretemasu-{#**yo**↘/∅}
oh that battery-Nom run.out.Ipfv.Prs.Polite-*yo*/∅
'Oh, it's out of battery.'

d. (The phone belongs to Yoshio's girlfriend.)
K: A, sore, kanojo-ga kinoo wasurete-itta
oh that girlfriend-Nom yesterday forget-go.Pst
yatsu-desu-{#**yo**↘/∅}
one-Cop.Prs.Polite-*yo*/∅

The use of *yo*↘ is fine in (15a) and (16a, b), but sounds odd (unfairly accusing, unreasonably hostile) in (15b) and (16c, d). The difference here is that in the former set of discourses the speaker is pointing out a misconception that the hearer *could have avoided* utilizing his previous knowledge, reasoning ability, and/or powers of observation, while in the latter the speaker is pointing out a misconception that the hearer could not reasonably be expected to avoid.

One may argue that (15b) and (16c, d) sound strange because they are too abrupt or rude. It is, however, natural to assume that pointing out an avoidable misconception incurs a more serious risk of threatening the hearer's face (in Brown and Levinson's 1987 sense) than pointing out an unavoidable misconception. Indeed, the situations in (15a) and (16a, b) intuitively appear to be more embarrassing for the hearer than those of (15b) and (16c, d). Thus, one would expect that a higher level of politeness is called for in (15a) and (16a, b) than in (15b) and (16c, d), rather than the other way round.

Note that McCready's (2009) analysis discussed above fails to account for the described contrast too. There is no intuitive reason to believe, for example, that the propositional content of (15a) (the proposition that Morino is taking a day off today as he does on other Sundays) is more informative than that of (15b) (the proposition that Morino is taking a day off because he has a cold).

3.4 A Note on Inseparability of Particle Meaning and Intonation Meaning

From the next section onward, I will propose alternative analyses of what Davis calls the "guide to action" and "correction" uses of *yo* and further discuss four additional uses of it. Before proceeding, I would like to make clear my position on how intonation interacts with the interpretation of *yo* (and discourse particles in general). Contra Davis (2011), I consider it impossible to neatly separate the meaning of a discourse particle itself from the meaning of an intonational contour.

It is tempting to hypothesize that the meaning of *yo*↗ is a composition of those of *yo* itself and the question-rise contour, and the meaning of *yo*↘ is a composition of those of *yo* itself and the non-rising (flat) contour. It seems not feasible, however, to fully implement this idea to deal with the various uses of *yo*↗, *yo*↘, and *yo*↑↓ to be discussed below, let alone the contrasting functions of (i) bare (particle-less) clauses with different contours and (ii) clauses with many other particle/contour combinations (Japanese has quite a few discourse particles besides *yo*, which are compatible with different sets of contour types; see Oshima (2013) for an overview). Maintaining the compositional view of particle/intonation combination would require making stretched and ad hoc moves, including assigning to each contour multiple and specific meanings that manifest themselves only in combination with certain discourse particles.

I will thus assume that the combination of a clause type and a contour is the basic unit that carries a conventionalized function, where clause types include bare clauses, *yo*-marked clauses, *ne*-marked clauses, etc. This is of course not to say that there cannot be any semantic commonality or resemblance between clauses that share the same clause type but differ in intonation, or between clauses that share the same contour but belong to different clause types. My view is that there is a great deal of randomness and arbitrariness, as well as a good deal of systematicity, in the discourse functions of clause type/contour combinations.

4 *Yo* with the Rising Intonation: Required and Permitted Actions

Yo in its "guide to action" use indicates that the utterance conveys information that is relevant to and might affect what the hearer should do or be. This information, however, does not need to determine, or imply that it is determined, what it is.[4] To

[4]A similar characterization of *yo*↗ is presented by Inoue (1997, p. 64), who suggests that [ϕ *yo*↗] indicates that ϕ holds true in the circumstances surrounding the speaker and hearer, and further

capture this property of *yo*↗, I propose that it instructs to add the propositional content to the modal base for priority modality relativized to the hearer.

Priority modality is a term covering deontic modality (in the narrow sense, concerning rules, laws, morality, and the like), bouletic modality (concerning desires), and teleological modality (concerning goals), and is synonymous to deontic modality in the broad sense (Portner 2007). Following Kratzer (1991) *inter alia*, I assume that modal expressions in natural language are interpreted with respect to two contextually provided conversational grounds (sets of propositions): the modal base and the ordering source. For priority modality, it is generally understood that the modal base is circumstantial, i.e., consists of *relevant facts*, and the ordering source is *what the laws, rules, moral codes, etc., provide*. Note that the modal base for priority modality generally cannot be identified with the set of all known facts (i.e., the common ground). To illustrate why, take the modal statement "John should be in New York now", which can be true when in actuality John is in San Francisco. If the modal base contains the proposition that John is in San Francisco, then the proposition that John is in New York holds in none of the worlds best-ranked according to the ordering source, so that it is wrongly predicted that the modal statement has to be false.

Priority modality, in general terms, has to do with what should and may hold true in view of certain rules, desires, goals, etc. I introduce the term (agent-)relativized priority modality to refer to a variety of priority modality that has to do with what a particular agent should and may *make* true (roughly, required and permitted actions for the agent; cf. Portner 2007, pp. 370–373). The proposition that there is peace in the nation of X is likely to be a deontic necessity, but not a deontic necessity relativized to an average citizen of X (or of any other nation). It could be, on the other hand, a deontic necessity relativized to the head of state of X; that is, it could be a duty for him to keep peace in or bring peace to X. The set of relevant facts differs for what should be the case in a given context and for what a certain agent should make the case in the same context. To exemplify, suppose that John witnessed a robbery. Whether John should make it the case that the robber is arrested (e.g., by arresting him) depends on factors such as whether John is a police officer, whether he is properly armed, and whether he is running after another criminal. The truth of the (non-relative) deontic statement that the robber should be arrested, on the other hand, is not contingent on such factors.

Let us suppose that bare declaratives (declaratives without *yo*) canonically have a discourse function (context change potential) to add their propositional content to the common ground (Heim 1983), and further that the context consists of the common ground (CG), the modal base (f), and the ordering source (g):

(17) *The discourse function of a bare declarative*
Where C is a context of the form $< CG, f, g >$,
$C + \phi_{decl} = \; < CG', f, g >$, where $CG' = CG \cup \{[\![\phi_{decl}]\!]\}$.

(Footnote 4 continued)
poses to the hearer the question: "What are you going to do in these circumstances?"; see also Izuhara (2003, p. 5).

Table 1 Means to update conversational grounds

	Modal base	Ordering source
Priority modality	**Declaratives with *yo*↗**	Imperatives
Epistemic modality	Regular declaratives	Evidentials

The discourse function of a declarative with *yo* in its "guide to action" use differs from that of a bare declarative in two respects: (i) it requires that the common ground and the modal base be ones appropriate for hearer-relativized priority modality (in particular, the former is required to be circumstantial), and (ii) it adds the propositional content to the modal base, as well as to the common ground.

(18) *The discourse function of a declarative with* yo↗
Where C is a context of the form $< CG, f, g >$,
(i) $C + [\phi_{decl}\,\text{yo}\nearrow]$ is defined only if f and g are concerned with priority modality relativized to the hearer;
(ii) If defined, $C + [\phi_{decl}\,\text{yo}\nearrow] = < CG', f', g >$, where $CG' = CG \cup \{[\![\phi_{decl}]\!]\}$ and $f' = f \cup \{[\![\phi_{decl}]\!]\}$.

In typical cases, a declarative with *yo*↗ has a double function: it informs the hearer of the propositional content, and further points out that it is relevant to what the hearer should and may do. Uyeno's (1972, pp. 72–73) remark that *yo* serves to draw the hearer's attention to the propositional content, and miyazaki et al.'s (2002, p. 266) remark that an utterance with *yo* presents the propositional content as something the hearer should be aware of, appear to point to the same idea.

A declarative with *yo*↗ may also be uttered in a context where its propositional content is already in the common ground (e.g., *Kimi-wa mada miseinen-da-yo*↗ 'You are still under age.'; Kinsui 1993; Takubo and Kinsui 1997). In such a case, it still carries out the second function, and thus, unlike the corresponding bare declarative, is not necessarily redundant.

A proposition added to the priority modal base affects what should and may be (made) the case, either by itself or in conjunction with other propositions; otherwise, it would be irrelevant and cannot be felicitously added to the modal base. Expansion of the modal base, however, does not guarantee that a contextual decision problem, if there is one, is solved in the post-update context. In (14), for example, the speaker will not know the answer to the contextual decision problem: "Should the hearer buy a phone now?" until further information is added to the common ground, such as how the yet-to-be-released model of phone differs from the one currently available.

Note that it is not a new idea that some types of utterances explicitly update conversational backgrounds. Portner (2007) argues that imperatives update the ordering source for priority modality, and suggests that evidentials update the one for epistemic modality. The modal base for epistemic modality is standardly considered to be the same as the common ground (i.e., the set of all known facts), so regular declaratives suffice to update it. Declaratives with *yo*↗ fit in the remaining quadrant (Table 1).

5 *Yo* with the Non-rising Intonation: Blame on Ignorance

It was observed above, with the data in (15) and (16), that an utterance with *yo*↘ is infelicitous in a context where the hearer cannot be reasonably expected to know the propositional content beforehand,[5] and also that corrective statements need not to be accompanied by *yo* (with the rising or non-rising intonation).

I propose that the function of *yo*↘ is essentially to blame the hearer for his failure to recognize the propositional content. Mccready's (2005) analysis, mentioned but not adopted in McCready (2009), pursues this idea (see also Nihongo Kijutsu Bunpoo Kenkyuukai 2003, p. 204); $must_d$ in (19a) is a deontic (priority) necessity operator.

(19) *McCready's (2005) analysis*
$[\![yo(\phi)]\!]=$

a. Presupposition: $\mathcal{B}_S\neg\mathcal{B}_H\phi; \mathcal{B}_S must_d \mathcal{B}_H\phi$
(i.e.: The speaker believes that the hearer does not believe ϕ and the speaker believes that the hearer must come to believe ϕ.)

b. Semantics: $\sigma||sassert(\phi)||\sigma'$
(i.e.: Update the information state with ϕ; in case of incompatibility, first downdate the information state and then update.)

It seems to me that the "presupposition" here can be simplified to "$\neg\mathcal{B}_H(\phi)$; $must_d$ $\mathcal{B}_H\phi$" without changing its effect.

The 2005 version of McCready's analysis fares better with the data in (15) and (16) than the 2009 version. The utterances (15a) and (16a, b) can, if the speaker dares, be naturally followed by a remark like: "Silly you! You should have realized that", while the same does not hold for (15b) or (16c, d). It is counterintuitive, however, to suppose that the utterer of [ϕ *yo*↘] *presupposes* (i.e., takes it for granted that both interlocutors believe) that (the speaker believes that) the hearer should come to believe ϕ at the time of utterance. In the context of (12), for example, obviously the speaker does not expect the hearer to believe that (the speaker believes that) he (= the hearer) should come to believe that the movie starts at 8.

The semantic contribution of *yo*↘, on the other hand, is not part of regular assertion, either. This can be shown by observing that the message conveyed by *yo*↘ cannot be a target of negation. (20B), for example, can only be taken as an attempt to refute the factual claim that the movie starts at 7, and not the message that B should have known that the movie starts at 7.

[5] This property of *yo*↘ is addressed by Hasunuma (1996), who proposes that *yo*↘ directs the discourse participants to fill gaps or fix flaws in their understanding *using their existing knowledge and/or commonsensical reasoning*. My analysis (to be presented below) departs from hers in claiming that the information update ("filling gaps and fixing flaws") is carried out by the utterance itself (rather than the hearer's inference/reasoning) and that *yo*↘ merely conveys that the update *could have been done* with the hearer's previous knowledge, commonsensical reasoning, etc.

(20) A: Eiga-wa shichi-ji-kara-da-**yo**↘
movie-Top 7-o'clock-from-Cop-*yo*
'The movie starts at 7.'
B: Iya, sore-wa chigau.
no that-Top wrong.Prs
'No, that's not so.'

I propose that the semantic contribution of *yo*↘ belongs to the level of conventional implicature (Potts 2005, 2007; McCready 2010). Declaratives with *yo*↘, like bare declaratives and declaratives with *yo*↗, instruct to update the common ground with the propositional content. In addition, they conventionally implicate that the hearer should have realized the propositional content beforehand. Conveying such a message can be sensible only when the hearer had a chance to know the propositional content (but failed to take advantage of it). In the cases of (15a) and (16c, d), the hearer did not have such a chance, and thus it is odd to use *yo*↘.

It is worth noting that the proposed functions of *yo*↗ and *yo*↘ are both concerned with the hearer's duties (the "guide to action" use indicates that the statement is relevant to what the hearer should do, and the "blame on ignorance" use indicates that the hearer failed to do something that he should have done). This commonality can be taken as a conceptual link between the two distinct uses of *yo*.

6 Some Other Uses of *yo*

It is possible to find occurrences of *yo*↗ and *yo*↘ that do not conform to the analysis provided above. *Yo*↘, in particular, has quite a wide range of meanings. To obtain a full understanding of *yo*, it is essential to acknowledge its multi-functionality. It is worth stressing on this point, because in the existing literature on *yo*, it is often implicitly assumed that *yo* is mono-functional or has only a small number of functions. While a uniform analysis is to be preferred provided it can consistently account for the full range of data, close examination of facts reveals that one needs to acknowledge that *yo* (especially *yo*↘) is heavily polysemous.

Below, I describe and discuss some additional uses of *yo*. The task to develop formal analyses of them is beyond the scope of the current work and will be left to future research; it is worth noting, however, that some if not all of the uses discussed in this section seem to be amenable to the theory of conventional implicature/expressive meaning in line with Potts (2005, 2007).

6.1 The "Affection" Use of yo *with the Rising Contour*

Some utterances with *yo*↗ cannot be straightforwardly taken to provide information relevant to what the hearer should and may do (a similar remark is made by Miyazaki et al. 2002, pp. 266–267). Examples (21)–(23) illustrate such cases.

(21) (**Situation**: A comes into his office on the sixth floor. He sees his colleague B, says hello, and then reports what he saw on the way.)

A: Nanka, ikkai-ni keisatsu-ga kiteta(-**yo**↗)
somehow ground.floor-Dat police-Nom come.Ipfv.Pst-*yo*
'I don't know why, but there were some cops on the street floor.'

(22) (**Situation**: A is talking on the phone with his friend B, who moved to Osaka two months ago.)

A: Osaka-no kurashi-wa doo?
Osaka-Gen life-Top how
'How do you like the life in Osaka?'

B: Un, kekkoo tanoshii(-**yo**↗)
yeah fairly fun.Prs-*yo*
'Yeah, it's pretty fun.'

(23) (**Situation**: A is B's mother. B is leaving home.)

A: Obentoo motta?
box.lunch take.Pst
'Do you have your box lunch with you?'

B: Motta(-**yo**↗) Ja, ittekimasu.
take.Pst-*yo* then bye
'I've got it. See you later.'

I suggest that this kind of *yo* merely serves as a marker of affection, and indicates that the speaker is enjoying having verbal interaction with the hearer. I further hypothesize that this second use—the "affection" use, to name it—was derived from the "guide to action" use. *Yo* in its guide to action use is typically used to suggest the hearer to take a certain action, and this action can be "having (further) conversation with the speaker". Expressing a wish to have verbal interaction with the hearer is a natural and common way to express affection to him. It seems natural to consider that this effect of *yo* became conventionalized and gave rise to the "affection" use.[6]

Indeed, one may argue that the occurrences of *yo* in cases like (21) and (22) serve the "guide to action" function, urging the hearer to (continue to) have verbal interaction with the speaker (note that this account cannot be applied to the case of (23), where speaker B closes up the conversation right after his using *yo*). The facts that they are omissible in the given contexts, and that *yo* in its guide to action generally cannot be left out (see (9) and (10)), suggest however that they are serving the derived interpersonal function as an affection marker.

6.2 *Varied Functions of* yo *with the Non-Falling Contour*

Besides the "blame on ignorance" use, *yo*↘ has at least three distinct uses.

[6] I hasten to note, however, that this putative process is a mere stipulation, which is yet to be empirically tested based on diachronic data.

6.2.1 Emotion toward the propositional content

Yo can be used to convey that the speaker feels a heightened emotion toward the propositional content (Tanaka and Kubozono 1999, pp.122–123). The emotion involved can be either positive, as in (24), or negative, as in (25).

(24) Kimi-ga tetsudatte-kureta okage-de hayaku owatta-**yo**↘
you-Nom help-Ben.Pst thanks.to early finish.Pst-*yo*
'I was able to finish the work early thanks to your help.'

(25) (**Situation**: The hearer knows that the speaker lost his wallet at the ballpark two weeks ago.)
Saifu, yappari mitsukarakatta-**yo**↘
wallet as.expected be.found.Neg.Pst-*yo*
'I couldn't find my wallet, as I thought.'

In a case like (26), the emotion involved could be a mere surprise that does not involve positive or negative evaluation.

(26) (**Situation**: The speaker looks out of the window and sees it snowing.)
Wa, yuki-ga futteru-**yo**↘
wow snow-Nom fall.Ipfv.Prs-*yo*
'Wow, it's snowing.'

6.2.2 Exclamatory expression of the speaker's mental state/impression

Yo can also be used to add an exclamatory tone to an utterance where the speaker reports his own emotion, feeling, or impression.

(27) a. Arigatoo, ureshii-**yo**↘
thank.you happy.Prs-*yo*
'Thank you, oh am I happy!'
b. Ano toki-wa ureshikatta-**yo**↘
that time-Top happy.Pst-*yo*
'Was I happy then!'

(28) a. Aitsu-ni-wa hontoo-ni hara-ga tatsu-**yo**↘
that.person-Dat-Top really stomach-Nom stand.Prs-*yo*
'I am really angry at him!'
b. Ano toki-wa hara-ga tatta-**yo**↘
that time-Top stomach-Nom stand.Pst-*yo*
'Was I angry then!'

(29) a. Ano mizuumi-wa hontoo-ni kirei-datta-**yo**↘
that lake-Top really beautiful-Cop.Pst-*yo*
'Was that lake beautiful!'
b. Anna kirei-na mizuumi-wa hoka-ni nai-**yo**↘
that.much beautiful-Cop.Attr lake-Top else absent.Prs-*yo*
'No other lake is as beautiful as that one.'
c. Anna kirei-na mizuumi-wa hajimete mita-**yo**↘
that.much beautiful-Cop.Attr lake-Top for.the.first.time see.Pst-*yo*
'I had never seen such a beautiful lake.'

Note that this use is to be distinguished from the "emotion toward the propositional content" use mentioned above, in that the emphasized emotion is not one toward the propositional content. In (27), for example, *yo* does not convey that the speaker has a strong emotion toward the fact that he feels/felt happy.[7]

6.2.3 Intention/plan

Yo↘ is used in utterances where the speaker explains his intention or plan.

(30) (**Situation**: The speaker and the hearer are at a restaurant.)
Rinji shuunyuu-ga atta-kara kyoo-wa boku-ga ogoru-**yo**↘
one.time income-Nom exist.Pst-because today-Top I-Nom treat.Prs-*yo*
'I will buy your dinner, because I had a little windfall.'

[7]It is possible, as in (i), to use *yo* in its "emotion toward the propositional content" use in an utterance that describes the speaker's emotion, thereby conveying that the speaker has a strong emotion (second-order emotion, so to speak) towards the fact that he feels that emotion.

(i) (**Situation**: The speaker has just heard that an annoying neighbor of his is leaving town.)
Are, ore gakkari-shiteru-**yo**↘ (Nande-daroo.)
oh I disappointed-do.Ipfv.Prs-*yo* why-Cop.Presumptive
'Oh, I am feeling disappointed. (I wonder why.)'

(31) (**Situation**: After work, the speaker is invited by his colleagues to join them to go to a bar.)

Kyoo-wa tsukareteru-kara moo kaeru-**yo**↘
today-Top tired.Ipfv.Prs-because already return.Prs-*yo*
'I am tired today, so I would rather go straight home.'

(32) (**Situation**: The speaker lost a game of tennis to the hearer.)

Kuyashii-naa. Tsugi-wa zettai-ni kachimasu-**yo**↘
vexing.Prs-DP next-Top definitely win.Prs.Polite-*yo*
'How vexing! I will win next time, no matter what.'

This kind of *yo*↘ can be understood as a marker of a commissive illocutionary act.[8] It is worth noting that this use, like the "guide to action" and "correction" uses, has to do with the notion of duty (although it is concerned with the speaker's duties rather than the hearer's).[9]

7 *Yo* with the Rise-Fall Contour

The functions of *yo* accompanied by the rise-fall contour (*yo*↑↓) can be understood as variants of those of *yo*↘. With the rise-fall contour instead of the non-rising (flat) contour, the utterance carries an added childish tone and conveys the speaker's want for the hearer's sympathy and/or understanding.[10]

(33b), (34b), (35b), and (36) respectively illustrate variants of the "blame on ignorance" use, the "emotion toward the propositional content" use, the "exclamatory

[8]Commissive acts are those illocutionary acts that commit the speaker to some future course of action (Searle 1979); promises and offers are paradigmatic examples.

[9]Thanks to the anonymous reviewer for directing my attention to this point.

[10]Note that the rise-fall contour does not always convey a childish tone. To illustrate, (ib), where the discourse particle *ne* is accompanied by the rise-fall contour, is fully felicitous in a conversation between two adults who are socially distant.

(i) Kyoo-wa kaze-ga tsuyoi-desu-{a. ne↑/b. ne(e)↑↓}
today-Top wind-Nom strong.Prs-Polite-DP
'It is windy today, isn't it.'

In comparison to the version with the insisting-rise contour (Kori's "↑"; Venditti's "H%") presented as (ia), (ib) conveys additional emotiveness, but it does not sound childish or indicate the speaker's want for sympathy/understanding (Oshima 2013).

expression of the speaker's mental state/impression" use, and the "intention/plan" use.[11]

(33) (**Situation**: in reply to: "You should go out and get some exercise")
Soto-wa samui-{a. **yo**↘/b. **yo**(**o**)↑↓}
outside-Top cold.Prs-*yo*
'It's cold outside.'

(34) Saifu-ga mitsukaranai-{a. **yo**↘/b. **yo**(**o**)↑↓}
wallet-Nom be.found.Neg.Prs-*yo*
'I can't find my wallet.'

(35) (**Situation**: The speaker receives a phone call from his boss.)
Mata yobidashi-da. Ki-ga omoi-{a. **yo**↘/b. **yo**(**o**)↑↓}
again summon-Cop.Prs mind-Nom heavy.Prs-*yo*
'He wants me to come again. I feel dismal.'

(36) (**Situation**: A child says to his mother, who is holding a balloon.)
Boku-ga motsu-**yo**(**o**)↑↓
I-Nom hold.Prs-*yo*
'I'll hold it.'

The use of *yo*↑↓ is not appropriate in contexts where it is clear that the speaker is not asking for sympathy; this point is illustrated in (37) (cf. (24)).

(37) #Kimi-ga tetsudatte-kureta okage-de hayaku owatta-**yo**(**o**)↑↓
you-Nom help-Ben.Pst thanks.to early finish.Pst-*yo*
(I was able to finish the work early thanks to your help.)

8 Summary

This chapter presented an analysis of two central functions *yo* occurring in declarative clauses. *Yo* with the rising contour has a function to add the propositional content to the modal base of priority modality relativized to the hearer, thereby indicating that it is relevant to what the hearer should and may do. *Yo* with the non-rising contour has a function to indicate that the hearer should have recognized the propositional content beforehand.

It was also pointed out (i) that *yo* with the rising contour has a distinct use as an affection marker, (ii) that *yo* with the non-rising contour has at least three distinct uses (the "emotion toward the propositional content" use, the "exclamatory expression of the speaker's mental state/impression" use, and the "intention/plan" use), and (iii) that *yo* with the rise-fall contour has similar functions as *yo* with the non-rising contour but conveys an additional emotional tone.

[11]The rise-fall contour often involves lengthening of the final vowel (Tanaka and Kubozono 1999, pp. 119–120). In informal writing, this lengthening is often reflected by an added vowel letter or long vowel mark (*choo'onpu*).

The functions of *yo* are diverse (note that *yo* also occurs in non-declaratives, carrying out yet other functions; e.g., Shirakawa 1993; Davis 2011) and this work hardly addressed the links between them. The task to examine the conceptual and diachronic relations between the different uses of *yo* will be left to future research.

References

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Reissue ed.). Cambridge: Cambridge University Press.

Davis, C. (2011). Constraining interpretation: Sentence final particles in Japanese. Ph.D. thesis, University of Massachusetts.

Gunlogson, C. (2003). *True to form: Rising and falling declaratives as questions in English*. New York: Routledge.

Hasunuma, A. (1996). Shuujoshi yo no danwa kinoo [The discourse functions of the sentence-final particle yo]. In I. Ueda, Y. Sunakawa, K. Takami, H. Noda, & A. Hasunuma (Eds.), *Gengo tankyuu no ryooiki [Domains of linguistic inquiry]* (pp. 383–395). Tokyo: Daigaku Shorin.

Heim, I. (1983). On the projection problem for presuppositions. In M. Barlow, D. Flickinger, & M. T. Wescoat (Eds.), *Proceedings of WCCFL 2* (pp. 114–125). Stanford: Stanford University.

Inoue, M. (1997). "Moshimoshi, kippu o otosaremashitayo": Shuujoshi "yo" o tsukau koto no imi ["Excuse me, you dropped your train ticket": The meaning of the use of the sentence-final particle "yo"]. *Gekkan Gengo*, *26*, 62–71.

Izuhara, E. (2003). Shuujoshi "yo" "yone" "ne" saikou [A study of ending particles "yo" "yone" "ne"]. *The Journal of Aichi Gakuin University: Humanities and Sciences*, *51*, 1–15.

Kinsui, S. (1993). Shuujoshi yo to ne [Sentence-final particles yo and ne]. *Gekkan Gengo*, *22*(4), 118–121.

Kori, S. (1997). Nihongo no intoneeshon: Kata to kinoo [Intonation in Japanese: Patterns and functions]. In T. Kunihiro, H. Hirose & M. Kono (Eds.), *Akusento, intoneeshon, rizumu to poozu [Accent, intonation, rhythm and pause]* (pp. 169–202). Tokyo: Sanseido.

Koyama, T. (1997). Bunmatsushi to bunmatsu intoneeshon [Sentence final particles and sentence final intonation]. In Onsei Bunpoo Kenkyuukai (ed.), *Bunpoo to onsei [Speech and grammar]* (vol. 1, pp. 97–119). Tokyo: Kurosio Publishers.

Kratzer, A. (1991). Modality. In A. von Stechow & D. Wunderlich (Eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung* (pp. 651–656). Berlin: Walter de Gruyter.

McCready, E. (2005). The dynamics of particles. Ph.D. thesis, University of Texas.

McCready, E. (2009). Particles: Dynamics vs. utility. In Y. Takubo, T. Kinuhata, S. Grzelak, & K. Nagai (Eds.), *Japanese/Korean linguistics* (Vol. 16, pp. 466–481). Stanford: CSLI Publications.

McCready, E. (2010). Varieties of conventional implicature. *Semantics and Pragmatics*, *3*(8), 1–57.

Miyazaki, K., Adachi, T., Noda, H., & Takanashi, S. (2002). *Modaritii [Modality]*. Tokyo: Kurosio Publishers.

Nihongo Kijutsu Bunpoo Kenkyuukai (2003). *Gendai nihongo bunpoo [Contemporary Japanese grammar]* (Vol. 4). Tokyo: Kurosio Publishers.

Oshima, D. Y. (2013). Nihongo ni okeru intoneeshongata to shuuzyosi kinoo no sookan ni tsuite [On the correlation between the intonation types and the functions of discourse particles in Japanese]. *Forum of International Development Studies*, *43*, 47–63.

Portner, P. (2007). Imperatives and modals. *Natural Language Semantics*, *15*, 351–383.

Potts, C. (2005). *The logic of conventional impricatures*. Oxford: Oxford University Press.

Potts, C. (2007). The expressive dimension. *Theoretical Linguistics*, *33*, 165–197.

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press.

Shirakawa, H. (1993). "Hatarakikake" "toikake" no bun to shuujoshi "yo" no kinoo [Imperatives, interrogatives, and the function of the sentence-final particle yo]. *Bulletin of the Department of Teaching Japanese as a Second Language, Hiroshima University*, *3*, 7–14.

Takubo, Y., & Kinsui, S. (1997). Discourse management in terms of mental spaces. *Journal of Pragmatics*, *28*, 741–758.

Tanaka, S., & Kubozono, H. (1999). *Nihongo no hatsuon kyooshitsu: Riron to renshuu [Introduction to Japanese pronunciation: Theory and practice]*. Tokyo: Kurosio Publishers.

Uyeno, T. (1972). Shuujoshi to sono shuuhen [On Japanese sentence particles]. *Nihongo Kyooiku*, *17*, 62–77.

Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 172–200). Oxford: Oxford University Press.

A Question of Priority

Robert van Rooij and Katrin Schulz

Abstract Properties as set of individuals, or of features? Worlds, or propositions? Time-points, or events? Preference, or choice? Natural kinds, or similarity? In modern analytic philosophy it is standard to take (i) individuals as basic, and properties as defined in terms of them; (ii) worlds as basic, and propositions as defined in terms of them; (iii) time-points as basic, and intervals as constructions out of them; (iv) preference as basic, and optimal choice as defined in terms of them; and (v) natural kinds as basic, and similarities as defined in terms of them. In this chapter we show that in all cases the other direction is possible as well. Most of the constructions used are well-known. But by putting them collectively on the table we hope to show that the constructions have something in common, and that it is not always clear which perspective is ontologically less committing.

1 Properties: Sets of Individuals or of Features?

Logic started with Aristotelian syllogistic reasoning. Almost all modern textbooks use Venn diagrams, and thus an *extensional semantics*, to decide whether a syllogistic inference is valid. It is basically assumed that the terms denote non-empty sets of individuals. Somewhat more generally, the idea is to start with a partially ordered set $\langle \mathcal{A}, \leq \rangle$ (where '$\leq$' is a reflexive, transitive, and anti-symmetric relation on $\mathcal{A}$) such that all subsets B of $\mathcal{A}$ have a greatest lower bound. (b is the greatest lower bound of B iff (i) $\forall x \in B{:}b \leq x$, and (ii) $\forall y \in \mathcal{A}$: if $\forall x \in B{:}y \leq x$, then $b \leq y$). The greatest lower bound of $\mathcal{A}$ itself we denote by $\perp$. We assume that every

We would like to thank the anonymous reviewers for this volume, as well as the volume editors

R. van Rooij (✉) · K. Schulz
Institute for Logic, Language and Computation, Universiteit van Amsterdam,
Oude Turfmarkt 141-147, Amsterdam, The Netherlands
e-mail: R.A.M.vanRooij@uva.nl

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*, Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_13,

term T of the language denotes an element of $A - \{\bot\}$, i.e., $[[T]] \in \mathcal{A} - \{\bot\}$. A universal sentence like 'All S are P', represented as SaP, is counted as true iff $[[S]] \leq [[P]]$. A particular sentence like 'An S is P', represented by SiP, is counted as true iff the greatest lower bound of $[[S]]$ and $[[P]]$ is not equal to $\bot$, $glb\{[[S]], [[P]]\} \neq \bot$.

According to Leibniz, an extensional semantics is not the most natural way to interpret the meaning of syllogistic terms and syllogistic inference. Rather—or so he proposed—we should start out with an *intensional* semantics. According to it, SaP is true if the *intension* of P is contained in the *intension* of S. Following Rescher (1954), we will think of the intension of S as the set of attributes associated with S, and will call a semantics *intensional* iff it doesn't crucially refer to individuals. One way to think of this—in fact, this is arguably exactly how Leibniz thought of it—is to start out with a (semi-)lattice, instead of with a partial order. Indeed, let us start with a semi-lattice $\langle \mathcal{A}, \circ \rangle$, where '$\circ$' is a binary operation on $\langle \mathcal{A}$ that is idempotent ($\forall x \in \mathcal{A}{:}x \circ x = x$), commutative ($\forall x, y \in A{:}x \circ y = y \circ x$), and associative ($\forall x, y, z \in \langle \mathcal{A}{:}(x \circ y) \circ z = x \circ (y \circ z)$). In terms of this structure, we can say that SaP is true iff $[[S]] \circ [[P]] = [[S]]$. To also interpret particular sentences, we need to assume that the semi-lattice also has an *extreme element* e, meaning that $\forall x \in \mathcal{A}{:}x \circ e = e$. Now we can say that SiP is true iff $[[S]] \circ [[P]] \neq e$.

What should be the intuitive interpretation of '$\circ$'? Meet, or join? In fact, it doesn't matter much. It is clear that if we interpret $\langle \mathcal{A}, \circ \rangle$ as a meet semi-lattice, it corresponds exactly with the partially ordered set $\langle \mathcal{A}, \leq \rangle$ closed under greatest lower bounds. The extreme element e of the semi-lattice would intuitively correspond with $\bot$, a primitive notion of *inconsistency*. But we might as well interpret $\langle \mathcal{A}, \circ \rangle$ as a join semi-lattice, which would correspond with the same partially ordered set closed, this time, under smallest upper bounds. In fact, if we want to interpret Leibniz' semantics intensionally, the latter interpretation is the way to go. But what should in this case the extreme element e be associated with? By making use of an explicit intensional interpretation, we will explain that also now e should correspond with a primitive notion of inconsistency.

To give a set theoretic intensional semantics for syllogistics without negative terms, we have to start at least with a primitive set of attributes $\mathcal{A}$ and an interpretation function '$[[\cdot]]$' that assigns sets of attributes to terms. It is quite clear how to provide a semantics for sentences of the form SaP. This universal sentence is true iff $[[P]] \subseteq [[S]]$. But the problem is how to provide a semantics for particular sentences: SiP. The first idea that came to Leibniz's mind given in set-theoretic terms would be to say that SiP is true iff $[[S]] \cap [[P]] \neq \emptyset$. But this idea is clearly non-sensical: some bike is red, but there is nothing in the intension of 'red' that is also in the intension of 'bike', or so it seems. Or even more obviously, the sentence 'No gold is silver' is obviously true. According to the above suggestion this is true iff there is no attribute, or property, that gold and silver share. But there is obviously one: metal. What has to be assumed, rather, is the following idea: the intensions of 'red' and 'bike' are *not incompatible*. What this means is that also for our intensional interpretation, we must assume that a *primitive* relation of *(in)compatibility*. And the relation is one *between primitive features*.[1]

The fact that we have to assume such a notion of (in)compatibility already suggests why Leibniz had a hard time to come up with a satisfying characteristics for even simple syllogistic logic. Just like Wittgenstein when he was writing his Tractatus, also Leibniz thought of his simple terms, or attributes, as being *logical independent* of each other, i.e., their being mutually compatible with all other simples (cf. Ishiguro 1972, p. 54): only if the simples are logically independent of each other is it possible to construct a language where inference and equivalence can be checked 'from the surface'. To check validity we don't have to know what the interpretation of the different terms is. But if all terms are interpreted by sets of these simple attributes, a sentence like 'No gold is silver' can never be true. In our following interpretation, we assume an incompatibility relation $\perp$.

Let $M = \langle \mathcal{A}, [[\cdot]], \perp \rangle$ be a model, with $\mathcal{A}$ a set of attributes, $[[\cdot]]$ an interpretation function which assigns to each primitive term T a subset of A, $[[T]] \subseteq \mathcal{A}$, and $\perp$ a symmetric and irreflexive relation between elements of $\mathcal{A}$. If for two elements $x, y \in \mathcal{A}$ it holds that $x \perp y$, we say that the attributes x and y are incompatible. We will denote by Δ the set of subsets of $\mathcal{A}$ which contain such mutually incompatible elements: $\Delta = \{S \subseteq \mathcal{A}: \exists x, y \in S: x \perp y\}$. We assume that for each primitive term T, $[[T]] \notin \Delta$, and that the set of supersets of $[[T]]$ does not equal the set of all maximally consistent sets of $\mathcal{A}$. Now we say that $[[SaP]] = 1$ iff $[[S]] \cup [[P]] = [[S]]$ iff $[[S]] \supseteq [[P]]$, and $[[SiP]] = 1$ iff $[[S]] \cup [[P]] \notin \Delta$. Thus, SiP is true iff S and P do not contain mutually incompatible attributes. SoP and SeP are interpreted as the negations of SaP and SiP, respectively. If we say that $\phi_1, \ldots, \phi_n \models \psi$ iff for all models in which the premisses are true, the conclusion is true as well, this semantics validate all and only all arguments in classical syllogistic style if and only if they are traditionally counted as valid.

2 Worlds and Individuals: Primitives, or Maximal Sets?

In modal logic it is standard to think of worlds as primitive entities. Lewis (1973, 1986) even believed that they are universes that really exist. But if we want to prove completeness results one does this by thinking of worlds as maximally consistent sets of sentences or propositions. In that case one *defines* worlds in terms of other primitives: propositions and a notion of (in)consistency. And indeed, a number of authors have proposed to think of possible worlds in exactly this way, as total state descriptions (e.g. Carnap 1947). Lewis (1973) used a primitive similarity relation between possible worlds to account for counterfactual conditionals. Proponents of premisse semantics (Veltman 1976; Kratzer 1981) have argued that it is more illuminative

[1] If we think of the extensional counterpart, this means that 'some bike is red' is true not because there actually exists a red bike, but rather that it is possible that such a bike exists. And indeed, what Leibniz considers to be the extension of a term (a set of individuals scattered around all worlds) is very much what in possible worlds semantics is its intension (cf. Leibniz 1966 and Ishiguro 1972, p. 49).

to *construct* such a comparative similarity relation in terms of sets of propositions. Turner (1981) went even so far that these propositions themselves should be taken as primitive, instead of the possible worlds. We have seen above that Leibniz favored an interpretation of syllogistic logic in which the terms denote sets of features, instead of sets of individuals. But how then to think of individuals? Although it is not very clear whether he takes individuals to be primitive as well, he at least also thinks of individuals as maximally consistent sets of properties (Leibniz 1686). Indeed, in agreement with the ontology of syllogistic reasoning, Leibniz doesn't seem to make a clear ontological distinction between properties and individuals, it is just that individuals are the maximally consistent ones.

For Leibniz, all properties were on a par. Indeed, properties were taken to be closed under logical operations like complementation, disjunction, and conjunction.[2] The early Russell was strongly influenced by Leibniz, and during his 'realistic' period when Russell believed in the existence of real universals, he also wanted to define (in Russell 1912) individuals in terms of properties. However, in constrast to Leibniz, he didn't want to use *all* properties, only the natural ones: the universals. In distinction with standard properties, universals are *not* taken to be closed under complementation and disjunction. Thus, also Russell defined individuals somehow as maximal sets of properties, constrained by a notion similar—though not identical—to that of consistency. The notion now was that of the primitive relation of '*compresence*'. Even though constructivists have to take a notion like 'consistency' to be primitive, we have an intuitive understanding of what it is supposed to be. For 'compresence' this is somewhat more difficult. Still, the notion of compresence is very much like consistency: it is a reflexive and symmetric relation, though now restricted to properties. Intuitively, the compresence relation should just like consistency not be transitive: we can imagine two distinct individuals, one having only universals P and Q, and a second only having Q and R. If compresence were transitive, it would mean not only that we also had an individual with universals P and R, but also that if we think of individuals as *maximal* compresent sets of properties, there would in fact be only one individual having all properties P, Q, *and* R. There could be no two different individuals sharing a single property. But thinking of the 'compresence' relation as being non-transitive still gives rise to problems. Intuitively, we can imagine three distinct individuals, one having only universals P and Q, one only having properties Q and R, and a third only having properties P and R. But if we think of individuals as *maximal* sets of compresent properties, this is impossible: in such cases there also *has to be* a fourth individual having all three properties. For the construction of individuals out of natural properties this problem didn't receive a lot of attention. The very similar problem discovered in the trial to construct universals, or natural properties, out of particulars together with a primitive similarity relation, however, received a lot of attention. We will discuss this problem in a later section. Before that, however, we will consider first a more successful constructive move made by Russell to think of instants as maximal sets of intervals.

[2] This is what he believed, but he was not able to work out a full semantics for syllogisms with complex terms.

3 Instants, or Events?

Events It is standard to take instant time points, and the 'before'-relation between them, to be basic, and to define intervals in terms of them as convex sets of instants. Russell (1914, 1936), however, proposed to go the other way around: temporal instants should be constructed from what he calls events. His motivation is that he wants to show that our conception of abstract instants is derived from (reconstructed from) the events and the temporal relationships we perceive. Similar things have been done by Whitehead and Wiener (1914), and more recently by Kamp (1979), van Benthem (1991), and Thomason (1984), and after that in AI by people like Allen and Hayes (1985). They all start with an *event structure*. An event structure, $\langle E, \prec \rangle$ is just a set of events that is temporally ordered: $e \prec e'$ means that e is temporally completely before e'. According to some, events should just be strictly partially ordered (irreflexive and transitive). Most authors, however, assume something stronger. They assume that events give rise to what is known as an *interval order*:

Definition 1 An *interval order* is a structure $\langle X, R \rangle$, with R a binary relation on X that is irreflexive, and satisfies the interval order condition (IO):
(IR) $\forall x{:}\neg R(x, x)$.
(IO) $\forall x, y, v, w{:}(R(x, y) \wedge R(v, w)) \rightarrow (R(x, w) \vee R(v, y))$.

Notice that any interval order is also a strict partial order, because from (IR) and (IO) one can immediately derive that the ordering is also transitive.

The difference between starting with a strict partial order or with an interval order is that according to the former it might be unclear how some events are temporally related to one another, while this is (almost) impossible according to the latter approach. To see this, let us define a relation '$\sim$' as follows: $e \sim e'$, iff$_{def}$ $e \not\prec e' \wedge e' \not\prec e$. Notice that from this definition it follows that (i) $\sim$ is reflexive and symmetric, but need not be transitive, (ii) $\prec$ and $\sim$ are disjoint, and (iii) $\prec \cup \sim$ is complete. If we take $\langle E, \prec \rangle$ to be an interval order, '$\sim$' intuitively represents 'temporal overlap'. But if $\langle E, \prec \rangle$ is just a strict partial order, it might also be that $e \sim e'$ because the events are temporally *incomparable*. To illustrate this, it is possible (Fishburn 1970) to represent interval orders in terms of, yes, *intervals* of the real line: f is a function from events to intervals of $\mathbf{R}$. $x \prec y$ then means that all points of $f(x)$ are before all points of $f(y)$, and $x \sim y$ means that $f(x)$ and $f(y)$ have a non-empty intersection. Such a representation is *not* possible for strict partial orders: even if $x \prec y$ and $v \prec w$ it might still be possible that both x *and* y are incomparable with both v *and* w.

In terms of event structures we can define a relation of *temporal inclusion* between events '$\sqsubseteq$'. If $\langle E, \prec \rangle$ is an interval order, we can define $e \sqsubseteq e'$ iff$_{df}$ iff $\forall e''{:}e'' \sim e \rightarrow e'' \sim e'$. It is easy to see that now '$\sqsubseteq$' is reflexive and transitive, and that $\sqsubseteq$ means 'temporally included' as intended if '$\sim$' means 'temporal overlap'. But the latter is only the case if $\langle E, \prec \rangle$ is an interval order.

We can also define a notion of temporal inclusion between events if $\langle E, \prec \rangle$ is a strict partial order. Define $e \sqsubseteq e'$ iff$_{def}$ $\forall e''[e' \prec e'' \rightarrow e \prec e''] \wedge \forall e''[e'' \prec e' \rightarrow$

$e'' \prec e$]. It is easy to prove that '$\sqsubseteq$' is reflexive and transitive, and thus a *pre-order*. But it need not satisfy antisymmetry, and thus '$\sqsubseteq$' does not (necessarily) give rise to a partial order.

Events structures can be *atomic* or give rise to *endless descent*. Event structure $\langle E, \prec\rangle$ is atomic iff $\forall e \in E{:}\exists e' \sqsubseteq e{:}\forall e'' \sqsubseteq e'{:}e'' = e'$. An event structure gives rise to endless descent iff $\forall e \in E{:}\exists e' \sqsubseteq e{:}e' \neq e$.

Intervals In terms of the pre-order, '$\sqsubseteq$', we can define a new relation, '$\approx$', as follows: $e \approx e'$ if$_{def}$ $e \sqsubseteq e' \wedge e' \sqsubseteq e$. It is obvious that this relation is an equivalence relation. Now we can define *intervals* as equivalence classes of events, and we can derive a new structure, $\langle [E]_\approx, <, \sqsubseteq^*\rangle$, with $I < J$ iff$_{def}$ $\exists e \in I, e' \in J{:}e \prec e'$ and $I \sqsubseteq^* J$ iff$_{def}$ $\exists e \in I, e' \in J{:}e \sqsubseteq e'$. One can show that if $\langle E, \prec\rangle$ is a strict partial order/interval order, then (i) $\langle [E]_\approx, \sqsubseteq^*\rangle$ is a partial order, and (ii) $\langle [E]_\approx, <\rangle$ is a strict partial order/interval order. An interval structure is atomic, or gives rise to endless chains iff the corresponding event structure is.

Instants Now Russell defines instants as maximal sets of pairwise overlapping intervals.

Definition 2 Let $\Sigma_I = \langle [E]_\approx, <\rangle$ be an interval structure. An *instant* t is a subset of $[E]_\approx$ such that: (i) $\forall I, J \in t{:}I \sim J$ and (ii) $\forall I \notin t{:}\exists J \in t{:}I \not\sim J$ (in other words, an instant is a *maximal* subset of $[E]_\approx$ where $\forall I, J \in t{:}I \sim J$).

We denote the set of instants of Σ_I as $T(\Sigma_I)$.

Definition 3 Let t and t' be any two instants of $T(\Sigma_I)$. Then: $t <^* t'$ iff$_{def}$ $\exists I \in t{:}\exists J \in t'{:}I < J$. We call $\tau(\Sigma_I) = \langle I(\Sigma_I), <^*\rangle$ the instant structure derived from Σ_I.

Most important in Russell's construction is the following theorem:

Theorem 1 *$\tau(\Sigma_I)$ is a linear order, if Σ_I is an interval order. (It is a strict partial order if Σ_I is).*

A linear order demands that all elements of a set are comparable. In this sense, linear orders are very informative. Intuitively, however, it seems that the ordering between instants is stronger than a linear order. For one thing, the order seems to be *dense*. An ordering $\langle X, R\rangle$ is dense iff $\forall x, y{:}R(x, y) \rightarrow \exists z{:}R(x, z) \wedge R(z, y)$. The set of natural numbers is not dense, because there is no natural number between, say, 1 and 2, but the set of rational numbers is. Russell showed that the density of the ordering relation between instants can be derived from some constraints on the ordering between events (intervals). To account for density, Russell (1936) proposed the following constraint on interval orders:

Let $\Sigma_I = \langle [E]_\approx, <\rangle$ be an interval order. The following condition suffices to make the ordering on instants $\tau(\Sigma_I) = \langle I(\Sigma_I), <^*\rangle$ being dense[3]:

[3] Interestingly, Allen and Hayes define the notion of 'meet' ':', as follows:
$I{:}J$ iff$_{def}$ $I < J \wedge \neg\exists K, L(I < K \wedge K \sim L \wedge L < J)$.

$$\text{For all } I, J \in [E]_{\approx}{:}I < J \rightarrow \exists K, L \in [E]_{\approx}{:}I < K \sim L < J.$$

Although this condition is sufficient, it is *not necessary* for $\tau(\Sigma_I)$ being dense.[4] If an ordering relation is not dense, it is *discrete*. There exists an alternative way to say that $\tau(\Sigma_I) = \langle I(\Sigma_I), <^* \rangle$ is dense/discrete. It is easy to show that in case $\langle E, < \rangle$ is atomic, the ordering relation between instants is discrete, and in case the interval structure gives rise to endless descent, the ordering relation between instants is dense.

From instants to intervals and back Suppose we start with a primitive instant structure $\langle T, <^{**} \rangle$, where T is a set of instants, and $<^{**}$ a primitive relation between instants. It is very natural to assume that $\langle T, <^{**} \rangle$ is a strict partial order or even a linear order. Now we can define intervals as non-empty convex subsets of T, where I is *convex* iff $x \in I \wedge z \in I \wedge x <^{**} y <^{**} z \rightarrow y \in I$. More in particular, assume that i and j are points, then we can define an interval as follows: $\{t \in T{:}i <^{**} t \wedge t \leq^{**} j\}$.[5] Let us now take $I(T)$ to be the set of intervals so constructed. Define the interval structure based on $\sigma_T = \langle T, < \rangle$, $\iota(\sigma_T)$, as the structure $\langle I(T), <, \sqsubseteq \rangle$ where $I < J$ iff $\forall i \in I, j \in J{:}i <^{**} j$ and $I \sqsubseteq J$ iff $I \subseteq J$. Now one can show that in the structure $\langle I(T), <, \sqsubseteq \rangle$, '$<$' is an interval order if '$<^{**}$' is a linear order, $\sqsubseteq$ a partial order that satisfies (CONJ), and where '$<$' and '$\sqsubset$' satisfy (MON) and (CONV). Moreover, $\langle I(T), < \rangle$ is atomic if $\langle T, <^{**} \rangle$ is discrete, and it has endless descent if $\langle T, <^{**} \rangle$ is dense.

Above we have seen that from interval orders we can derive a linearly ordered instant structure. Thus, this can also be done for $\iota(\Sigma_T)$. Call the result $T(\iota(\Sigma_T))$. From $T(\iota(\Sigma_T))$ we can derive an interval order again, and it is possible to show that $\iota(\Sigma_T)$ is isomorphic to this new interval order. From this new interval order we can derive an instant structure again, and it is possible to show that this new instant structure is isomorphic to $T(\iota(\Sigma_T))$. This process can be continued indefinitely.

4 Orderings, or Choice?

Maximizing choice functions In the theory of choice it is standard to take a comparative preference order to be basic. Analoguously, in the possible world theory of counterfactuals of Lewis (1973), a comparative similarity relation between worlds is taken to be primitive. However, a much discussed topic in the theory of choice is how a preference order among options can be derived on the assumption that the notion of *choice* is primitive. In the semantic analysis of counterfactuals

---

[4] For a necessary condition, see Lück (2006). In this paper it is also proven under which circumstances one can generate a *continuous* order of instants.

[5] Of course, it is not necessarily to define intervals as having open beginnings and closed ends. The other way is possible as well. Just to assume that it any convex set is an interval doesn't give rise to endless descent even if $\langle T, <^{**} \rangle$ is dense.

(and of belief revision, for instance) a similar question has been addressed: can we define an ordering from natrual constraints on choice functions? And, of course, this is how Stalnaker (1968) started.

Assuming a choice function that selects an element from each finite set of options, one can easily show how we can generate a linear order by putting constraints on how this function should behave on different sets of options. Let us define a *choice structure* to be a triple $\langle X, O, C\rangle$, where X is a non-empty set, the set O consists of all finite subsets of A, and the choice function C assigns to each finite set of options $o \in O$ an element of o, $C(o)$, satisfying the following condition:

(LIN) $\forall o, o' \in O$:If $(C(o) \in o'$ and $C(o') \in o)$, then $C(o) = C(o')$.

If we say that $x > y$, iff$_{def}$ $C(\{x, y\}) = x$, one can easily show that the ordering as defined above gives rise to a *linear order*.

Arrow (1959) already showed how we can generate a strict weak ordering by putting other constraints.

Definition 4 A *strict weak order* is a structure $\langle X, P\rangle$, with P a binary relation on X that is irrreflexive (IR), transitive (TR), and almost connected (AC):
(IR) $\forall x{:}\neg P(x, x)$.
(TR) $\forall x, y, z{:}(P(x, y) \wedge P(y, z)) \rightarrow P(x, z)$.
(AC) $\forall x, y, z{:}P(x, y) \rightarrow (P(x, z) \vee P(z, y))$.

In this case, the choice function C assigns to each finite set of options $o \in O$ a *subset* of o, $C(o)$. Arrow (1959) stated the following principle of choice (C), and the constraints (A1) and (A2) to assure that the choice function behaves in a 'consistent' way:

(C) $\forall o \in O{:}C(o) \neq \emptyset$.
(A1) If $o \subseteq o'$, then $o \cap C(o') \subseteq C(o)$.
(A2) If $o \subseteq o'$ and $o \cap C(o') \neq \emptyset$, then $C(o) \subseteq C(o')$.

If we say that $x > y$, iff$_{def}$ $x \in C(\{x, y\}) \wedge y \notin C(\{x, y\})$, one can easily show that the ordering as defined above gives rise to a *strict weak order*.

Condition (A1) is better known as Sen (1971) *Property* α. Condition (A2) is also known as Sen's Property β^+. Taken together with "(EMPTY) If $o \subseteq o'$ and $C(o') = \emptyset$, then $C(o) = \emptyset$" (also assumed by Lewis, and which follows from (C)), it implies both (II) and (III) discussed below. Arrow formulated the choice function as the combination of (A1) and (A2), and called it the axiom of independence of irrelevant alternatives:

(A) If $o \subseteq o'$ and $o \cap C(o') \neq \emptyset$, then $C(o') \cap o = C(o)$.

While condition (A1) expresses some kind of 'contraction consistency' in proceding from larger menus to smaller ones, the following condition proceeds from smaller menus to larger ones:

(II) $C(o) \cap C(o') \subseteq C(o \cup o')$.[6]

The following axiom is known as *Aizerman's axiom*:

(III) If $o \subseteq o'$ and $C(o') \subseteq o$, then $C(o) \subseteq C(o')$

Taken together with (A1), the superset axiom implies (III). Condition (III) is independent of condition (II), even in the presence of condition (A1).

Suppose that O consists of all finite subsets of I. We can state facts like the following: Ordering $<$ is *acyclic*, if C satisfies EMPTY and (A1); the ordering is *transitive*, if C also satisfies (III). If C satisfies EMPTY and (A2), then $<$ is almost connected. Another sufficient condition for $<$ to be almost connected is for C to be closed under arbitrary union and satisfying (EMPTY) and (A2). Thus, for natural properties the preference relation has, there correspond 'natural' constraints on choice functions. It is not clear what should be taken as primitive.

Satisficing choice functions We would like to derive the meaning of 'better than' in terms of the meaning of 'best'—as is assumed if agents are taken to be utility maximizers—, but rather to derive the meaning of 'better than' in terms of the context-dependent meaning of 'good'.[7] What is crucial for the interpretation of the results of our chapter is that although 'good' seems to obey axiom (A2), axiom (A1) seems much too strong: (A1) demands that if both x and y are considered to be good in the context of $\{x, y, z\}$, both should considered to be good in the context $\{x, y\}$ as well. But that is exactly what we don't want for a context dependent notion of 'good': in the latter context, we want it to be possible that only x, or only y, is considered to be good. We should conclude that if we want to characterize the behavior of 'good', we should give up on (A1). Unfortunately, by just constraints (C) and (A2) we cannot guarantee that the comparative relation 'better than' behaves as desired. In particular, we cannot guarantee that it behaves almost connected.

To assure that the comparative behaves as desired, we add to (C) and (A2) the Upward Difference-constraint (UD), proposed by van Benthem (1982). To state this constraint, we define the notion of a difference pair: $\langle x, y\rangle \in D(o)$ iff$_{def}$ $x \in C(o)$ and $y \in (o - C(o))$. Now we can define the constraint:

(UD) $o \subseteq o'$ and $D(o') = \emptyset$, then $D(o) = \emptyset$.

In fact, van Benthem (1982) states the following constraints: No Reversal (NR), Upward Difference (UD), and Downward Difference (DD) (where o^2 abbreviates

---

[6] This axiom is a finitary version of Sen's Property γ.

[7] Interestingly enough, this is exactly analogue to what Klein (1980) intended to do in linguistics: the meaning of 'taller than' (or 'better than') should be defined in terms of the meaning of 'tall' (or 'good'), not that of 'tallest' (or 'best').

$o \times o$, and $D^{-1}(o) =_{def} \{\langle y, x\rangle : \langle x, y\rangle \in D(o)\}$):

(NR) $\forall o, o' \in O : D(o) \cap D^{-1}(o') = \emptyset$.
(UD) $o \subseteq o'$ and $D(o') = \emptyset$, then $D(o) = \emptyset$.
(DD) $o \subseteq o'$ and $D(o) = \emptyset$, then $D(o') \cap o^2 = \emptyset$.

One can show that if constraints (NR), (UD) and (UD) are satisfied, the preference relation '>' as defined before still has the same properties as before: it is still predicted to be a strict weak order. In van Rooij (2011) it is shown that other constraints on the satisfying choice function gives rise to other ordering structures, making it less than obvious, again, to assume that either the one (preference order), or the other (choice), should be taken as primitive.

5 Universals, or Similarity?

There are not many laws named after philosophers. But Leibniz' law is an exception: it states that 'two' objects are *identical* if and only if they are indescernible, i.e. when they share *all* their properties. The notion of 'similarity' is closely related with that of 'identity', and Leibniz expressed ideas about this notion as well: he claimed that x is similar to y if and only if x and y share *at least one* property. Goodman (1972) famously argued that the notion of similarity thus defined is useless. Assuming that properties are sets, and that all sets are on equal footing, it indeed follows immediately that any two objects have a property (set) in common. On a similar assumption one can also easily prove that even the comparative notion of similarity 'y is more similar to x than z' is useless: there are exactly as many sets of which x and y are elements than there are of which x and z are elements. Thus, we should not work with any notion of 'overall similarity'. At best, we should have a *relative* notion of similarity: 'x is similar to y *in respect* r'. Unfortunately, according to Goodman, once we introduce such 'respects' the notion of 'similarity' plays no role anymore: the respects do all the work. Suppose, for instance, that we say that we take r to be 'red'. What then would be the use of similarity? 'x is similar to y *in respect* r' would now be true just because both x and y are red.

There is a lot to say about Goodman's arguments, most obviously his equation of properties with sets. But we are not so much interested in the issue how to *define* a notion of similarity. Our major concern is what we can do with it, once we have such a notion. We will discuss whether we can explain natural properties, or universals, in terms of them (Lewis 1983; Armstrong 1989).

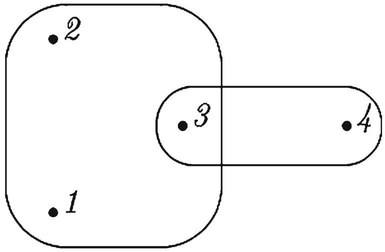
The problem of universals What makes it that we can 'group' several objects or individuals together under a general term, and that it is more natural to divide the world up in one way than in another? This is basically the very old problem of universals that is with us ever since Plato (1941) and Aristotle (1941) wrestled with it. For Plato and Aristotle the answer to the problem was (relatively) simple: we divide the things around us up in the way we do because this is the way *reality* is cut to its joints. According to *realists* like them, we classify the world up in cats, dogs, humans and trees, because this is the way the world is divided, and x and y are both cats, because they both have a property in common: cathood. This picture is indeed very natural. The animal kingdom, for instance, is divided in species and what these species are has to be *discovered* by the methods of classification. Individuals do not belong to the same species because they are similar, but they are similar because they belong to the same species, just as Leibniz had it. Thus, a particular classification could be shown to be wrong. Natural as it may seem, the realist' position also gives rise to difficulties: how should we think of these species exactly? Is it something that exists in addition to the cats themselves? And can it even exist without being instantiated? This last question seems unnatural for species, but is more natural for the properties denoted by other general terms, including adjectives. Plato answered all these questions positively, but never made it very clear what it means to 'have' a property. Aristotle, on the other hand, was a more down-to-earth philosopher than Plato, and didn't like uninstantiated universals. But could, as a result, not make very clear what he assumed species and other universals to be.

Suspicious of abstract objects for which they could not determine identity conditions, nominalists rejected the existence of universals. According to them, nature produces individuals and nothing more. Species have no actual existence in nature, they are not the 'objects' we denote by general terms. General terms have just been invented in order that we be able to refer to a great numbers of individuals collectively. General terms can be used because we classify the world. This classification is conventional and can be neither right nor wrong. It is not a theory, but merely a way of summarizing information in an inteligible form. One assesses its value by consideration of its *usefulness*.

Still, it appears that the conventional classification of the objects around us in groups has to be based on something. The animals of the world can be classified in infinitely many ways. How come that all cultures divide up this class in similar ways in terms of cats, dogs, etc.? One suggestion would be that this says a good deal about the ways we humans are able to classify. This is no doubt true, but if we gathered this ability by natural selection (Quine 1969), it must say a good deal about the world as well: why else could this way to categorize be so useful? A nominalist, however, doesn't want to say that this means that our categories correspond with real species out there in the world, as the realist has it. He will say at most that what we find in the world is a notion of 'similarity'. His project is to explain how our categorization and our use of general terms works by distingishing 'natural' groupings from 'unnatural' ones, by defining 'natural' groups (or sets) in terms of a primitive notion of similarity.

Resemblance Nominalism According to nominalism, properties are just sets. But not all sets are *real* properties, the properties that *cut nature at its joints*. These real properties are defined in terms of similarity. This type of nominalism has a long history, but the first one who seriously tried to work it out was Carnap in his *Logische Aufbau der Welt* (1923; 1928). In this famous book he proposed that real properties are *maximal resemblance sets*.

Fig. 1 Set of properties



Let us start with a similarity structure $\langle X, R\rangle$, where R is some kind of similarity relation between the objects in X. Let us call each set like P for whom it holds that $\forall x \in P{:}\forall y \in P{:}R(y, x)$ a *resemblance set*, an element of $\mathcal{R}$. The maximal resemblance set are those elements of $\mathcal{R}$ such that there is no other element $Q \in \mathcal{R}$ such that $P \subset Q$. Thus, a maximal resemblance set is a *maximal* set of individuals each of which resemble each other. Can we think of such a maximal resemblance set as a natural property?

Suppose that similarity is an equivalence relation '$\approx$'. The similarity structure is then one of the form $\langle X, \approx\rangle$. As is well-known we can determine properties as equivalence classes: $Q = \{[x]_{\approx}{:}x \in X\}$. We can also go from partition to equivalence relation: $x \approx y$ iff $\exists q \in Q{:}x, y \in q$. But it is interesting to observe that the set of maximal similarity sets as defined above is exactly the set of equivalence classes that partition X, if the similarity relation is an equivalence relation. Is it natural to call these sets *natural properties*? Not really. For (i) there are intuitively natural properties P and P' that can overlap each other, $P \cap P' \neq \emptyset$, and (ii) there are intuitively natural properties P and P' where the one is a proper subset of the other, $P \subset P'$. Neither of those possibilities is allowed, if we assumed that the similarity relation involved is an equivalence relation, i.e., a relation that is *reflexivie*, *symmetric*, and *transitive*.

Carnap was aware of this, and for this reason he assumed that the similarity relation '$\sim$' is just reflexive and symmetric. In that case the similarity structure $\langle X, \sim\rangle$ is sometimes called a *tolerant space*. In terms of it one can determine similarity sets and maximal similarity sets as before.[8] But now the set of maximal similarity sets need not form a partition. (Note: Russell's construction of instants as maximal sets of overlapping events is special case. Though now this is a partition). Instead, what results is just a *cover*, where Q is a **cover** of X iff (a) Q is a set of subsets of X; (b) $\emptyset \notin Q$ and (c) $\forall x \in X{:}\exists q \in Q{:}x \in q$ (i.e. $\bigcup Q = X$). Now we say that $\langle X, Q\rangle$ is a **property structure** over X iff (i) $X \neq \emptyset$ and (ii) Q is a cover of X. Then we can determine a similarity relation as follows: $x \sim y$ iff$_{df}$ $\exists q \in Q{:}x, y \in q$.

Consider the set of objects $X = \{1, 2, 3, 4\}$ with the set of properties Q consisting of two properties: $\{1, 2, 3\}$ and $\{3, 4\}$ (Fig. 1).

This set of properties gives rise to the following similarity structure (closed under reflexivity and symmetry): $\langle X, \{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 1, 3\rangle, \langle 3, 4\rangle\}\rangle$ (Fig. 2).

[8] The existence of maximal similarity sets is, in general, guaranteed by Zorn's Lemma.

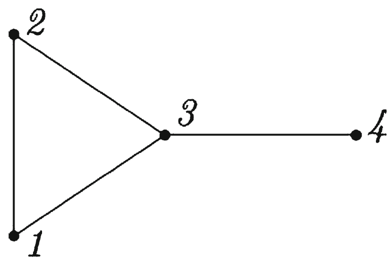


Fig. 2 Corresponing similarity relation

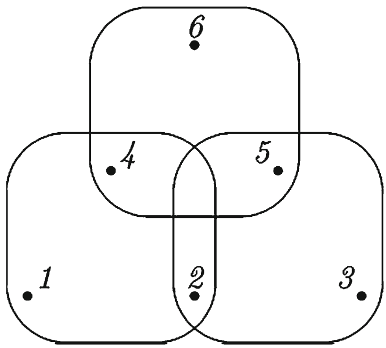


Fig. 3 Imperfect community 1: properties

In this case, Q is the set of maximal similarity sets w.r.t. the similarity structure. Unfortunately, it is not in general the case that we can recover the original set of properties as the set of maximal similarity sets. We have the *no-uniqueness* problem: not any set of properties can be adequatly represented by a similarity set. There are two problems: *the problem of imperfect community* and the *companionship difficulty*.

Imperfect community: Suppose that we start with cover $Q = \{\{1, 2, 4\}, \{2, 3, 5\}, \{4, 5, 6\}\}$. The similarity relation is then $\langle\{1, 2, 3, 4, 5, 6\}, \{\langle 1, 2\rangle, \langle 1, 4\rangle, \langle 2, 3\rangle, \langle 2, 5\rangle, \langle 3, 5\rangle, \langle 2, 4\rangle, \langle 4, 5\rangle, \langle 4, 6\rangle, \langle 5, 6\rangle\}\rangle$. (as before, closed under reflexivity and symmetry). Let $r(x)$ be the function that maps x to the maximal similarity sets in which x is part. Now $\{r(x): x \in X\} = \{\{1, 2, 4\}, \{2, 4, 5\}, \{2, 3, 5\}, \{4, 5, 6\}\} \neq Q$. Goodman states that this Q exihibits the difficulty of imperfect community': the similarity relation defined via Q gives rise to some maximal similarity sets that are not properties (Figs. 3,4).

In the above example we had more maximal similarity sets than original properties. But at other times, the converse problem will appear: we end up with less maximal similarity sets than we had original properties. As a minimal example, consider cover $Q = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$ (Fig.5). The similarity structure then is $\langle\{1, 2, 3\}, \{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 1, 3\rangle\}\rangle$. But the set of maximal similarity sets derived from this is just $\{\{1, 2, 3\}\} \neq Q$ (Fig.6).

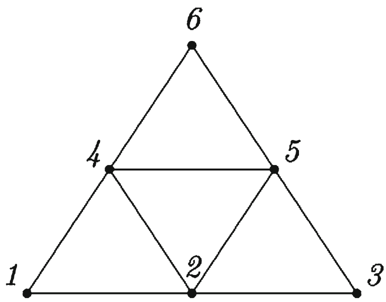


Fig. 4 Imperfect community 1: similarity

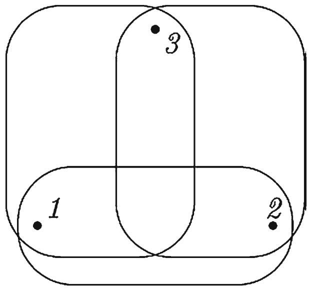


Fig. 5 Imperfect community 2: properties

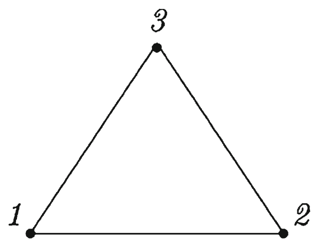


Fig. 6 Imperfect community 2: similarity

Companionship difficutly. Let us we start with cover $Q = \{\{1, 2\}, \{1, 2, 3\}\}$. The similarity structure then is $\langle\{1, 2, 3\}, \{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 1, 3\rangle\}\rangle$. But the set of maximal similarity sets derived from this is just $\{\{1, 2, 3\}\} \neq Q$.

Goodman (1953) famously argued that because of these problems it is hopeless to start with a binary similarity relation as basic: in this way we cannot *recover* the basic properties. But in a sense this criticism is *not fair*. For a real nominalist, there are no basic properties, and similarity is all there is. So, in a sense the above problems cannot even be stated. What can be stated at most is that not all covers of X can be sets of basic properties. But in fact, as pointed out by Leitgeb (2007), this is true for purely cardinality reasons:

The non-uniqueness problem is unsolvable, due to *cardinality problems*: the set of all covers of X is much greater than the set of all possible similarity relations on X. Thus, a unique binary similarity relation *cannot* be found for each cover. The problem is solvable, however, if we limit ourselves to covers Q that satisfy certain constraints. Berge (1989) proved that if Q satisfies the following constraints, constructing properties as maximal similarity sets is ok and we don't have a non-uniqueness problem.

$\langle X, \sim\rangle$ is faithful w.r.t. $\langle X, Q\rangle$ iff

(i) $\forall X, Y, Z \in Q, \exists P \in Q{:}(X \cap Y) \cup (X \cap Z) \cup (Y \cap Z) \subseteq P$, (imperfect community) and
(ii) $\neg\exists X, Y \in Q{:}X \subset Y$ (companionship).

These constraints show the **main problem** of Carnap's analysis: mainly due to (ii) such a Q cannot be thought of as set of natural properties, because it can't account for laws.

Goodman and Quine concluded that Carnap's binary similarity relation was *too weak*, we need at least a 3-place *comparative* similarity relation, or a 4-place similarity relation. Moreover, Goodman argued that we need similarity *in respect* (Gärdenfors 2000).

Rodriguez-Pereyra Rodriguez-Pereyra claims (1999; 2002) that we can solve Goodman's problem by making use of a more general similarity relation: similarity also relating *pairs*.

Suppose 1,2, and 3 all resemble each other (have property P) and so do 4 and 5 (have property Q). Then the pairs $\langle 1, 2\rangle$ and $\langle 2, 3\rangle$ resemble each other, but $\langle 1, 2\rangle$ and $\langle 4, 5\rangle$ do not. (In this whole section it is assumed that this is equivalent to say that the sets $\{1, 2\}$ and $\{2, 3\}$ resemble each other, but $\{1, 2\}$ and $\{4, 5\}$ do not.) But then also $\langle\langle 1, 2\rangle, \langle 2, 3\rangle\rangle$ resembles $\langle\langle 1, 2\rangle, \langle 1, 3\rangle\rangle$ but not $\langle\langle 1, 2\rangle, \langle 4, 5\rangle\rangle$. (Meaning that the sets $\{\{1, 2\}, \{2, 3\}\}$ and $\{\{1, 2\}, \{1, 3\}\}$ resemble each other, but the sets $\{\{1, 2\}, \{2, 3\}\}$ and $\{\{1, 2\}, \{4, 5\}\}$ do not.) Thus, $\langle x, y\rangle$ resembles$_1$ $\langle u, v\rangle$ iff x and y resemble both u and v. Similarly $\langle X, Y\rangle$ resembles$_1$ $\langle U, V\rangle$ iff X and Y resemble both U and V.

Definition: A set of objects P is a *perfect community* iff (i) all its members resemble each other, (ii) all pairs of members of P resemble each other, (iii) all pairs of pairs of members of X resemble each other etc.

Take the *imperfect community*: $Q = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$. If we want to represent it with just a similarity relation between individuals it gives rise to similarity structure $\langle\{1, 2, 3\}, \{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 1, 3\rangle\}\rangle$. The only maximal similarity set here is $\{1, 2, 3\}$, i.e., we cannot recover Q.

The *perfect community* $Q' = \{X'_0 = \{1, 2, 3\}\}$ gives rise to the same similarity structure and can be reconstructed.

But now let us also assume that we have a similarity relation between pairs. Such a relation can contain more information then one that only relates individuals. In particular, we can construct a similarity structure $\langle X, \sim'\rangle$ that faithfully represents the above mentioned perfect community Q' as follows: $\langle\{1, 2, 3\}, \{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 1, 3\rangle,$ $\langle\langle 1, 2\rangle, \langle 2, 3\rangle\rangle, \langle\langle 1, 2\rangle, \langle 1, 3\rangle\rangle, \langle\langle 1, 3\rangle, \langle 2, 3\rangle\rangle\}\rangle$. The similarity structure $\langle X, \sim\rangle$ that faithfully represents Q, on the other hand, is just $\langle\{1, 2, 3\}, \{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 1, 3\rangle\}\rangle$ and does not contain similarities between pairs. Notice that $P = \{1, 2, 3\}$ is a perfect community with respect to $\langle X, \sim' \rangle$ but not with respect to $\langle X, \sim \rangle$, because there are at least two pairs of members of P that do not resemble each other: $\langle 1, 2\rangle \not\sim \langle 1, 3\rangle$. Thus, we can recover Q by looking not at the maximal similarity sets, but rather at the perfect communities.

There is a simpler way to arrive at the same result. The reason is that Pereira Rodriguez seems to assume that if $\langle x, y\rangle \sim \langle x, z\rangle$ then it also holds that $\langle x, y\rangle \sim \langle x, z\rangle$ and $\langle y, z\rangle \sim \langle x, z\rangle$.[9] Now suppose that we start with a PR-similarity structure $\langle X, \sim_{pr}\rangle$. Define for each $x \in X$ the set $s^i(x)$ for each positive natural number i as follows: $s^1(x) = \{\{x, y\}{:}y \in X \ \& \ x \sim_{pr} y\}$; $s^{n+1}(x) = \{X \cup Y{:}X, Y \in s^n(x) \ \& \ X \sim_{pr} Y\}$. $s(x)$ is now defined as the fixed point of this sequence. The set of properties Q is now defined as $Q = \{s(x){:}x \in X\}$.

To go the other way, we have to define the similarity relation in terms of a cover Q such that after recovering the similarity relation again, we can recapture the same cover Q again. Define for each $x \in X$, $f(x)$ as $\{q \in Q{:}x \in q\}$. For pairs and higher we define it as follows: $f(\langle x, y\rangle) = f(x) \cap f(y)$. Notice that in the perfect community (i) $\forall x \in X{:}f(x) = \{\{1, 2, 3\}\}$, but also (ii) $f(\langle 1, 2\rangle) = f(\langle 1, 3\rangle) = f(\langle 2, 3\rangle) = \{\{1, 2, 3\}\}$. In the imperfect community, however, $f(1) = \{\{1, 2\}, \{1, 3\}\} \neq f(2) = \{\{1, 2\}, \{2, 3\}\} \neq f(3) = \{\{1, 3\}, \{2, 3\}\}$. But this means that $f(\langle 1, 2\rangle) = \{\{1, 2\}\} \neq \{\{2, 3\}\} = f(\langle 2, 3\rangle)$. In general, we can determine the similarity relation from a cover Q by defining the similarity as follows: $x \sim^Q y$ iff$_{df}$ $f^Q(x) \cap f^Q(y) \neq \emptyset$, where x can either be an element of X, or a pair, or a pair of pairs, etc. Then we would like it to be the case that for any cover Q of X, $\sim^Q$ as defined above faithfully represents the cover: we look at all maximal perfect communities that we get from $\sim^Q$ and see whether this is the same as Q. This is not yet the case, because we haven't solved yet the companionship difficulty.

Companionship difficulty. Let us start with cover $Q = \{\{1, 2\}, \{1, 2, 3\}\}$. The similarity structure $\langle X, \sim^Q\rangle$ that we derive now is $\langle\{1, 2, 3\}, \{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 1, 3\rangle, \langle\langle 1, 2\rangle, \langle 2, 3\rangle\rangle, \langle\langle 1, 2\rangle, \langle 1, 3\rangle\rangle, \langle\langle 1, 3\rangle, \langle 2, 3\rangle\rangle\}\rangle$. But the set of perfect communities derived from this is just $\{\{1, 2, 3\}\} \neq Q$. Thus, we have not solved the companionship difficulty yet. For the companionship difficulty, Rodriguez needs *degrees of resemblance*. He says that $sim(x, y) = n$ iff x and y share n properties. Thus, let us make a distinction between $\sim_1$ and $\sim_2$. In general, we say that $x \sim^Q_n y$ iff$_{df}$ $|f(x) \cap f(y)| \geq n$. Notice that $f(1) \cap f(2) = \{\{1, 2\}, \{1, 2, 3\}\}$, whereas $f(1) \cap f(3) = \{\{1, 2, 3\}\}$. Thus, whereas $1 \sim^Q_2 2$, it is not the case that $1 \sim^Q_2 3$, although $1 \sim^Q_1 3$. Now we represent cover Q by the following similarity structure $\langle X, \sim^Q_1, \sim^Q_2\rangle$: $\langle\{1, 2, 3\}, \{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 1, 3\rangle, \langle\langle 1, 2\rangle, \langle 2, 3\rangle\rangle, \langle\langle 1, 2\rangle, \langle 1, 3\rangle\rangle, \langle\langle 1, 3\rangle, \langle 2, 3\rangle\rangle\}, \{\langle 1, 2\rangle\}\rangle$. We determine the perfect communities w.r.t. each $\sim_n$. Notice that $\{1, 2\}$ and $\{1, 2, 3\}$ are perfect communities of level 1, whereas only $\{1, 2\}$ is a perfect community of level 2. Let us say that P is a basic property if there is an n such that (i) P is a perfect community of level n and (ii) there is no perfect community P' of level n such that $P \subset P'$. Thus, we now see that $\langle X, \sim^Q_1, \sim^Q_2\rangle$ gives rise to the following basic properties: $\{\{1, 2\}, \{1, 2, 3\}\}$. This is the same as Q, as desired.

[9] This is so, because in order for $f(\langle 1, 2\rangle) \cap f(\langle 2, 3\rangle) \neq \emptyset$ it must be that $\exists X : X \in f(1) \cap f(2) \cap f(3)$ such that $\{1, 2, 3\} \subseteq X$, see below.

Alternatively, define for each $x \in X$ the set $s_n^i(x)$ for each positive natural number i and level m as follows: $s_m^1(x) = \{\{x, y\}: y \in X \,\&\, x \sim_m y\}$; $s_m^{n+1}(x) = \{X \cup Y: X, Y \in s^n(x) \,\&\, X \sim_m Y\}$. $s_m(x)$ is now defined as the fixed point of this sequence. The set of properties Q of level m is now defined as $Q = \{s_m(x): x \in X\}$.

Of course, once we allow for degrees we basically take resemblance to be a 3- or even 4-place relation: '*b* is more similar to *b* than *c*' and '*a* is more similar to *b* as *c* is similar to *d*' become meaningful. Rodriguez-Pereyra's solution of the companionship difficulty is not the most interesting feature of his proposal. What is interesting about his proposal is how he solves the problem of imperfect communities. However, even if Rodriguez-Pereryra can solve both of these problems, that is still not enough to guarantee a 1–1 relation between sets of properties and similarity structures.

Mere intersection difficulty. Is it the case that every cover Q can be faithfully represented by a similarity structure? This depends on whether covers are closed under intersection yes or no. If yes (which I think is natural), we are ready. If no, we have a problem. Consider the following cover of $X = \{1, 2, 3, 4, 5, 6\}$: $Q = \{\{1, 2, 3\}, \{1, 2, 3, 4, 6\}, \{1, 2, 3, 4, 5\}\}$. Notice that from this Q we derive a similarity structure $\langle X, \sim_1^Q, \sim_2^Q, \sim_3^Q\rangle$ from which we derive the following set of properties: $\{\{1, 2, 3\}, \{1, 2, 3, 4, 6\}, \{1, 2, 3, 4, 5\}, \{1, 2, 3, 4\}\}$: $\{1, 2, 3\}$ is a maximal perfect community of level 3, $\{1, 2, 3, 4, 6\}$ and $\{1, 2, 3, 4, 5\}$ are maximal communities of level 1, and $\{1, 2, 3, 4\}$ is a maximal community of level 2. But this is not the same as Q, because $\{1, 2, 3, 4\}$ does not belong to Q. Notice that this set is the intersection of $\{1, 2, 3, 4, 6\}$ and $\{1, 2, 3, 4, 5\}$.

Similarity relation between sets Can't we simplify what Rodriguez Pereyra wanted to do? Why not simply say—as already suggested above—that we allow for similarity relations between *sets* of individuals? Thus, not only can there be similarity relations between different sets of exactly two elements, but there can be similarity relations between sets of all non-zero cardinality. In this way, we don't need to go any 'higher', i.e., assume similarity relations between pairs of pairs of .... pairs of individuals. So, here is the idea.[10] Suppose we start with a cover Q. Now we define the similarity relation between non-empty sets as follows: $p \sim p'$ iff$_{def}$ $\exists q \in Q: p, p' \subseteq q$.[11] Notice that this similarity relation is certainly reflexive and symmetric. Is it also transitive? No, for consider $Q = \{\{1, 2, 3\}, \{3, 4\}\}$. In this case we have, for instance, that $\{1\} \sim \{3\}$ and that $\{3\} \sim \{4\}$, but it is not the case that $\{1\} \sim \{4\}$. Thus, we can go from a cover to a similarity structure, although this similarity is now between sets, rather than between individuals. Of course, this new similarity relation would be isomorphic to a similarity relation between individuals, if we limited ourselves to singleton sets. But we did not, and thus our new similarity relation contains possibly more information.

[10] Only after writing this chapter we discovered Paseau (2012), where something very similar was worked out very precisely. Paseau argues that resemblance similarity can be saved, but that the cost of assuming similarity relations between *sets* of individuals is probably a too high price to pay for a nominalist.

[11] The empty set will be similar to no other set.

Is this extra information enough to solve the imperfect community problem? Let us look at our simplest example again: $Q = \{\{1, 2\}, \{1, 3\}, \{1, 3\}\}$. The similarity relation that this gives rise to contains (if we forget about the reflexive relations) only $\{1\} \sim \{2\}$, $\{1\} \sim \{3\}$, and $\{2\} \sim \{3\}$. This is different with the cover $Q' = \{\{1, 2, 3\}\}$, which gives rise to a similarity relation also connecting $\{1, 2\} \sim \{2, 3\}$, $\{1, 2\} \sim \{3\}$, and $\{1, 3\} \sim \{2, 3\}$, for instance. But whether we have solved the imperfect community problem depends on how we are now going to define (sparse, or natural) properties.

The proposal is very straightforward. First, we are going to define what it means to be a similarity* set. A similarity* set X is now not just a set that obeys constraint (i) all its elements (or better, singleton sets) are similar to each other, but also constraint (ii) for all non-empty subsets p, p' of $\bigcup X$ it holds that $p \sim p'$.[12] After this strengthening of the notion of a similarity set, we go on as before. First, we collect all maximal similarity* sets. Let us call MAX the set of all maximally similarity* sets. The properties induced by the similarity structure are then defined as $\{\bigcup X : X \in MAX\}$.

Let us first see whether the similarity relation induced by cover Q indeed avoids inducing imperfect community $\{1, 2, 3\}$ as a property. We have seen that the similarity relation Q gives rise to contains only $\{1\} \sim \{2\}$, $\{1\} \sim \{3\}$, and $\{2\} \sim \{3\}$. Notice that although $\{1, 2, 3\}$ is a similairty set, it is not a similarity* set, because it doesn't hold, for instance, that $\{1, 2\} \sim \{3\}$. With $\{1, 2, 3\}$ out of the way, we can see that now the maximal similarity* sets are $\{1, 2\}$, $\{1, 3\}$, and $\{2, 3\}$, just as desired. It is different for the similarity relation induced by $Q' = \{\{1, 2, 3\}\}$. Because all non-empty subsets p and p' of $\{1, 2, 3\}$ are by construction similar to each other, we end up with only one maximal similarity* set, X, such that $\bigcup X = \{1, 2, 3\}$.

What we would really like to know, however, is whether the set of properties $Q' = \{\bigcup X : X \in MAX\}$ induced from the similarity relation that is itself induced by cover Q is such that $Q' = Q$. To prove this, we should prove that $Q \subseteq Q'$ and that $Q' \subseteq Q$.

Let us start with an arbitrary element $q \in Q$. it holds by construction for non-empty subsets p and p' of q that $p \sim p'$. But this means that the set X of all non-empty subsets of q is a similarity* set. Will X also be a *maximal* similarity* set? If we forget about covers like $Q = \{\{1, 2\}, \{1, 2, 3\}\}$ that give rise to Goodman's companionship difficulty, it is clear that it is. But obviously $\bigcup X = q$, and thus $q \in Q'$.

To go the other way around, let us assume that $q' \in Q'$. This means that there is a maximal similarity* set X such that $q' = \bigcup X$. Form the fact that X is a similarity* set it follows that (i) $\forall p, p' \in X : p \sim p'$, and (ii) $\forall p, p' \subseteq \bigcup X : (p \neq \emptyset \wedge p' \neq \emptyset) \rightarrow p \sim p'$. Because if $p \in X$, it follows that $p \subseteq \bigcup X$, we can do with only condition (ii). But the similarity relation defined in terms of Q was exactly defined like (ii), with an element Q instead of $\bigcup X$. This means that $\bigcup X = q'$ is an element of Q.

[12] Because if $p \in X$, it follows that $p \subseteq \bigcup X$, we can do with only condition (ii). Notice that (ii) entails that all individuals in $\bigcup X$ resemble each other.

But a resemblance nominalist doesn't really start with a cover Q, but rather with a similarity relation. Can he really start out with a set of individuals S and then a arbitrary similarity relation between non-empty subsets of S? Suppose, for instance, that $S = \{1, 2, 3\}$, and that the similarity relation only connected $\{1, 2\} \sim \{3\}$ (ignoring reflexivity and symmetry as usual). It is clear that such a similarity relation does only give rise to singleton sets as similarity* sets. But what would it mean then that $\{1, 2\} \sim \{3\}$? It would have no meaning. We propose to get rid of such similarity relations by the following constraint: $\forall p, p'$: if $p \sim p'$, then $\forall r_{\neq\emptyset}, r'_{\neq\emptyset} \subseteq p \cup p'$:$r \sim r'$. Notice that this constraint is indeed obeyed in the above examples.

We believe that to solve the companionship difficulty staring from covers like $Q = \{\{1, 2\}, \{1, 2, 3\}\}$, we can use a trick very similar to what Rodriguez-Pereryra made use of: just assume that we we have more than 1 similarity relation, or better, perhaps, that we start with a three- or four-place comparative similarity relation: in this case, all non-empty subsets of $\{1, 2\}$ are at least as similar to each other as all non-empty subsets of $\{1, 2, 3\}$ are, and $\{1\}$ and $\{2\}$ are more similar to each other than $\{1\}$ and $\{3\}$ are, for instance. One way to go is to start with a four-place comparative similarity relation $q <_p r$ (q is more similar to p than r is), and define $<_p$ in terms of original cover Q as follows: if $\exists q, q' \in Q$:$q \subset q'$, then $\forall p, p'_{\neq\emptyset} \subseteq q$:$\forall r_{\neq\emptyset} \subseteq p$:$\forall s_{\neq\emptyset} \subseteq q' - q$:$p' <_p (r \cup s)$.

6 Conclusion

In this chapter we have shown, or reminded, the reader that it is in many cases less than obvious which notions should be taken as primitive and what should be constructed out of what. In many cases, both directions are possible. Some constructions give rise to technical difficulties, but these can in many cases be solved by assuming a somewhat richer ontology—as for instance in the previous section—or stronger constraints on the initial orderings—as in the case of events. Which direction the construction should go (if we want construction at all) depends on how 'natural' the primitives and constraints on the constructions one starts out with to get what one wants are taken to be. To think, for instance, of propositions as primitives, and to define worlds and similarities between them in terms of it, allows one to make more distinctions, but is somewhat harder to work with. The same holds for the view according to which properties are sets of features. But once one has decided on taking features to be basic, it is also natural to think of individuals in terms of (sets of sets of) features as well. It was not the purpose of this chapter to argue what the primitives should be, but just to point out certain options.

References

Allen, J. F., & Hayes, P. J. (1985). A common-sense theory of time. In A. Joshi (Ed.), *Proceedings of the the Ninth Joint Conference on Artificial Intelligence*, Los Altos.
Aristotle. (1941). Basic works. In R. McKeon (Ed.) New York: Random House.
Armstrong, D. M. (1989). *Universals: An opinionated introduction*. Oxford: Westview Press.
Arrow, K. (1959). Rational choice functions and orderings. *Economica*, *26*, 121–127.
Berge, C. (1989). *Hypergraphs*. Amsterdam: North Holland.
Carnap, R. (1923), Die Quasizerlegung. Ein Verfahren zur Ordnung nichthomogener Mengen mit den Mitteln der Beziehungslehre. Unpublished manuscript RC-081-04-01, University of Pittsburgh.
Carnap, R. (1928). *Der logische Aufbau der Welt*. Berlin: Weltkreis.
Carnap, R. (1947). *Meaning and necessity*. Chicago: University of Chicago.
Fishburn, P. C. (1970). Intransitive indifference with unequal indifference intervals. *Journal of Mathematical Psychology*, *7*, 144–149.
Gärdenfors, P. (2000). *Conceptual spaces*. Cambridge: MIT.
Glashoff, K. (to appear). An intensional Leibniz semantics for Aristotelian logic. *The Review of Symbolic Logic*.
Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects* (pp. 437–446). Indianapolis: The Bobbs-Merrill Company Inc.
Goodman, N. (1953). *The structure of appearance* (2nd ed.). Indianapolis: The Bobbs-Merrill Company Inc.
Ishiguro, H. (1972). *Leibniz' Philosophy of Logic and Language*. London: Duckworth.
Kamp, H. (1979). Events, instants and temporal reference. In R. Bäuerle et al. (Eds.), *Semantics form different points of view* (pp. 131–175). Berlin: Springer.
Klein, E. (1980). The semantics of positive and comparative adjectives. *Linguistics and Philosophy*, *4*, 1–45.
Kratzer, A. (1981). Partition and revision: The semantics of counterfactuals. *Journals of Philosophical Logic*, *10*, 242–258.
Leibniz, G. (1686). Discours de Métaphysique, (first published in English by Open Court, 1902).
Leibniz, G. (1966). Rules from which a decision can be made, by means of numbers, about the validity of inferences and about the forms and moods of categorical syllogisms. In G. H. R. Parkinson (ed.), Leibniz: Logical papers (pp. 25–32). Oxford: Clarendon Press.
Leitgeb, H. (2007). A new analysis of Quasianalysis. *Journal of Philosophical Logic*, *36*, 181–226.
Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
Lewis, D. (1983). New work on the theory of universals. *Australian Journal of Philosophy*, *64*, 85–88.
Lewis, D. (1986). *On the plurality of worlds*. Oxford: Blackwell.
Lück, U. (2006). Continuous time goes by Russell. *Notre Dame Journal of Formal Logic*, *47*, 397–434.
Mormann, T. (2009). New work for Carnap's quasi-analysis. *Journal of Philosophical Logic*, *38*, 249–282.
Paseau, A. (2012). Resemblance theories of properties. *Philosophical Studies*, *157*, 361–382.
Plato. (1941). *Republic* (F. Cornford, Trans). Oxford: Oxford University Press.
Quine, W. V. (1969). Natural kinds. In W. V. Quine (Ed.), *Ontological relativity and other essays* (pp. 114–138). New York: Columbia University Press.
Rescher, N. (1954). Leibniz's interpretation of his logical calculi. *Journal of Symbolic Logic*, *19*, 1–13.
Rodriguez-Pereyra, G. (1999). Resemblance nominalism and the imperfect community. *Philosophy and Phenomenological Research*, *59*, 965–982.
Rodriguez-Pereyra, G. (2002). *Resemblance Nominalism: A solution to the problem of universals*. Oxford: Clarendon Press.
Russell, B. (1912). The problems of philosophy. London: Home University Library.

Russell, B. (1914). *Our knowledge of the external world*. London: Routledge.
Russell, B. (1936). On order in time. *Proceedings of the Cambridge Philosophical Society*, *3*, 216–228.
Sen, A. K. (1971). Choice functions and revealed preference. *The Review of Economic Studies*, *38*, 307–317.
Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory*, American Philosophical Quarterly Monograph Series (pp. 98–112). Oxford: Blackwell.
Thomason, S. K. (1984). On constructing instants from events. *Journal of Philosophical Logic*, *13*, 85–96.
Turner, R. (1981). Counterfactuals without possible worlds. *Journals of Philosophical Logic*, *10*, 453–493.
van Benthem, J. (1982). Later than late: On the logical origin of the temporal order. *Pacific Philosophical Quarterly*, *63*, 193–203.
van Benthem, J. (1991). *The logic of time*. Synthese Library.
van Rooij, R. (2011). Semi-orders and satisficing behavior. *Synthese*, *179*, 1–12.
Veltman, F. (1976). Prejudices, presuppositions, and the theory of counterfactuals. In J. Groenendijk & M. Stokhof (Eds.), *Amsterdam papers in formal grammar. Proceedings of the 1st Amsterdam Colloquium* (pp. 248–281), Amsterdam.
Wiener, N. (1914). A contribution to the theory of relative position. *Proceedings of the Cambridge Philosophical Society*, *17*, 441–449.

Measurement-Theoretic Foundations of Dynamic Epistemic Preference Logic

Satoru Suzuki

Abstract In this chapter, we propose a new version of sound and complete dynamic epistemic preference logic (DEPL). Both preference logic and dynamic epistemic logic have gained considerable attention in linguistics, computer science and philosophy. Recently van Benthem and Liu proposed to integrate preference logic with dynamic epistemic logic. They called the resulting logic 'dynamic epistemic upgrade logic (DEUL)'. DEUL is designed only to deal with the dynamic interactions between knowledge and preferences originating from decision makings under certainty. On the other hand, DEPL is designed to deal with the dynamic interactions between knowledge and preferences originating from decision makings under certainty, risk, uncertainty and ignorance. So DEPL has much wider scope of application than DEUL. Providing DEPL with measurement-theoretic semantics enables it to have such wide scope.

Keywords Dynamic epistemic logic · Expected utility maximisation · Measurement theory · Preference logic · Representation theorem

1 Introduction

In this chapter, we propose a new version of sound and complete dynamic epistemic preference logic (DEPL). The notion of preference plays an important role in many disciplines, including philosophy and economics.[1] Some of notable recent developments in ethics make substantial use of preference logic.[2] In computer science, preference logic has become an indispensable device. Recently using Boutilier's idea

[1] Hansson (2006) conducts a comprehensive survey of preference in general

[2] For a comprehensive survey of preference logic, see Hansson (2001).

S. Suzuki (✉)
Faculty of Arts and Sciences, Komazawa University, 1-23-1, Komazawa,
Setagaya-ku, Tokyo154-8525, Japan
e-mail: bxs05253@nifty.com

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*,
Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_14,

(1994) that preferences between propositions can be defined in terms of two sorts of modalities, one of which is a universal modality, Van Benthem et al. (2005) reduced preference logic to modal logic. On the other hand, epistemic logic gets its start with the recognition that the expressions like ‘know that’ have systematic properties that are suitable for logical analysis. In addition to its relevance to traditional philosophical problems, epistemic logic has many applications in computer science and economics. Knowledge has not only static properties but also dynamic ones. ‘Dynamic epistemic logic’ is an umbrella term for a number of extensions of epistemic logics with dynamic operators that enables us to formalise reasoning information changes.[3] Dynamic epistemic logic has gained considerable attention in formal linguistics, computer science and philosophy. Recently Van Benthem and Liu (2007) proposed to integrate van Benthem et al.’s preference logic with dynamic epistemic logic. They called the resulting logic ‘dynamic epistemic upgrade logic (DEUL)’. DEUL enables us to reason logically about the dynamic interactions between knowledge and preferences. Decision problems can be classified into the following four types: decision making under

1. certainty,
2. risk,
3. uncertainty, and
4. ignorance.

DEUL is designed only to deal with the dynamic interactions between knowledge and preferences originating from decision makings under *certainty*. On the other hand, DEPL is designed to deal with the dynamic interactions between knowledge and preferences originating from decision makings under certainty, risk, uncertainty, and ignorance. So DEPL has much wider scope of application than DEUL. Providing DEPL with measurement-theoretic semantics enables it to have such wide scope. *The aim of this chapter is to propose a new version of sound and complete dynamic epistemic preference logic* (DEPL) *that deals with the dynamic interactions between knowledge and preferences originating from decision makings under certainty, risk, uncertainty, and ignorance, by using measurement theory*. Measurement theory is a theory that provides measurement with its mathematical foundation.[4] On the other hand, there are at least two kinds of decision theory:

1. evidential decision theory,[5] and
2. causal decision theory.[6]

The former is designed for decision makings that have statistical or evidential connections between actions and outcomes. The latter is designed for decision makings

[3] For a comprehensive survey of dynamic epistemic logic, consult Van Ditmarsch et al. (2007).

[4] The mathematical foundation of measurement had not been studied before Hölder (1901) developed his axiomatisation for the measurement of mass. Krantz (1971); Suppes (1989) and Luce (1990) are seen as milestones in the history of measurement theory. For a comprehensive survey of measurement theory, see Roberts (1979).

[5] For a comprehensive survey of evidential decision theory, see Jeffrey (1990).

[6] For a comprehensive survey of causal decision theory, see Joyce (1999).

that have causal connections between actions and outcomes. Both theories adopt conditional expected utility maximisation as a main decision rule. Jeffrey (1990) is a typical example of the former. Ramsey (1926) is a typical example of the latter. Ramsey regarded degree of desire as attitude toward consequences but degree of belief as *propositional attitude*. Moreover, he regarded preference as attitude toward an ordered pair of gambles, that is, hybrid entities composed of outcomes and propositions. Jeffrey (1990) developed an alternative to Ramsey's theory. He regarded both degree of desire and degree of belief as propositional attitudes. Moreover, he regarded preference as propositional attitude (attitude toward an ordered pair of propositions). In this sense we call Jeffrey's a *mono-set theory*. Its initial axiomatisation was provided in terms of measurement theory by Bolker (1967) on the mathematics developed in Bolker (1966). Jeffrey (1978) modified Bolker's axioms to accommodate null propositions. Domotor (1978) axiomatised a *finite* version of mono-set theory. Mono-set theories are more suitable for the semantics of logic than non-mono-set ones like Ramsey's, for regarding propositions as the semantic values of sentences is simpler than regarding gambles as those when we wish to provide logic with its semantics. Especially, Domotor's theory is the most suitable for the semantics of logic of these three mono-set theories, for constructing the syntactic analogues of the axioms of Domotor's theory is easier than of the other two theories. Like Bolker's and Jeffrey's, Domotor's theory has a conjoint structure. In them, preferences are decomposable into beliefs and desires. There are two fundamental problems in measurement theory:

1. the representation problem: justifying the assignment of numbers to objects or propositions,
2. the uniqueness problem: specifying the transformation up to which this assignment is unique.

A solution to the former can be furnished by a *representation theorem*, which establishes that the chosen numerical system preserves the relations of the relational system. Representation theorems of [conditional] expected utility maximisation have the following form:

> If [and only if] an agent's preferences satisfy such-and-such conditions, there exist a probability function and a utility function such that he should act as a [conditional] expected utility maximiser.

Among mono-set measurement theories, Domotor's representation theorem is the only known one of conditional expected utility maximisation that has the "only if" part. So only by virtue of Domotor's representation theorem, an observer can explain ascribing the logical properties to the agent's preferences originating from decision makings under certainty, risk, uncertainty or ignorance in terms of his beliefs and desires via expected utility maximisation.

The structure of this chapter is as follows. In Sect. 2, we prepare the projective-geometric concepts for the measurement-theoretic settings, define preference space and preference assignment, state necessary and sufficient conditions for representation: Connectedness and Projectivity, and show a Domotor-type representation theorem. In Sect. 3, we define the language $\mathcal{L}_{\mathsf{EPL}}$ of EPL, define a multi-agent

Domotor-type structured Kripke model $\mathfrak{M}$ for knowledge and preference, provide EPL with a truth definition, provide EPL with a proof system, prove the soundness of EPL in the usual way, and prove the completeness of EPL by filtration. In Sect. 4, we define the language $\mathcal{L}_{\mathsf{DEPL}}$ of DEPL, define the updated multi-agent Domotor-type structured Kripke model $\mathfrak{M}_{\varphi}$ for knowledge and preference, provide DEPL with a truth definition, provide DEPL with a proof system, and prove the soundness of DEPL in the usual way, provide a translation function, and prove the completeness of DEPL by means of it.

2 Measurement-Theoretic Settings

The point of this section is as follows:

- We would like to sate necessary and sufficient conditions for *Domotor's representation theorem*.
- We can state them in terms of *exterior product*, *symmetric product* and *four-fold exterior product*.
- They all can be defined in terms of *four-fold Cartesian product*.

2.1 Projective-Geometric Concepts

We need some projective-geometric concepts to state Domotor's representation theorem. We define the preliminary concepts to the measurement-theoretic settings as follows:

Definition 1 (*Preliminary Concepts*) $\mathbf{W}$ is a nonempty set of possible worlds. Let $\mathcal{F}$ denote a Boolean field of subsets of $\mathbf{W}$. We call $A \in \mathcal{F}$ a proposition.

We define a characteristic function as follows:

Definition 2 (*Characteristic Function I*) A characteristic function $\widehat{\ } : \mathcal{F} \to \{0, 1\}^{\mathbf{W}}$ is a function where for any $A \in \mathcal{F}$ we have $\widehat{A} : \mathbf{W} \to \{0, 1\}$ such that

$$\widehat{A}(w) := \begin{cases} 1 & \text{if } w \in A, \\ 0 & \text{otherwise,} \end{cases}$$

for any $w \in \mathbf{W}$.

Because it is impossible to characterise multiplication of probabilities and utilities in terms of union, intersection and preferences, we need a Cartesian product $\times.\widehat{\ }$ is defined also on Cartesian products of propositions:

Definition 3 (*Characteristic Function II*)

$$\widehat{(A \times B)}(w_1, w_2) := \begin{cases} 1 & \text{if } w_1 \in A \text{ and } w_2 \in B, \\ 0 & \text{otherwise,} \end{cases}$$

for any $w_1, w_2 \in \mathbf{W}$.

By means of $\times$, we define an exterior product $\widehat{A} \wedge \widehat{B}$ as follows:

Definition 4 (*Exterior Product*) $\widehat{A} \wedge \widehat{B}$ is a 3-valued random variable defined by

$$\widehat{A} \wedge \widehat{B} := (A \times B)\widehat{} - (B \times A)\widehat{},$$

where '$-$' denotes subtraction.

Remark 1 In short the exterior product is an anti-symmetric Cartesian product.

Remark 2 Intuitively, $\widehat{A} \wedge \widehat{B}$ can be measured by weighted utility difference $P(A)P(B)(U(B) - U(A))$.

Roughly speaking, projective geometry is represented in the language of *quadruples* of points. It suggests that we combine exterior products by means of a symmetric product $\odot$ as follows:

$$\begin{aligned}
&(\widehat{A} \wedge \widehat{B}) \odot (\widehat{C} \wedge \widehat{D}) \\
&:= (\widehat{A} \wedge \widehat{B}) \wedge (\widehat{C} \wedge \widehat{D}) + (\widehat{C} \wedge \widehat{D}) \wedge (\widehat{A} \wedge \widehat{B}) = \\
&(A \times B \times C \times D)\widehat{} + (B \times A \times D \times C)\widehat{} + (C \times D \times A \times B)\widehat{} + (D \times C \times B \times A)\widehat{} \\
&-(A \times B \times D \times C)\widehat{} - (B \times A \times C \times D)\widehat{} - (C \times D \times B \times A)\widehat{} - (D \times C \times A \times B)\widehat{},
\end{aligned}$$

where '$+$' denotes addition.

Remark 3 Intuitively, $(\widehat{A} \wedge \widehat{B}) \odot (\widehat{C} \wedge \widehat{D})$ can be measured by weighted products of utility differences $P(A)P(B)P(C)P(D)(U(B) - U(A))(U(D) - U(C))$. Symmetric products can contribute to describing multiplication in [conditional] expected utility theory in terms of measurement theory.

By means of symmetric products, we define a four-fold exterior product $\triangle(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{D})$ as follows:

Definition 5 (*Four-Fold Exterior Product*) $\triangle(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{D})$ is a 25-valued random variable defined by

$$\begin{aligned}
&\triangle(\widehat{A}, \widehat{B}, \widehat{C}, \widehat{D}) := \\
&(\widehat{A} \wedge \widehat{B}) \odot (\widehat{C} \wedge \widehat{D}) + (\widehat{A} \wedge \widehat{C}) \odot (\widehat{D} \wedge \widehat{B}) + (\widehat{A} \wedge \widehat{D}) \odot (\widehat{B} \wedge \widehat{C}) = \\
&(A \times B \times C \times D)\widehat{} + (B \times A \times D \times C)\widehat{} + (C \times D \times A \times B)\widehat{} + (D \times C \times B \times A)\widehat{} \\
&- (A \times B \times D \times C)\widehat{} - (B \times A \times C \times D)\widehat{} - (C \times D \times B \times A)\widehat{} - (D \times C \times A \times B)\widehat{} \\
&+ (A \times C \times D \times B)\widehat{} + (C \times A \times B \times D)\widehat{} + (D \times B \times A \times C)\widehat{} + (B \times D \times C \times A)\widehat{} \\
&- (A \times C \times B \times D)\widehat{} - (C \times A \times D \times B)\widehat{} - (D \times B \times C \times A)\widehat{} - (B \times D \times A \times C)\widehat{} \\
&+ (A \times D \times B \times C)\widehat{} + (D \times A \times C \times B)\widehat{} + (B \times C \times A \times D)\widehat{} + (C \times B \times D \times A)\widehat{} \\
&- (A \times D \times C \times B)\widehat{} - (D \times A \times B \times C)\widehat{} - (B \times C \times D \times A)\widehat{} - (C \times B \times A \times D)\widehat{}.
\end{aligned}$$

2.2 Preference Space and Preference Assignment

We define preference space and preference space assignment as follows:

Definition 6 (*Preference Space and Preference Space Assignment*)

- **A** is a finite set of agents.
- $W_{a,w} \subseteq \mathbf{W}$ should be interpreted to mean a set of worlds that $a \in \mathbf{A}$ takes into consideration at $w \in \mathbf{W}$.
- $\preceq_w$ is a weak preference relation on $\mathcal{F}_{a,w} \times \mathcal{F}_{a,w}$.
- $A \preceq_{a,w} B$ is interpreted to mean that a does not prefer A to B at w.
- $\sim_{a,w}$ and $\prec_{a,w}$ are defined as follows:
  - $A \sim_{a,w} B := A \preceq_{a,w} B$ and $B \preceq_{a,w} A$,
  - $A \prec_{a,w} B := A \preceq_{a,w} B$ and $A \not\succeq_{a,w} B$.
- For any $w \in \mathbf{W}$, $(W_{a,w}, \mathcal{F}_{a,w}, \preceq_{a,w}, \widehat{\ }, \times, +, -)$ is called a preference space.
- Let **PS** denote the set of all preference spaces.
- $\rho : \mathbf{W} \to \mathbf{PS}$ is called a preference space assignment.

2.3 Conditions for Representation

We can state necessary and sufficient conditions for representation as follows:

- $A \preceq_{a,w} B$ or $B \preceq_{a,w} A$ (**Connectedness**),
- If ($A_i \preceq_{a,w} B_i$ and $C_i \preceq_{a,w} D_i$ for any $i < n$),
  then (if $A_n \preceq_{a,w} B_n$, then $D_n \preceq_{a,w} C_n$),
  where $\sum_{i \leq n} (\widehat{A_i} \wedge \widehat{B_i}) \odot (\widehat{C_i} \wedge \widehat{D_i}) = \triangle(\widehat{A_n}, \widehat{B_n}, \widehat{C_n}, \widehat{D_n})$ (**Projectivity**).

Remark 4 Projectivity essentially says that given an equality

$$\sum_{i \leq n} P_{a,w}(A_i) P_{a,w}(B_i) P_{a,w}(C_i) P_{a,w}(D_i)(U_{a,w}(B_i) - U_{a,w}(A_i))(U_{a,w}(D_i) - U_{a,w}(C_i)) = 0,$$

the conditions $U_{a,w}(A_i) \leq U_{a,w}(B_i)$ with i between 1 and n and $U_{a,w}(C_i) \leq U_{a,w}(D_i)$ with i between 1 and $n-1$ necessitate $U_{a,w}(D_n) \leq U_{a,w}(C_n)$. Zero on the right-hand side comes from the fact that the measure of $\triangle(\widehat{A_n}, \widehat{B_n}, \widehat{C_n}, \widehat{D_n})$ happens to be equal to zero:

$$\begin{aligned} &P_{a,w}(A_n) P_{a,w}(B_n) P_{a,w}(C_n) P_{a,w}(D_n)((U_{a,w}(B_n) \\ &- U_{a,w}(A_n))(U_{a,w}(D_n) - U_{a,w}(C_n)) \\ &+ (U_{a,w}(C_n) - U_{a,w}(A_n))(U_{a,w}(B_n) - U_{a,w}(D_n)) \\ &+ (U_{a,w}(D_n) - U_{a,w}(A_n))(U_{a,w}(C_n) - U_{a,w}(B_n))) = 0. \end{aligned}$$

2.4 Domotor's Representation Theorem

Domotor proved the following representation theorem:

Theorem 1 (Representation, Domotor (1978)) *When* $\mathbf{W}$ *is finite, for any* $w \in \mathbf{W}$ *and any* $a \in \mathbf{A}$, $(W_{a,w}, \mathcal{F}_{a,w}, \preceq_{a,w}, \widehat{\ }, \times, +, -)$ *satisfies Connectedness and Projectivity iff there are* $P_{a,w} : \mathcal{F} \to \mathbb{R}$ *and* $U_{a,w} : \mathcal{F}_{a,w} \backslash \emptyset \to \mathbb{R}$ *such that the following conditions hold for any* $A, B \in \mathcal{F}_{a,w} \backslash \emptyset$:

- $(W_{a,w}, \mathcal{F}_{a,w}, P_{a,w})$ *is a finitely additive probability space,*
- $A \preceq_{a,w} B$ *iff* $U_{a,w}(A) \leq U_{a,w}(B)$,
- *If* $A \cap B = \emptyset$, $U_{a,w}(A \cup B) = P_{a,w}(A|A \cup B)U_{a,w}(A) + P_{a,w}(B|A \cup B)U_{a,w}(B)$,
- *When* $A \in \mathcal{F}_{a,w}$, *if* $P_{a,w}(A) = 0$, *then* $A = \emptyset$.

Remark 5 Domotor did not obtain the uniqueness result. But it does not matter when we provide EPL with its model.

Remark 6 Domotor's representation theorem follows from Scott's separation theorem.

Theorem 2 (Separation, Scott (1964)) *Let* I *be a finite-dimensional real linear vector space and let* $\emptyset \neq G \subset H \subset I$, *where* $H = -H = \{-v : v \in H\}$ *is finite and all its elements have rational coordinates with respect to a given basis. Then there exists a linear functional* $F : I \to \mathbb{R}$ *such that for any* $v \in H$

$$F(v) \geq 0 \textit{ iff } v \in G$$

iff for any $v, v_i \in H$ $(1 \leq i \leq n)$ *we have both*

(1) $v \in G or - v \in G$,

and

(2) *If* $v_i \in G$ *for any* $i < n$, *then* $- v_n \in G$, *where* $\sum_{i \leq n} v_i = 0$.

Remark 7 (1) corresponds to Connectedness and (2) corresponds to Projectivity. Scott's separation theorem is based on the general criterion for the solvability of a finite set of linear inequalities.

2.5 Significance of Domotor-Type Representation Theorem and Merit of DEPL

Based upon Luce and Raiffa (1957, p. 13) with a slight modification, decision problems can be classified into the following four types. We say that an agent is in the realm of decision making under:

1. *Certainty* if each leads to a specific outcome with the probability of 1 that is known to him,
2. *Risk* if each action leads to one of a set of possible specific outcomes each of which occurs with a probability that is known to him,
3. *Uncertainty* if each action leads to one of a set of possible specific outcomes, some of which occur with a probability that is known to him, but the other of which occur with a probability that is unknown to him, and
4. *Ignorance* if each action leads to one of a set of possible specific outcomes each of which occurs with a probability that is unknown to him.

When an observer considers an agent to be a decision maker under certainty, the dominating rule for decision making is *utility maximisation*. We provide an example of decision makings under certainty:

Example 1 (Decision Making under Certainty) The dinner guest is to provide the wine (white/red/rosé) when chicken is to be served. The guest's utility matrix might then be this:

	Chicken
White	1
Red	0
Rosé	0.5

The preference ordering of these acts then is this:

$$\text{Red} \prec \text{Rosé} \prec \text{White}$$

when $\prec$ is well-defined.

Cantor proved the following theorem about utility maximisation.

Theorem 3 (Representation and Uniqueness, Cantor (1895)) *Suppose* $\mathbf{W}$ *is a countable set and* $\preceq^*$ *is a binary relation on* $\mathbf{W}$. *Then* $(\mathbf{W}, \preceq^*)$ *is a weak order iff there is a function* $U^* : \mathbf{W} \to \mathbb{R}$ *satisfying*

$$w_1 \preceq^* w_2 \textit{ iff } U^*(w_1) \leq U^*(w_2).$$

Moreover, U^* *is unique up to a positive affine transformation.*

If an agent's desire state can be represented by a utility function, then by virtue of Theorem 3, an observer can explain ascribing the logical properties to the agent's preferences originating only from decision makings under certainty in terms of his desires via utility maximisation. On the other hand, when an observer considers an agent to be a decision maker under risk, the dominating rule for decision making is *[conditional] expected utility maximisation*. Jeffrey (1990, pp. 3–4, pp. 26–27) provides an example of decision makings under risk:

Example 2 (Decision Making under Risk) The dinner guest who is to provide the wine has forgotten whether chicken, beef or herring is to be served. He has no cellular phone, has a bottle of white, a bottle of red and a bottle of rosé, and can bring one of them in an oversized pocket since he is going by bicycle. The consequence matrix might well be the following:

	Chicken	Beef	Herring
White	White wine with chicken	White wine with beef	White wine with herring
Red	Red wine with chicken	Red wine with beef	Red wine with herring
Rosé	Rosé wine with chicken	Rosé wine with beef	Rosé wine with herring

The guest's utility matrix might then be this:

	Chicken	Beef	Herring
White	1	-1	1
Red	0	1	-1
Rosé	0.5	0	-1

If he takes the probabilities of chicken, beef, and herring to 0.4, 0.4, and 0.2 regardless of which act performs, his probability matrix will be as follows (left-hand columns):

	Chicken	Beef	Herring	
White	0.4	0.4	0.2	0.2
Red	0.4	0.4	0.2	0.2
Rosé	0.4	0.4	0.2	0

so that the expected utilities of bringing white, red, and rosé are as in the column at the right. Then with the given probability and utility matrices, the preference ordering of these acts is this:

$$\text{Rosé} \prec \text{White} \sim \text{Red}$$

when $\prec$ and $\sim$ are well-defined.

Next, we can provide an example of decision makings under uncertainty by slightly modifying Example 2:

Example 3 (Decision Making under Uncertainty) The dinner guest might not know some of the probabilities of chicken, beef, and herring. Then his probability matrix might be as follows:
so that the expected utilities of bringing white, red, and rosé are as in the column at the right.

	Chicken	Beef	Herring	
White	x_1	y_1	0.2	$x_1 - y_1 + 0.2$
Red	x_2	y_2	0.2	$y_2 - 0.2$
Rosé	x_3	y_3	0.2	$0.5x_3 - 0.2$

Finally, we can provide an example of decision makings under ignorance by slightly modifying Example 2:

Example 4 (Decision Making under Ignorance) The dinner guest might not know any of the probabilities of chicken, beef, and herring. Then his probability matrix might be as follows:

	Chicken	Beef	Herring	
White	x_1	y_1	z_1	$x_1 - y_1 + z_1$
Red	x_2	y_2	z_2	$y_2 - z_2$
Rosé	x_3	y_3	z_3	$0.5x_3 - z_3$

so that the expected utilities of bringing white, red, and rosé are as in the column at the right.

In both decision makings under uncertainty like Example 3 and decision makings under ignorance like Example 4, we cannot *generally* fix preference orderings of acts by calculating the expected utilities of them. So in this chapter, we follow another route. We pursue conditions of an agent's preferences necessary and sufficient for there existing a probability function and a utility function such that he should act as a [conditional] expected utility maximiser. In mono-set measurement theories, Domotor's representation theorem is the only known one that can furnish such necessary and sufficient conditions. All other representation theorems of [conditional] expected utility maximisation, such as Bolker (1967) and Jeffrey (1978), can furnish only sufficient conditions for it. Decision makings under certainty are degenerate cases of decision makings under risk where the probabilities are 0 or 1. Moreover, in Domotor's representation theorem, whether an agent knows the probabilities of outcomes or not, Connectedness and Projectivity of $\preceq_{a,w}$ can guarantee that there are probability function and utility function such that $\preceq_{a,w}$ respects the equality and inequality of conditional expected utility. So the preference relation that satisfies Connectedness and Projectivity can cover preferences originating from decision makings under certainty, risk, uncertainty and, ignorance. Therefore, if an agent's belief state can be represented by a probability function and his desire state can be represented by a utility function, then only by virtue of Domotor's representation theorem, an observer can explain ascribing the logical properties to the agent's preferences originating from decision makings under certainty, risk, uncertainty and, ignorance in terms of his beliefs and desires via conditional expected utility maximisation. Because the preference relations in the model of DEUL are mere *quasi-orders* (reflexive and transitive), DEUL cannot deal with the dynamic interactions between knowledge and preferences originating from decision makings under other circumstances than certainty. On the other hand, because the preference relations in the model of DEPL satisfy Connectedness and Projectivity, DEPL can deal with the dynamic interactions between knowledge and preferences originating from decision makings under certainty, risk, uncertainty and ignorance. So DEPL has much wider scope of application than DEUL.

3 Epistemic Preference Logic EPL

3.1 Language

The language of EPL $\mathcal{L}_{\mathsf{EPL}}$ is defined as follows:

Definition 7 (*Language*) Let **S** denote a set of sentential variables, **A** a finite set of agents, $\mathbf{K}_a$ an epistemic operator, $\mathbf{WPR}_a$ a weak preference relation symbol and **FCP** a four-fold Cartesian product symbol. $\mathcal{L}_{\mathsf{EPL}}$ is given by the following rule:

$$\varphi ::= s \mid \top \mid \neg\varphi \mid \varphi_1 \& \varphi_2 \mid \mathbf{K}_a(\varphi) \mid \mathbf{WPR}_a(\varphi_1, \varphi_2) \mid \mathbf{FCP}(\varphi_1, \varphi_2, \varphi_3, \varphi_4),$$

where $s \in \mathbf{S}$ and $a \in \mathbf{A}$.

- $\mathbf{WPR}(\varphi_1, \varphi_2)$ should be interpreted to mean that φ_2 is not preferred to φ_1.
- We define an indifference relation symbol **IND** and a strict value preference relation symbol **SPR** as follows:

$$\mathbf{IND}(\varphi_1, \varphi_2) := \mathbf{WPR}(\varphi_1, \varphi_2) \& \mathbf{WPR}(\varphi_2, \varphi_1),$$
$$\mathbf{SPR}(\varphi_1, \varphi_2) := \neg\mathbf{WPR}(\varphi_2, \varphi_1).$$

- The set of all well-formed formulae of $\mathcal{L}_{\mathsf{EPL}}$ is denoted by $\Phi_{\mathcal{L}_{\mathsf{EPL}}}$.

3.2 Semantics

DAG In order to state $\sum_{i \leq n} (\widehat{A_i} \wedge \widehat{B_i}) \odot (\widehat{C_i} \wedge \widehat{D_i}) = \triangle(\widehat{A_n}, \widehat{B_n}, \widehat{C_n}, \widehat{D_n})$ of Projectivity in logical terms, we use **FCP**. To provide **FCP** with a truth definition, we use a *directed acyclic graph (DAG)*. We got a hint about this idea from Naumov (2006). We define directedness as follows:

Definition 8 (*Directedness*) A graph G is directed if G consists of a nonempty set **W** of vertices (possible worlds) and an irreflexive accessibility relation R on **W**. G is denoted as $(\mathbf{W}, R)$.

We define a path as follows:

Definition 9 (*Path*) A sequence $[w_1, \ldots, w_{n+1}]$ of vertices is a path of length n in G from w_1 to w_{n+1} if $(w_i, w_{i+1}) \in R$ for $i = 1, \ldots, n$.

By means of a path, we define a cycle.

Definition 10 (*Cycle*) A cycle of length n is a path $[w_1, \ldots, w_n, w_1]$ from w_1 to w_1.

By means of a cycle, we define acyclicity as follows:

Definition 11 (*Acyclicity*) G is acyclic if G contains no cycles.

By means of directedness and acyclicity, we define a directed acyclic graph (DAG) as follows:

Definition 12 (*DAG*) G is a directed acyclic graph (DAG) if G is both directed and acyclic.

We define some concepts:

Definition 13 (*Parent, Child, Ancestor and Descendant*) w_1 is a parent of w_2 and w_2 is a child of w_1 if $(w_1, w_2) \in R$. w_1 is an ancestor of w_2 and w_2 is a descendant of w_1 if there is a path from w_1 to w_2.

Definition 14 (*Ancestral Ordering*) $[w_1, \ldots, w_n]$ is an ancestral ordering of the vertices in $\mathbf{W}$ if for each $1 \leq i \leq n$ all the ancestors of w_i are ordered before w_i.

DAGs have the following important property.

Proposition 1 (Ancestral Ordering and DAG) *There exists an ancestral ordering of the vertices in* $\mathbf{W}$ *iff* G *is a DAG.*

Model By developing the idea of Naumov (2006) and that of Halpern (2003), we define a multi-agent Domotor-type structured Kripke model $\mathfrak{M}$ for knowledge and preference as follows:

Definition 15 (*Model*)

- $\mathfrak{M}$ is a sextuple $(\mathbf{W}, R_{\mathbf{FCP}}, L, \{\approx_a\}_{a \in \mathbf{A}}, V, \rho)$, where
  - $\mathbf{W}$ is a nonempty set of possible worlds,
  - $R_{\mathbf{FCP}}$ is a relation of **FCP** on $\mathbf{W} \times \mathbf{W}$,
  - $(\mathbf{W}, R_{\mathbf{FCP}})$ is a directed acyclic graph (DAG),
  - $L : R_{\mathbf{FCP}} \to \{\pi_1, \pi_2, \pi_3, \pi_4\}$ is a function that assigns labels to the edges of the graph,
  - any two edges leaving the same vertex have different labels,
  - any vertex either has π_1-, π_2-, π_3- and π_4-labeled outgoing edges or none of them,
  - $\approx_a$ is an equivalence relation on $\mathbf{W} \times \mathbf{W}$,
  - V is a truth assignment to each $s \in \mathbf{S}$ for each $w \in \mathbf{W}$,

– ρ is a preference space assignment that assigns to each $a \in \mathbf{A}$ and each $w \in \mathbf{W}$ $(W_{a,w}, \mathcal{F}_{a,w}, \preceq_{a,w}, \widehat{\ }, \times, +, -)$ that satisfies Connectedness and Projectivity in which

$W_{a,w} := \{w' : w \approx_a w'\}$ (**Consistency**),

$\mathcal{F}_{a,w}$ is a Boolean algebra of subsets of $W_{a,w}$ with $\emptyset$ as zero element and $W_{a,w}$ as unit element,

for all $a \in \mathbf{A}$ and $w_1, w_2 \in \mathbf{W}$, if $w_1 \approx_a w_2$, then $\rho(a, w_1) = \rho(a, w_2)$ (**World-Dependent Preference**),

- For any $w_1 \in \mathbf{W}$, by $\pi_i(w_1)$ $(i = 1, 2, 3, 4)$ we mean the unique $w_2 \in \mathbf{W}$ such that $R_{\mathbf{FCP}}(w_1, w_2)$ and $L(w_1, w_2) = \pi_i$ if such world exists.

Remark 8 Some important aspects of the interactions between knowledge and preference can be caught by Consistency and World-Dependent Preference. Consistency postulates that an agent assigns preference only to worlds that he considers accessible (equivalent). World-Dependent Preference postulates that the choice of preference space is the same in all worlds the agent considers accessible (equivalent).

Truth We can provide EPL with the following truth definition:

Definition 16 (*Truth*) The notion of $\varphi \in \Phi_{\mathcal{L}_{\mathsf{EPL}}}$ being true at $w \in \mathbf{W}$ in $\mathfrak{M}$, in symbols $(\mathfrak{M}, w) \models_{\mathsf{EPL}} \varphi$, is inductively defined as follows:

- $(\mathfrak{M}, w) \models_{\mathsf{EPL}} s$ iff $V(w)(s) = \mathbf{true}$,
- $(\mathfrak{M}, w) \models_{\mathsf{EPL}} \top$,
- $(\mathfrak{M}, w) \models_{\mathsf{EPL}} \varphi_1 \& \varphi_2$ iff $(\mathfrak{M}, w) \models_{\mathsf{EPL}} \varphi_1$ and $(\mathfrak{M}, w) \models_{\mathsf{EPL}} \varphi_2$,
- $(\mathfrak{M}, w) \models_{\mathsf{EPL}} \neg\varphi$ iff $(\mathfrak{M}, w) \not\models_{\mathsf{EPL}} \varphi$,
- $(\mathfrak{M}, w_1) \models_{\mathsf{EPL}} \mathbf{K}_a(\varphi)$ iff $(\mathfrak{M}, w_2) \models_{\mathsf{EPL}} \varphi$ for all w_2 such that $w_1 \approx_a w_2$,
- $(\mathfrak{M}, w) \models_{\mathsf{EPL}} \mathbf{FCP}(\varphi_1, \varphi_2, \varphi_3, \varphi_4)$ iff $(\mathfrak{M}, \pi_1(w)) \models_{\mathsf{EPL}} \varphi_1$ and $(\mathfrak{M}, \pi_2(w)) \models_{\mathsf{EPL}} \varphi_2$ and $(\mathfrak{M}, \pi_3(w)) \models_{\mathsf{EPL}} \varphi_3$ and $(\mathfrak{M}, \pi_4(w)) \models_{\mathsf{EPL}} \varphi_4$,
- $(\mathfrak{M}, w_1) \models_{\mathsf{EPL}} \mathbf{WPR}_a(\varphi_1, \varphi_2)$ iff $[\![\varphi_1]\!]^{\mathfrak{M}}_{a,w_1} \preceq_{a,w_1} [\![\varphi_2]\!]^{\mathfrak{M}}_{a,w_1}$, where $[\![\varphi]\!]^{\mathfrak{M}}_{a,w_1} := \{w_2 \in \mathbf{W} : w_1 \approx_a w_2 \text{ and } (\mathfrak{M}, w_2) \models_{\mathsf{EPL}} \varphi\}$.

If $(\mathfrak{M}, w) \models_{\mathsf{EPL}} \varphi$ for all $w \in \mathbf{W}$, we write $\mathfrak{M} \models \varphi$ and say that φ is valid in $\mathfrak{M}$. If φ is valid in all multi-agent Domotor-type structured Kripke models for knowledge and preference, we write $\models_{\mathsf{EPL}} \varphi$ and say that φ is valid.

Remark 9 Later we will provide the truth condition of a syntactic counterpart of Projectivity by means of Definitions 17 and 18.

Significance of FCP FCP is a kind of modal operator. In $\mathfrak{M}$, we have assumed that each possible world w has its proper four $R_{\mathbf{FCP}}$-accessible worlds ($\pi_1(w)$, $\pi_2(w)$, $\pi_3(w)$ and $\pi_4(w)$) or none of them, where π_i is defined by $R_{\mathbf{FCP}}$ and L. We have given the truth condition of $\mathbf{FCP}(\varphi_1, \varphi_2, \varphi_3, \varphi_4)$ at w in $\mathfrak{M}$ in terms of the truth of φ_1 at $\pi_1(w)$ in $\mathfrak{M}$, the truth of φ_2 at $\pi_2(w)$ in $\mathfrak{M}$, the truth of φ_3 at $\pi_3(w)$ in $\mathfrak{M}$ and the truth of φ_4 at $\pi_4(w)$ in $\mathfrak{M}$. Because $(\mathbf{W}, R_{\mathbf{FCP}})$ is a DAG, Proposition 1 guarantees that there exists an **FCP**-ancestral ordering of the vertices in **W**.

3.3 Syntax

Syntactic Counterpart of Projectivity We devise a syntactic counterpart of Projectivity. By developing the idea of Segerberg (1971), we define a syntactic counterpart of Projectivity. Assume that

$$(3.3.1)\quad \sum_{i\leq n}(\widehat{A_i}\wedge \widehat{B_i})\odot(\widehat{C_i}\wedge \widehat{D_i}) = \triangle(\widehat{A_n}, \widehat{B_n}, \widehat{C_n}, \widehat{D_n}).$$

Then by means of Definition 5, we get

$$(3.3.2)\quad \sum_{i\leq n-1}(\widehat{A_i}\wedge \widehat{B_i})\odot(\widehat{C_i}\wedge \widehat{D_i}) - (\widehat{A_n}\wedge \widehat{C_n})\odot(\widehat{D_n}\wedge \widehat{B_n}) - (\widehat{A_n}\wedge \widehat{D_n})\odot(\widehat{B_n}\wedge \widehat{C_n}) = 0$$

Through transposition and simplification combined with the definition of $\odot$, we have

$$
\begin{aligned}
(3.3.3)\quad & (A_1\times B_1\times C_1\times D_1)\widehat{} + \cdots + (A_{n-1}\times B_{n-1}\times C_{n-1}\times D_{n-1})\widehat{} \\
& + (B_1\times A_1\times D_1\times C_1)\widehat{} + \cdots + (B_{n-1}\times A_{n-1}\times D_{n-1}\times C_{n-1})\widehat{} \\
& + (C_i\times D_1\times A_1\times B_1)\widehat{} + \cdots + (C_{n-1}\times D_{n-1}\times A_{n-1}\times B_{n-1})\widehat{} \\
& + (D_1\times C_1\times B_1\times A_1)\widehat{} + \cdots + (D_{n-1}\times C_{n-1}\times B_{n-1}\times A_{n-1})\widehat{} \\
& + (A_n\times C_n\times B_n\times D_n)\widehat{} + (C_n\times A_n\times D_n\times B_n)\widehat{} \\
& + (D_n\times B_n\times C_n\times A_n)\widehat{} + (B_n\times D_n\times A_n\times C_n)\widehat{} \\
& + (A_n\times D_n\times C_n\times B_n)\widehat{} + (D_n\times A_n\times B_n\times C_n)\widehat{} \\
& + (B_n\times C_n\times D_n\times A_n)\widehat{} + (C_n\times B_n\times A_n\times D_n)\widehat{} \\
& - (A_1\times B_1\times D_1\times C_1)\widehat{} - \cdots - (A_{n-1}\times B_{n-1}\times D_{n-1}\times C_{n-1})\widehat{} \\
& - (B_1\times A_1\times C_1\times D_1)\widehat{} - \cdots - (B_{n-1}\times A_{n-1}\times C_{n-1}\times D_{n-1})\widehat{} \\
& - (C_1\times D_1\times B_1\times A_1)\widehat{} - \cdots - (C_{n-1}\times D_{n-1}\times B_{n-1}\times A_{n-1})\widehat{} \\
& - (D_1\times C_1\times A_1\times B_1)\widehat{} - \cdots - (D_{n-1}\times C_{n-1}\times A_{n-1}\times B_{n-1})\widehat{} \\
& - (A_n\times C_n\times D_n\times B_n)\widehat{} - (C_n\times A_n\times B_n\times D_n)\widehat{} \\
& - (D_n\times B_n\times A_n\times C_n)\widehat{} - (B_n\times D_n\times C_n\times A_n)\widehat{} \\
& - (A_n\times D_n\times B_n\times C_n)\widehat{} - (D_n\times A_n\times C_n\times B_n)\widehat{} \\
& - (B_n\times C_n\times A_n\times D_n)\widehat{} - (C_n\times B_n\times D_n\times A_n)\widehat{} = 0.
\end{aligned}
$$

For example, we can consider $\mathbf{FCP}(\varphi_1, \psi_1, \chi_1, \tau_1)$ to be a syntactic counterpart of $(A_1\times B_1\times C_1\times D_1)\widehat{}$. So in terms of (3.3.3), we define $\mathbf{DC}_i$ (the disjunction of conjunctions of **FCP**s) that is the heart of a syntactic counterpart of Projectivity as follows:

Definition 17 (*Disjunction of Conjunctions of* **FCP***s*) For any i $(0\leq i\leq 4n+4)$, $\mathbf{DC}_i$ is defined as the disjunction of all the following conjunctions:

$$d_1\mathbf{FCP}(\varphi_1, \psi_1, \chi_1, \tau_1)\&\ldots\&d_{n-1}\mathbf{FCP}(\varphi_{n-1}, \psi_{n-1}, \chi_{n-1}, \tau_{n-1})$$

$$\&d_n\mathbf{FCP}(\psi_1, \varphi_1, \tau_1, \chi_1)\&\ldots\&d_{2n-2}\mathbf{FCP}(\psi_{n-1}, \varphi_{n-1}, \tau_{n-1}, \chi_{n-1})$$

$$\&d_{2n-1}\mathbf{FCP}(\chi_1, \tau_1, \varphi_1, \psi_1)\&\ldots\&d_{3n-3}\mathbf{FCP}(\chi_{n-1}, \tau_{n-1}, \varphi_{n-1}, \psi_{n-1})$$

$$\&d_{3n-2}\mathbf{FCP}(\tau_1, \chi_1, \psi_1, \varphi_1)\&\ldots\&d_{4n-4}\mathbf{FCP}(\tau_{n-1}, \chi_{n-1}, \psi_{n-1}, \varphi_{n-1})$$

$$\&d_{4n-3}\mathbf{FCP}(\varphi_n, \chi_n, \psi_n, \tau_n)\&d_{4n-2}\mathbf{FCP}(\chi_n, \varphi_n, \tau_n, \psi_n)$$

$$\&d_{4n-1}\mathbf{FCP}(\tau_n, \psi_n, \chi_n, \varphi_n)\&d_{4n}\mathbf{FCP}(\psi_n, \tau_n, \varphi_n, \chi_n)$$

$$\&d_{4n+1}\mathbf{FCP}(\varphi_n, \tau_n, \chi_n, \psi_n)\&d_{4n+2}\mathbf{FCP}(\tau_n, \varphi_n, \psi_n, \chi_n)$$

$$\&d_{4n+3}\mathbf{FCP}(\psi_n, \chi_n, \tau_n, \varphi_n)\&d_{4n+4}\mathbf{FCP}(\chi_n, \psi_n, \varphi_n, \tau_n)$$

$$\&e_1\mathbf{FCP}(\varphi_1, \psi_1, \tau_1, \chi_1)\&\ldots\&e_{n-1}\mathbf{FCP}(\varphi_{n-1}, \psi_{n-1}, \tau_{n-1}, \chi_{n-1})$$

$$\&e_n\mathbf{FCP}(\psi_1, \varphi_1, \chi_1, \tau_1)\&\ldots\&e_{2n-2}\mathbf{FCP}(\psi_{n-1}, \varphi_{n-1}, \chi_{n-1}, \tau_{n-1})$$

$$\&e_{2n-1}\mathbf{FCP}(\chi_1, \tau_1, \psi_1, \varphi_1)\&\ldots\&e_{3n-3}\mathbf{FCP}(\chi_{n-1}, \tau_{n-1}, \psi_{n-1}, \varphi_{n-1})$$

$$\&e_{3n-2}\mathbf{FCP}(\tau_1, \chi_1, \varphi_1, \psi_1)\&\ldots\&e_{4n-4}\mathbf{FCP}(\tau_{n-1}, \chi_{n-1}, \varphi_{n-1}, \psi_{n-1})$$

$$\&e_{4n-3}\mathbf{FCP}(\varphi_n, \chi_n, \tau_n, \psi_n)\&e_{4n-2}\mathbf{FCP}(\chi_n, \varphi_n, \psi_n, \tau_n)$$

$$\&e_{4n-1}\mathbf{FCP}(\tau_n, \psi_n, \varphi_n, \chi_n)\&e_{4n}\mathbf{FCP}(\psi_n, \tau_n, \chi_n, \varphi_n)$$

$$\&e_{4n+1}\mathbf{FCP}(\varphi_n, \tau_n, \psi_n, \chi_n)\&e_{4n+2}\mathbf{FCP}(\tau_n, \varphi_n, \chi_n, \psi_n)$$

$$\&e_{4n+3}\mathbf{FCP}(\psi_n, \chi_n, \varphi_n, \tau_n)\&e_{4n+4}\mathbf{FCP}(\chi_n, \psi_n, \tau_n, \varphi_n)$$

such that exactly i of the d_j's and i of the e_j's are the empty string of symbols, the rest of them being the negation symbols.

By means of $\mathbf{DC}_i$, we define $\mathbf{DDC}^n_{i=1}(\varphi_i, \psi_i, \chi_i, \tau_i)$ (the ḋisjunction of ḋisjunctions of ċonjunctions of **FCP**s) that is a syntactic counterpart of Projectivity as follows:

Definition 18 (*Disjunction of Disjunctions of Conjunctions of* **FCP***s*)

$$\mathbf{DDC}^n_{i=1}(\varphi_i, \psi_i, \chi_i, \tau_i) := \mathbf{DC}_1 \vee \ldots \vee \mathbf{DC}_{4n+4}.$$

By means of Definitions 17 and 18, we can provide $\mathbf{DDC}^n_{i=1}(\varphi_i, \psi_i, \chi_i, \tau_i)$ with the following truth condition:

Proposition 2 (Truth Condition of $\mathrm{DDC}^n_{i=1}(\varphi_i, \psi_i, \chi_i, \tau_i)$)

$$(\mathfrak{M}, w) \models_{\mathsf{EPL}} \mathbf{DDC}^n_{i=1}(\varphi_i, \psi_i, \chi_i, \tau_i)$$

iff

$(((\mathfrak{M}, \pi_1(w)) \models_{\mathsf{EPL}} \varphi_1$ and $(\mathfrak{M}, \pi_2(w)) \models_{\mathsf{EPL}} \psi_1$ and $(\mathfrak{M}, \pi_3(w)) \models_{\mathsf{EPL}} \chi_1$ and $(\mathfrak{M}, \pi_4(w)) \models_{\mathsf{EPL}} \tau_1)$ and…and
$((\mathfrak{M}, \pi_1(w)) \models_{\mathsf{EPL}} \chi_n$ and $(\mathfrak{M}, \pi_2(w)) \models_{\mathsf{EPL}} \psi_n$ and $(\mathfrak{M}, \pi_3(w)) \models_{\mathsf{EPL}} \tau_n$ and $(\mathfrak{M}, \pi_4(w)) \models_{\mathsf{EPL}} \varphi_n))$

or…or

$(((\mathfrak{M}, \pi_1(w)) \not\models_{\mathsf{EPL}} \varphi_1$ or $(\mathfrak{M}, \pi_2(w)) \not\models_{\mathsf{EPL}} \psi_1$ or $(\mathfrak{M}, \pi_3(w)) \not\models_{\mathsf{EPL}} \chi_1$ or $(\mathfrak{M}, \pi_4(w)) \not\models_{\mathsf{EPL}} \tau_1)$ and…and
$((\mathfrak{M}, \pi_1(w)) \not\models_{\mathsf{EPL}} \chi_n$ or $(\mathfrak{M}, \pi_2(w)) \not\models_{\mathsf{EPL}} \psi_n$ or $(\mathfrak{M}, \pi_3(w)) \not\models_{\mathsf{EPL}} \tau_n$ or $(\mathfrak{M}, \pi_4(w)) \not\models_{\mathsf{EPL}} \varphi_n))$.

Proof System We provide EPL with the following proof system.

Definition 19 (*Proof System*)

- **Axioms of** EPL

(A1) All tautologies of classical sentential logic,

(A2) $\mathbf{WPR}_a(\varphi_1, \varphi_2) \vee \mathbf{WPR}_a(\varphi_2, \varphi_1)$ (**Syntactic Analogue of Connectedness**),

(A3) $\mathbf{DDC}^n_{i=1}(\varphi_i, \psi_i, \chi_i, \tau_i) \to ((\mathbf{WPR}_a(\varphi_1, \psi_1) \,\&\, \mathbf{WPR}_a(\chi_1, \tau_1) \,\&\, \ldots \,\&\, \mathbf{WPR}_a(\varphi_{n-1}, \psi_{n-1}) \,\&\, \mathbf{WPR}_a(\chi_{n-1}, \tau_{n-1})) \to (\mathbf{WPR}_a(\varphi_n, \psi_n) \to \mathbf{WPR}_a(\tau_n, \chi_n)))$
(**Syntactic Counterpart of Projectivity**),

(A4) $\mathbf{FCP}(\top, \top, \top, \top)$ (**Tautology and Four-Fold Cartesian Product**),

(A5) $\mathbf{FCP}(\varphi_1 \,\&\, \varphi_2, \psi_1 \,\&\, \psi_2, \chi_1 \,\&\, \chi_2, \tau_1 \,\&\, \tau_2) \to (\mathbf{FCP}(\varphi_1, \psi_1, \chi_1, \tau_1) \,\&\, \mathbf{FCP}(\varphi_2, \psi_2, \chi_2, \tau_2))$(**Conjunction and Four-Fold Cartesian Product 1**),

(A6) $(\mathbf{FCP}(\varphi_1, \mu, \nu, \xi) \,\&\, \mathbf{FCP}(\varphi_2, \mu, \nu, \xi)) \to \mathbf{FCP}(\varphi_1 \,\&\, \varphi_2, \mu, \nu, \xi)$ (**Conjunction and Four-Fold Cartesian Product 2**),

(A7) $(\mathbf{FCP}(\lambda, \psi_1, \nu, \xi) \,\&\, \mathbf{FCP}(\lambda, \psi_2, \nu, \xi)) \to \mathbf{FCP}(\lambda, \psi_1 \,\&\, \psi_2, \nu, \xi)$ (**Conjunction and Four-Fold Cartesian Product 3**),

(A8) $(\mathbf{FCP}(\lambda, \mu, \chi_1, \xi) \,\&\, \mathbf{FCP}(\lambda, \mu, \chi_2, \xi)) \to \mathbf{FCP}(\lambda, \mu, \chi_1 \,\&\, \chi_2, \xi)$ (**Conjunction and Four-Fold Cartesian Product 4**),

(A9) $(\mathbf{FCP}(\lambda, \mu, \nu, \tau_1) \,\&\, \mathbf{FCP}(\lambda, \mu, \nu, \tau_2)) \to \mathbf{FCP}(\lambda, \mu, \nu, \tau_1 \,\&\, \tau_2)$ (**Conjunction and Four-Fold Cartesian Product 5**),

(A10) $\neg\mathbf{FCP}(\varphi, \psi, \chi, \tau) \leftrightarrow (\mathbf{FCP}(\neg\varphi, \psi, \chi, \tau) \vee \mathbf{FCP}(\varphi, \neg\psi, \chi, \tau) \vee \mathbf{FCP}(\varphi, \psi, \neg\chi, \tau) \vee \mathbf{FCP}(\varphi, \psi, \chi, \neg\tau))$(**Negation and Four-Fold Cartesian Product**),

(A11) $\mathbf{K}_a(\varphi_1 \to \varphi_2) \to (\mathbf{K}_a(\varphi_1) \to \mathbf{K}_a(\varphi_2))$ (**K**),

(A12) $\mathbf{K}_a(\varphi) \to \varphi$ (**T**),

(A13) $\mathbf{K}_a(\varphi) \to \mathbf{K}_a\mathbf{K}_a(\varphi)$ (**Positive Introspection**),

(A14) $\neg\mathbf{K}_a(\varphi) \to \mathbf{K}_a\neg\mathbf{K}_a(\varphi)$ (**Negative Introspection**), (A15) $\mathbf{K}_a(\varphi) \leftrightarrow \mathbf{IND}_a(\varphi, \top)$ (**Syntactic Analogue of Consistency**),

(A16) $\mathbf{WPR}_a(\varphi_1, \varphi_2) \rightarrow \mathbf{K}_a(\mathbf{WPR}_a(\varphi_1, \varphi_2))$
(**Syntactic Analogue of World-Dependent Preference**).

• Inference Rules of EPL

(R1) $\dfrac{\varphi_1 \quad \varphi_1 \rightarrow \varphi_2}{\varphi_2}$ (**Modus Ponens**),

(R2) $\dfrac{\varphi \,\&\, \psi \,\&\, \chi \,\&\, \tau}{\mathbf{FCP}(\varphi, \psi, \chi, \tau)}$ (**Four-Fold Cartesian Product Necessitation**),

(R3) $\dfrac{\varphi}{\mathbf{K}_a(\varphi)}$ (**Knowledge Necessitation**),

(R4) $\dfrac{\varphi \leftrightarrow \psi \quad \chi' \text{ is like } \chi \text{ except for containing } \psi \text{ in some place where } \chi \text{ has } \varphi}{\chi \leftrightarrow \chi'}$
(**Replacement**).

A proof of $\varphi \in \Phi_{\mathcal{L}_{\mathsf{EPL}}}$ is a finite sequence of $\mathcal{L}_{\mathsf{EPL}}$-formulae having φ as the last formula such that either each formula is an instance of an axiom or it can be obtained from formulae that appear earlier in the sequence by applying an inference rule. If there is a proof of φ, we write $\vdash_{\mathsf{EPL}} \varphi$.

3.4 Metalogic

Soundness of EPL We can prove the soundness of EPL.

Theorem 4 (Soundness) *For any $\varphi \in \Phi_{\mathcal{L}_{\mathsf{EPL}}}$, if $\vdash_{\mathsf{EPL}} \varphi$, then $\models_{\mathsf{EPL}} \varphi$.*

Proof We omit the inductive proof. The only difficulty of which is to show that all of instances of (A3) are true in all models. This difficulty is removed as follows. Let $w \in \mathbf{W}$ be in $\mathfrak{M}$, and assume that for $\varphi_1, \ldots, \varphi_n, \psi_1, \ldots, \psi_n, \chi_1, \ldots, \chi_n, \tau_1, \ldots, \tau_n \in \Phi_{\mathcal{L}_{\mathsf{EPL}}}$,

(3.4.1) $(\mathfrak{M}, w) \models_{\mathsf{EPL}} \mathbf{DDC}^n_{i=1}(\varphi_i, \psi_i, \chi_i, \tau_i)$,

(3.4.2) $(\mathfrak{M}, w) \models_{\mathsf{EPL}} \mathbf{WPR}_a(\varphi_1, \psi_1) \& \mathbf{WPR}_a(\chi_1, \tau_1)$
$\& \ldots \& \mathbf{WPR}_a(\varphi_{n-1}, \psi_{n-1}) \& \mathbf{WPR}_a(\chi_{n-1}, \tau_{n-1})$,

(3.4.3) $(\mathfrak{M}, w) \models_{\mathsf{EPL}} \mathbf{WPR}_a(\varphi_n, \psi_n)$.

By means of Definitions 17 and 18 combined with (3.3.1), (3.3.2), (3.3.3) and (3.4.1), we have

$$\sum_{i\le n}((\llbracket\varphi_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\psi_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}) \odot ((\llbracket\chi_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\tau_i\rrbracket^{\mathfrak{M}})\widehat{})$$
$$= \triangle((\llbracket\varphi_n\rrbracket^{\mathfrak{M}})\widehat{}, (\llbracket\psi_n\rrbracket^{\mathfrak{M}})\widehat{}, (\llbracket\chi_n\rrbracket^{\mathfrak{M}})\widehat{}, (\llbracket\tau_n\rrbracket^{\mathfrak{M}})\widehat{}).$$

We define an additive measure as follows:

$$M_{a,w}(A) := P_{a,w}(A)U_{a,w}(A),$$

for any $A \in \mathcal{F}$. $M_{a,w}(A)$ can be converted with $P_{a,w}$ into unique liner functionals on canonical vector space $\mathbf{V}(\mathbf{W})$. Their exterior product $P_{a,w} \wedge M_{a,w} : \mathbf{V} \times \mathbf{V}$ is applicable to the equation involving characteristic functions of propositions. By using $P_{a,w} \wedge M_{a,w}$ we obtain

$$\sum_{i\le n}((\llbracket\varphi_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\psi_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}) \odot ((\llbracket\chi_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\tau_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{})$$
$$= \sum_{i\le n} P_w \wedge M_w((\llbracket\varphi_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\psi_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}) P_w \wedge M_w((\llbracket\chi_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\tau_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}).$$

By substituting $U_{a,w}$ for $M_{a,w}$ via $U_{a,w}(A) = \frac{M_{a,w}(A)}{P_{a,w}(A)}$, we get

$$\sum_{i\le n} P_{a,w} \wedge M_{a,w}((\llbracket\varphi_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\psi_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}) P_{a,w} \wedge M_{a,w}((\llbracket\chi_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\tau_i\rrbracket^{\mathfrak{M}}_{a,w})\widehat{})$$
$$= \sum_{i\le n} P_{a,w}(\llbracket\varphi_i\rrbracket^{\mathfrak{M}}_{a,w}) P_{a,w}(\llbracket\psi_i\rrbracket^{\mathfrak{M}}_{a,w}) P_{a,w}(\llbracket\chi_i\rrbracket^{\mathfrak{M}}_{a,w}) P_{a,w}(\llbracket\tau_i\rrbracket^{\mathfrak{M}}_{a,w})$$
$$\times (U_{a,w}(\llbracket\psi_i\rrbracket^{\mathfrak{M}}_{a,w}) - U_{a,w}(\llbracket\varphi_i\rrbracket^{\mathfrak{M}}_{a,w}))(U_{a,w}(\llbracket\tau_i\rrbracket^{\mathfrak{M}}_{a,w}) - U_{a,w}(\llbracket\chi_i\rrbracket^{\mathfrak{M}}_{a,w}))$$

In the same way we get

$$\triangle((\llbracket\varphi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}, (\llbracket\psi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}, (\llbracket\chi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}, (\llbracket\tau_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{})$$
$$= ((\llbracket\varphi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\psi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}) \odot ((\llbracket\chi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\tau_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{})$$
$$+ ((\llbracket\varphi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\chi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}) \odot ((\llbracket\tau_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\psi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{})$$
$$+ ((\llbracket\varphi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\tau_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}) \odot ((\llbracket\psi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\chi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{})$$
$$= P_{a,w} \wedge M_{a,w}((\llbracket\varphi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\psi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}) P_{a,w} \wedge M_{a,w}((\llbracket\chi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\tau_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{})$$
$$+ P_{a,w} \wedge M_{a,w}((\llbracket\varphi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\chi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}) P_{a,w} \wedge M_{a,w}((\llbracket\tau_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\psi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{})$$
$$+ P_{a,w} \wedge M_{a,w}((\llbracket\varphi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\tau_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{}) P_{a,w} \wedge M_{a,w}((\llbracket\psi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{} \wedge (\llbracket\chi_n\rrbracket^{\mathfrak{M}}_{a,w})\widehat{})$$
$$= P_{a,w}(\llbracket\varphi_n\rrbracket^{\mathfrak{M}}_{a,w}) P_{a,w}(\llbracket\psi_n\rrbracket^{\mathfrak{M}}_{a,w}) P_{a,w}(\llbracket\chi_n\rrbracket^{\mathfrak{M}}_{a,w}) P_{a,w}(\llbracket\tau_n\rrbracket^{\mathfrak{M}}_{a,w})$$
$$\times ((U_{a,w}(\llbracket\psi_n\rrbracket^{\mathfrak{M}}_{a,w}) - U_{a,w}(\llbracket\varphi_n\rrbracket^{\mathfrak{M}}_{a,w}))(U_{a,w}(\llbracket\tau_n\rrbracket^{\mathfrak{M}}_{a,w}) - U_{a,w}(\llbracket\chi_n\rrbracket^{\mathfrak{M}}_{a,w}))$$
$$+ (U_{a,w}(\llbracket\chi_n\rrbracket^{\mathfrak{M}}_{a,w}) - U_{a,w}(\llbracket\varphi_n\rrbracket^{\mathfrak{M}}_{a,w}))(U_{a,w}(\llbracket\psi_n\rrbracket^{\mathfrak{M}}_{a,w}) - U_{a,w}(\llbracket\tau_n\rrbracket^{\mathfrak{M}}_{a,w}))$$
$$+ (U_{a,w}(\llbracket\tau_n\rrbracket^{\mathfrak{M}}_{a,w}) - U_{a,w}(\llbracket\varphi_n\rrbracket^{\mathfrak{M}}_{a,w}))(U_{a,w}(\llbracket\chi_n\rrbracket^{\mathfrak{M}}_{a,w}) - U_{a,w}(\llbracket\psi_n\rrbracket^{\mathfrak{M}}_{a,w}))) = 0$$

So we have

$$(3.4.4)\quad \sum_{i\le n} P_{a,w}([\![\varphi_i]\!]^{\mathfrak{M}}_{a,w})P_{a,w}([\![\psi_i]\!]^{\mathfrak{M}}_{a,w})P_{a,w}([\![\chi_i]\!]^{\mathfrak{M}}_{a,w})P_{a,w}([\![\tau_i]\!]^{\mathfrak{M}}_{a,w})$$
$$\times(U_{a,w}([\![\psi_i]\!]^{\mathfrak{M}}_{a,w})-U_{a,w}([\![\varphi_i]\!]^{\mathfrak{M}}_{a,w}))(U_{a,w}([\![\tau_i]\!]^{\mathfrak{M}}_{a,w})-U_{a,w}([\![\chi_i]\!]^{\mathfrak{M}}_{a,w}))=0.$$

Because of (3.4.2) and (3.4.3) we get

$$[\![\varphi_1]\!]^{\mathfrak{M}}_{a,w}\preceq_{a,w}[\![\psi_1]\!]^{\mathfrak{M}}_{a,w},[\![\chi_1]\!]^{\mathfrak{M}}_{a,w}\preceq_{a,w}[\![\tau_1]\!]^{\mathfrak{M}}_{a,w},\dots,$$
$$[\![\varphi_{n-1}]\!]^{\mathfrak{M}}_{a,w}\preceq_{a,w}[\![\psi_{n-1}]\!]^{\mathfrak{M}}_{a,w},[\![\chi_{n-1}]\!]^{\mathfrak{M}}_{a,w}\preceq_{a,w}[\![\tau_{n-1}]\!]^{\mathfrak{M}}_{a,w},[\![\varphi_n]\!]^{\mathfrak{M}}_{a,w}\preceq_{a,w}[\![\psi_n]\!]^{\mathfrak{M}}_{a,w}.$$

So by virtue of Theorem 1, we have

$$(3.4.5)\quad U_{a,w}([\![\varphi_1]\!]^{\mathfrak{M}}_{a,w})\le U_{a,w}([\![\psi_1]\!]^{\mathfrak{M}}_{a,w}),U_w([\![\chi_1]\!]^{\mathfrak{M}}_{a,w})\le U_{a,w}([\![\tau_1]\!]^{\mathfrak{M}}_{a,w}),\dots,$$
$$U_w([\![\varphi_{n-1}]\!]^{\mathfrak{M}}_{a,w})\le U_{a,w}([\![\psi_{n-1}]\!]^{\mathfrak{M}}_{a,w}),U_w([\![\chi_{n-1}]\!]^{\mathfrak{M}}_{a,w})\le U_{a,w}([\![\tau_{n-1}]\!]^{\mathfrak{M}}_{a,w}),$$
$$U_w([\![\varphi_n]\!]^{\mathfrak{M}}_{a,w})\le U_{a,w}([\![\psi_n]\!]^{\mathfrak{M}}_{a,w}).$$

(3.4.4) and (3.4.5) force

$$U_{a,w}([\![\tau_n]\!]^{\mathfrak{M}}_{a,w})\le U_{a,w}([\![\chi_n]\!]^{\mathfrak{M}}_{a,w})$$

Then by virtue of Theorem 1, we get

$$[\![\tau_n]\!]^{\mathfrak{M}}_{a,w}\preceq_{a,w}[\![\chi_n]\!]^{\mathfrak{M}}_{a,w}.$$

Hence, we have

$$(\mathfrak{M},w)\models_{\mathsf{EPL}}\mathbf{WPR}_a(\tau_n,\chi_n),$$

which is what we wanted to establish.

Completeness of EPL We now turn to the task of proving the completeness of EPL. We prove it by using the ideas of Segerberg (1971) and modifying *filtration* in such a way that completeness can be established by *Domotor's representation theorem*. We cannot go into detail because of limited space, but the outline of the proof is as follows. We begin by defining some new concepts.

Definition 20 (*Stuffedness*) Suppose that Θ is a set of formulae such that Θ is closed under subformulae. Let

$$\Delta := \{\varphi : \text{for some}\psi, \mathbf{WPR}_a(\varphi,\psi)\in\Theta\text{or}\mathbf{WPR}_a(\psi,\varphi)\in\Theta\},$$

and let Δ' be the closure of Δ under Boolean compounds. If Θ also satisfies the condition that $\mathbf{WPR}_a(\varphi,\psi)\in\Theta$, for any $\varphi,\psi\in\Delta'$, we say that Θ is stuffed.

Definition 21 (*Value Formula*) The formulae in Δ' are called the value formulae of Θ.

Remark 10 There is no occurrence of $\mathbf{WPR}_a$ in value formulae.

Definition 22 (*Base*) We say that $\Psi_0 \subseteq \Phi_{\mathcal{L}_{\mathsf{EPL}}}$ is a base (with respect to EPL) for $\Psi \subseteq \Phi_{\mathcal{L}_{\mathsf{EPL}}}$ if for any $\varphi \in \Psi$ there is some $\varphi_0 \in \Psi_0$ such that $\vdash_{\mathsf{EPL}} \varphi \leftrightarrow \varphi_0$.

Definition 23 (*Logical Finiteness*) We say that Ψ is logically finite (with respect to EPL) if there is a finite base for Ψ.

Then we can prove the next lemma:

Lemma 1 (Logical Finiteness) *If* $\Psi \subseteq \Phi_{\mathcal{L}_{\mathsf{EPL}}}$ *is a finite set closed under subformulae, and if* Θ *is the smallest stuffed superset of* Ψ*, then* Θ *is logically finite.*

We define EPL-maximal consistency as follows:

Definition 24 (EPL-*Maximal Consistency*) A finite set $\{\varphi_1, \ldots, \varphi_n\} \subseteq \Phi_{\mathcal{L}_{\mathsf{EPL}}}$ is EPL-consistent iff $\nvdash_{\mathsf{EPL}} \neg(\varphi_1 \& \ldots \& \varphi_n)$. An infinite set of formulae is EPL-consistent iff all of its finite subsets are EPL-consistent. $\Gamma \subseteq \Phi_{\mathcal{L}_{\mathsf{EPL}}}$ is a EPL-maximal consistent set iff it is EPL-consistent and for any $\varphi \notin \Gamma$, $\Gamma \cup \{\varphi\}$ is EPL-inconsistent.

A canonical model for the modal part of EPL is defined as follows:

Definition 25 (*Canonical Model for Modal Part*) We define $\mathfrak{U}^C := (\mathbf{X}^C, R^C_{\mathbf{FCP}}, L^C, \{\approx^C_a\}_{a \in \mathbf{A}}, V^C)$ as a canonical model for the modal part of EPL in which

- $\mathbf{X}^C := \{\Gamma \subseteq \Phi_{\mathcal{L}_{\mathsf{EPL}}} : \Gamma \text{is} \mathsf{EPL}\text{-maximal consistent}\}$,
- for any $\Gamma, \Delta_1, \Delta_2, \Delta_3, \Delta_4 \in \mathbf{W}^C$, $R^C_{\mathbf{FCP}}(\Gamma, \Delta_1)$, $R^C_{\mathbf{FCP}}(\Gamma, \Delta_2)$, $R^C_{\mathbf{FCP}}(\Gamma, \Delta_3)$ and $R^C_{\mathbf{FCP}}(\Gamma, \Delta_4)$ iff for any $\varphi_1, \varphi_2, \varphi_3, \varphi_4 \in \Phi_{\mathcal{L}_{\mathsf{EPL}}}$, if $\mathbf{FCP}(\varphi_1, \varphi_2, \varphi_3, \varphi_4) \in \Gamma$, then $\varphi_1 \in \Delta_1, \varphi_2 \in \Delta_2, \varphi_3 \in \Delta_3$ and $\varphi_4 \in \Delta_4$,
- for any $\Gamma, \Delta \in \mathbf{X}^C$, $\Gamma \approx^C_a \Delta$ iff for any $\varphi \in \Phi_{\mathcal{L}_{\mathsf{EPL}}}$, if $\mathbf{K}_a\varphi \in \Gamma$, then $\varphi \in \Delta$, and
- for any $\Gamma \in \mathbf{X}^C$,

$$V^C(\Gamma)(s) := \begin{cases} \textbf{true} & \text{if } s \in \Gamma, \\ \textbf{false} & \text{otherwise.} \end{cases}$$

We define an equivalence relation modulo Θ on $\mathbf{X}^C$ as follows:

Definition 26 (*Equivalence Class*) Let Θ be a stuffed set of formulae that are logically finite with respect to EPL. We define, for $\Gamma, \Delta \in \mathbf{X}^C$,

$$\Gamma \equiv_\Theta \Delta \text{ iff } \Gamma \cap \Theta = \Delta \cap \Theta.$$

Then, $\equiv_\Theta$ is an equivalence relation modulo Θ on $\mathbf{X}^C$. We write $[\Gamma]_\Theta$ for the equivalence class of Γ under $\equiv_\Theta$.

A filtration of $\mathfrak{U}^C$ through Θ is defined as follows:

Definition 27 (*Filtration*) We define $\mathfrak{U}^\Theta := (\mathbf{X}^\Theta, R^\Theta_{\mathbf{FCP}}, L^\Theta, \{\approx^\Theta_a\}_{a \in \mathbf{A}}, V^\Theta)$ as a filtration of $\mathfrak{U}^C$ through Θ in which

- $\mathbf{X}^\Theta := \{[\Gamma]_\Theta : \Gamma \in \mathbf{X}^C\}$,

- $R^{\Theta}_{\mathbf{FCP}}$ is a binary relation on $\mathbf{X}^{\Theta}$ such that
  1. if $R^{C}_{\mathbf{FCP}}(\Gamma, \Delta)$, then $R^{\Theta}_{\mathbf{FCP}}([\Gamma]_{\Theta}, [\Delta]_{\Theta})$,
  2. if $R^{\Theta}_{\mathbf{FCP}}([\Gamma]_{\Theta}, [\Delta_1]_{\Theta})$,$R^{\Theta}_{\mathbf{FCP}}([\Gamma]_{\Theta}, [\Delta_2]_{\Theta})$,$R^{\Theta}_{\mathbf{FCP}}([\Gamma]_{\Theta}, [\Delta_3]_{\Theta})$, $R^{\Theta}_{\mathbf{FCP}}$ $([\Gamma]_{\Theta}, [\Delta_4]_{\Theta})$ and $\mathbf{FCP}(\varphi_1, \varphi_2, \varphi_3, \varphi_4) \in \Gamma \cap \Theta$, then $\varphi_1 \in \Delta_1, \varphi_2 \in \Delta_2, \varphi_3 \in \Delta_3$ and $\varphi_4 \in \Delta_4$,
- $L^{\Theta} : R^{\Theta}_{\mathbf{FCP}} \to \{\pi_1, \pi_2, \pi_3, \pi_4\}$ is a function such that $\pi_1([\Gamma]_{\Theta}) = [\Delta_1]_{\Theta}$, $\pi_2([\Gamma]_{\Theta}) = [\Delta_2]_{\Theta}, \pi_3([\Gamma]_{\Theta}) = [\Delta_3]_{\Theta}$ and $\pi_4([\Gamma]_{\Theta}) = [\Delta_4]_{\Theta}$,
- $\approx^{\Theta}_{a}$ is an equivalence relation on $\mathbf{X}^{\Theta}$ such that
  1. if $\Gamma \approx^{C}_{a} \Delta$, then $[\Gamma]_{\Theta} \approx^{\Theta}_{a} [\Delta]_{\Theta}$,
  2. if $[\Gamma]_{\Theta} \approx^{\Theta}_{a} [\Delta]_{\Theta}$ and $\mathbf{K}_a\varphi \in \Gamma$, then $\varphi \in \Delta$, and
- V^{Θ} is a function such that for any $s \in \Theta$,

$$V^{\Theta}([\Gamma]_{\Theta})(s) = V^{C}(\Gamma)(s).$$

Thus, for any $\xi \in \mathbf{X}^{\Theta}$,

$$[\![\varphi]\!]^{\mathfrak{U}^{\Theta}}_{a,\xi} := \{\eta : \xi \approx^{\Theta}_{a} \eta \text{ and } (\mathfrak{U}^{\Theta}, \eta) \models_{\mathsf{EPL}} \varphi\}$$

is well-defined for any φ that does not contain $\mathbf{WPR}_a$. We can prove the Lindenbaum lemma:

Lemma 2 (Lindenbaum) *Every* EPL*-consistent set of formulae is a subset of a* EPL*-maximal consistent set of formulae.*

We can prove the partial truth lemma:

Lemma 3 (Partial Truth) *If $\varphi \in \Theta$ and φ does not contain* $\mathbf{WPR}_a$*, then for any $\Gamma \in \mathbf{X}^{C}$,*

$$(\mathfrak{U}^{\Theta}, [\Gamma]_{\Theta}) \models_{\mathsf{EPL}} \varphi \textit{ iff } \varphi \in \Gamma.$$

We wish to supplement $\mathfrak{U}^{\Theta}$ with a preference space assignment ρ^{Θ} so as to obtain a multi-agent Domotor-type structured Kripke model $\mathfrak{U}^{\Theta}_{\sharp}$ for which the truth lemma holds for all formulae in Θ. Doing this contributes to solving the completeness problem of EPL. $\mathcal{F}^{\Theta}_{a,\xi}$ is defined as follows:

Definition 28 ($\mathcal{F}^{\Theta}_{a,\xi}$) For any $\xi \in \mathbf{X}^{\Theta}$, we define $\mathcal{F}^{\Theta}_{a,\xi}$ as the set of all $\alpha \subseteq X^{\Theta}_{a,\xi} := \{\eta : \xi \approx^{\Theta}_{a} \eta\}$ such that for some value formula $\varphi \in \Theta$, $\alpha = [\![\varphi]\!]^{\mathfrak{U}^{\Theta}}_{a,\xi}$.

We can prove the next lemma:

Lemma 4 (Boolean Algebra) *For any $\xi \in \mathbf{X}^{\Theta}$, $\mathcal{F}^{\Theta}_{a,\xi}$ is a Boolean algebra with $\emptyset$ as zero element and $X^{\Theta}_{a,\xi} = \{\eta : \xi \approx^{\Theta}_{a} \eta\}$ as unit element.*

$\preceq_{a,\xi}$ is defined as follows:

Definition 29 ($\preceq_{a,\xi}$) For any $\xi \in \mathbf{X}^{\Theta}$, we define $\alpha \preceq_{a,\xi} \beta$ to hold between elements $\alpha, \beta \in \mathcal{F}^{\Theta}_{a,\xi}$ iff there are value formulae $\varphi, \psi \in \Theta$ such that $\alpha = [\![\varphi]\!]^{\mathfrak{U}^{\Theta}}_{a,\xi}$, $\beta = [\![\psi]\!]^{\mathfrak{U}^{\Theta}}_{a,\xi}$ and $\mathbf{WPR}_a(\varphi, \psi) \in \Gamma$ for any $\Gamma \in \xi$.

We can prove the next lemma:

Lemma 5 ($\preceq_{a,\xi}$ and $\mathbf{WPR}_a$) *For any value formula $\varphi, \psi \in \Theta$ and any $\xi \in \mathbf{X}^{\Theta}$, $[\![\varphi]\!]^{\mathfrak{U}^{\Theta}}_{a,\xi} \preceq_{a,\xi} [\![\psi]\!]^{\mathfrak{U}^{\Theta}}_{a,\xi}$ iff, for any $\Gamma \in \xi$, $\mathbf{WPR}_a(\varphi, \psi) \in \Gamma$.*

The next lemma follows from Lemma 5.

Lemma 6 (Connectedness and Projectivity) *For any $\xi \in \mathbf{X}^{\Theta}$, $\preceq_{a,\xi}$ on $\mathcal{F}^{\Theta}_{a,\xi}$ satisfies Connectedness and Projectivity.*

Since we assumed that Θ is logically finite, $\mathbf{X}^{\Theta}$ is *finite*. Hence for any $\xi \in \mathbf{X}^{\Theta}$, $\mathcal{F}^{\Theta}_{a,\xi}$ is *finite*, so the next corollary follows from Theorem 1 and Lemmas 4 and 6.

Corollary 1 (Representation on $\mathcal{F}^{\Theta}_{a,\xi}$) For any $\xi \in \mathbf{X}^{\Theta}$, there are $P_{a,\xi} : \mathcal{F}^{\Theta}_{a,\xi} \to \mathbb{R}$ and $U_{a,\xi} : \mathcal{F}^{\Theta}_{a,\xi} \backslash \emptyset \to \mathbb{R}$ such that the following conditions hold for any $A, B \in \mathcal{F}^{\Theta}_{a,\xi} \backslash \emptyset$:

- $(X^{\Theta}_{a,\xi}, \mathcal{F}^{\Theta}_{a,\xi}, P_{a,\xi})$ is a finitely additive probability space,
- $A \preceq_{a,\xi} B$ iff $U_{a,\xi}(A) \leq U_{a,\xi}(B)$,
- If $A \cap B = \emptyset$, $U_{a,\xi}(A \cup B) = P_{a,\xi}(A|A \cup B)U_{a,\xi}(A) + P_{a,\xi}(B|A \cup B)U_{a,\xi}(B)$,
- When $A \in \mathcal{F}^{\Theta}_{a,\xi}$, if $P_{a,\xi}(A) = 0$, then $A = \emptyset$.

$\mathfrak{U}^{\Theta}_{\sharp}$ is defined as follows:

Definition 30 ($\mathfrak{U}^{\Theta}_{\sharp}$) We define $\mathfrak{U}^{\Theta}_{\sharp}$ as $(\mathbf{X}^{\Theta}, R^{\Theta}_{\mathbf{FCP}}, L^{\Theta}, \{\approx^{\Theta}_a\}_{a \in \mathbf{A}}, V^{\Theta}, \rho^{\Theta})$ in which ρ^{Θ} is a preference space assignment that assigns to each $a \in \mathbf{A}$ and each $\xi \in \mathbf{X}^{\Theta}$ $(X^{\Theta}_{a,\xi}, \mathcal{F}^{\Theta}_{a,\xi}, \preceq_{a,\xi}, \widehat{\ }, \times, +, -)$ that satisfies Connectedness and Projectivity in which

- $X^{\Theta}_{a,\xi} := \{\xi' : \xi \approx^{\Theta}_a \xi'\}$,
- $\mathcal{F}^{\Theta}_{a,\xi}$ is a Boolean algebra of subsets of $X^{\Theta}_{a,\xi}$ with $\emptyset$ as zero element and $X^{\Theta}_{a,\xi}$ as unit element,
- for any $a \in \mathbf{A}$ and $\xi_1, \xi_2 \in \mathbf{X}^{\Theta}$, if $\xi_1 \approx^{\Theta}_a \xi_2$, then $\rho^{\Theta}(a, \xi_1) = \rho^{\Theta}(a, \xi_2)$.

We can prove the full truth lemma:

Lemma 7 (Full Truth) *For any $\varphi \in \Theta$ and any $\Gamma \in \mathbf{X}^{C}$,*

$$(\mathfrak{U}^{\Theta}_{\sharp}, [\Gamma]_{\Theta}) \models_{\mathsf{EPL}} \varphi \textit{ iff } \varphi \in \Gamma.$$

Remark 11 This lemma is the announced improvement of Lemma 3.

We can prove the completeness of EPL as follows:

Theorem 5 (Completeness) For any $\varphi \in \Phi_{\mathcal{L}_{\mathsf{EPL}}}$, if $\models_{\mathsf{EPL}} \varphi$, then $\vdash_{\mathsf{EPL}} \varphi$.

Proof Suppose that $\nvdash_{\mathsf{EPL}} \varphi_0$. Then $\{\neg\varphi_0\}$ is a EPL-consistent set. By Lemma 2, $\{\neg\varphi_0\}$ is a subset of a EPL-maximal consistent set Γ. Evidently, $\varphi_0 \notin \Gamma$. Let Ψ be the set of subformulae of EPL which is finite and let Θ be the smallest stuffed extension of Ψ. By Lemma 1, Θ is logically finite with respect to EPL. If $\mathfrak{U}_\sharp^\Theta$ is constructed as above, it follows from Lemma 7 that $(\mathfrak{U}_\sharp^\Theta, [\Gamma]_\Theta) \not\models_{\mathsf{EPL}} \varphi_0$. Therefore, $\not\models_{\mathsf{EPL}} \varphi_0$.

Decidability of EPL We can prove the decidability of EPL by the finite model property lemma as follows:

Lemma 8 (Finite Model Property) EPL *has the finite model property that every non-theorem of* EPL *fails in a multi-agent Domotor-type structured Kripke model for knowledge and preference with only a finite number of elements.*

Theorem 6 (Decidability) EPL *is decidable.*

Proof Suppose that φ is not provable in EPL. By Lemma 8, φ fails in a multi-agent Domotor-type structured Kripke model $\mathfrak{U}_\sharp^\Theta$ for EPL with a finite number of elements. If we take a domain $\mathbf{X}^\Theta$ with, at most, that many elements, there are only a finite number of ways in which $R_{\mathbf{FCP}}^\Theta$, L^Θ, $\approx_a^\Theta$, and V^Θ can be defined, and there are also only a finite number of ways to define the preference space assignment ρ^Θ. Whether a defined relation, $\preceq_{a,\xi}$, satisfies Connectedness and Projectivity can be decided in a finite number of steps. Thus, we find, in at most a finite number of steps, a counter-model that falsifies the unprovable formula. In fact, we can compute an upper bound to the number of steps needed. Thus, EPL is decidable.

4 Dynamic Epistemic Preference Logic DEPL

4.1 Language

The language of DEPL $\mathcal{L}_{\mathsf{DEPL}}$ is defined as follows:

Definition 31 (*Language*) Let $\mathbf{S}$ denote a set of sentential variables, $\mathbf{A}$ a finite set of agents, $\mathbf{K}_a$ an epistemic operator, $\mathbf{WPR}_a$ a weak preference relation symbol, and [ ] an update operator. $\mathcal{L}_{\mathsf{DEPL}}$ is given by the following rule:

$$\varphi ::= s \mid \top \mid \neg\varphi \mid \varphi_1 \& \varphi_2 \mid \mathbf{K}_a(\varphi) \mid \mathbf{WPR}_a(\varphi_1, \varphi_2) \mid \mathbf{FCP}(\varphi_1, \varphi_2, \varphi_3, \varphi_4) \mid [\varphi_1]\varphi_2,$$

where $s \in \mathbf{S}$, $a \in \mathbf{A}$.

- $[\varphi_1]\varphi_2$ should be interpreted to mean that φ_2 is the case after everyone simultaneously and commonly learns that φ_1 is the case.
- The set of all well-formed formulae of $\mathcal{L}_{\mathsf{DEPL}}$ will be denoted by $\Phi_{\mathcal{L}_{\mathsf{DEPL}}}$.

4.2 Semantics

Updated Expected Utility and Updated Weak Preference Relation There are at least two modes of change that cause changes of preference:

1. valuational preference change,
2. doxastic preference change.

The former can be represented by change of the utility $U_{a,w_1}(\{w_2\})$ of $w_2 \in \mathbf{W}$. The latter can be represented by change of the probability function P_{a,w_1}. We provide an example of valuational preference changes:

Example 5 (Valuational Preference Change) In the situation of Example 2, the guest's original utility matrix might was this:

	Chicken	Beef	Herring
White	1	−1	1
Red	0	1	−1
Rosé	0.5	0	−1

Suppose that his original utility matrix changes to the following:

	Chicken	Beef	Herring
White	1	−1	1
Red	0	1	−1
Rosé	0.5	0.5	0

And suppose that the original probability matrix holds:

	Chicken	Beef	Herring	
White	0.4	0.4	0.2	0.2
Red	0.4	0.4	0.2	0.2
Rosé	0.4	0.4	0.2	0.4

Then the expected utilities of bringing white, red, and rosé will be as in the column at the right and the preference ordering of these acts will be this:

$$\text{White} \sim \text{Red} \prec \text{Rosé}.$$

We provide an example of doxastic preference changes:

Example 6 (Doxastic Preference Change) When the guest learns that beef is not to be served for dinner, his original probability matrix might change to the following: Then the expected utilities of bringing white, red, and rosé will be as in the column at the right and the preference ordering of these acts will be this:

	Chicken	Beef	Herring	
White	$\frac{2}{3}$	0	$\frac{1}{3}$	1
Red	$\frac{2}{3}$	0	$\frac{1}{3}$	$-\frac{1}{3}$
Rosé	$\frac{2}{3}$	0	$\frac{1}{3}$	0

$$\text{Red} \prec \text{Rosé} \prec \text{White}.$$

DEPL is based only on doxastic preference changes. In DEPL $U_{a,w_1}(\{w_2\})$ is fixed, but according as P_{a,w_1} changes, the expected utility U_{a,w_1} changes. In DEPL the change of probability function is executed by conditionalisation. Conditionalisation is defined as follows:

Definition 32 (*Conditionalisation*) Given $a \in \mathbf{A}$ and $w \in \mathbf{W}$, let $\mathcal{P}$ denote the set of all probability functions on $\mathcal{F}_{a,w}$. The function $\oplus : \mathcal{P} \times \mathcal{F}_{a,w} \to \mathcal{P}$ such that for any $A \in \mathcal{F}_{a,w}$,

$$\oplus(P_{a,w}, A)(B) := \begin{cases} \dfrac{P_{a,w}(A \cap B)}{P_{a,w}(A)} & \text{if } P_{a,w}(A) \neq 0, \\ \text{undefined} & \text{otherwise} \end{cases}$$

is called conditionalisation on A.

The updated expected utility is defined as follows:

Definition 33 (*Updated Expected Utility*) Given U_{a,w_1} and $A \in \mathcal{F}_{a,w_1}$, U_{a,w_1} such that for any $B \in \mathcal{F}_{a,w_1}$,

$$U_{a,w_1,A}(B) := \sum_{w_2 \in B} \oplus(P_{a,w_1}, A)(\{w_2\}) U_{a,w_1}(\{w_2\}) = U_{a,w_1}(A \cap B)$$

is called the updated expected utility on A.

The updated weak preference relation is defined as follows:

Definition 34 (*Updated Weak Preference Relation*) When $U_{a,w,A}$ defined by $U_{a,w}$ the existence of which is guaranteed by Theorem 1 is given, $\preceq_{a,w,A}$ such that for any $B, C \in \mathcal{F}_{a,w}$,

$$B \preceq_{a,w,A} C \text{ iff } U_{a,w,A}(B) \leq U_{a,w,A}(C)$$

is called the updated weak preference relation on A.

From Definition 33 and 34 and Theorem 1, the next corollary follows.

Corollary 2 (Original Weak Preference Relation and Updated Weak Preference Relation) Given $a \in \mathbf{A}$, $w \in \mathbf{W}$ and $A \in \mathcal{F}_{a,w}$, if $(W_{a,w}, \mathcal{F}_{a,w}, \preceq_{a,w}, \times, +, -)$ satisfies Connectedness and Projectivity, then for any $B, C \in \mathcal{F}_{a,w}$,

$$B \preceq_{a,w,A} C \text{ iff } A \cap B \preceq_{a,w} A \cap C$$

holds.

Truth By virtue of Corollary 2, we can provide DEPL with the following truth definition:

Definition 35 (*Truth*) When $\mathfrak{M} := (\mathbf{W}, R_{\mathbf{FCP}}, L, \{\approx_a\}_{a\in\mathbf{A}}, V, \rho)$ is given, the notion of $\varphi \in \Phi_{\mathcal{L}_{\mathsf{DEPL}}}$ being true at $w \in \mathbf{W}$ in $\mathfrak{M}$, in symbols $(\mathfrak{M}, w) \models_{\mathsf{DEPL}} \varphi$, is inductively defined as follows:

- $(\mathfrak{M}, w) \models_{\mathsf{DEPL}} s$ iff $V(w)(s) = \mathbf{true}$,
- $(\mathfrak{M}, w) \models_{\mathsf{DEPL}} \top$,
- $(\mathfrak{M}, w) \models_{\mathsf{DEPL}} \varphi_1 \& \varphi_2$ iff $(\mathfrak{M}, w) \models_{\mathsf{DEPL}} \varphi_1$ and $(\mathfrak{M}, w) \models_{\mathsf{DEPL}} \varphi_2$,
- $(\mathfrak{M}, w) \models_{\mathsf{DEPL}} \neg\varphi$ iff $(\mathfrak{M}, w) \not\models_{\mathsf{DEPL}} \varphi$,
- $(\mathfrak{M}, w_1) \models_{\mathsf{DEPL}} \mathbf{K}_a(\varphi)$ iff $(\mathfrak{M}, w_2) \models_{\mathsf{DEPL}} \varphi$ for all w_2 such that $w_1 \approx_a w_2$,
- $(\mathfrak{M}, w) \models_{\mathsf{DEPL}} \mathbf{FCP}(\varphi_1, \varphi_2, \varphi_3, \varphi_4)$ iff $(\mathfrak{M}, \pi_1(w)) \models_{\mathsf{DEPL}} \varphi_1$ and $(\mathfrak{M}, \pi_2(w)) \models_{\mathsf{DEPL}} \varphi_2$ and $(\mathfrak{M}, \pi_3(w)) \models_{\mathsf{DEPL}} \varphi_3$ and $(\mathfrak{M}, \pi_4(w)) \models_{\mathsf{DEPL}} \varphi_4$,
- $(\mathfrak{M}, w_1) \models_{\mathsf{DEPL}} \mathbf{WPR}_a(\varphi_1, \varphi_2)$ iff $[\![\varphi_1]\!]^{\mathfrak{M}}_{a,w_1} \preceq_{a,w_1} [\![\varphi_2]\!]^{\mathfrak{M}}_{a,w_1}$,
- $(\mathfrak{M}, w) \models_{\mathsf{DEPL}} [\varphi_1]\varphi_2$ iff $(\mathfrak{M}, w) \models_{\mathsf{DEPL}} \varphi_1$ implies $(\mathfrak{M}_{\varphi_1}, w) \models_{\mathsf{DEPL}} \varphi_2$, where $\mathfrak{M}_{\varphi_1}$ is the updated multi-agent Domotor-type structured Kripke model for knowledge and preference obtained from replacing each $\approx_a$ with its updated equivalence relation $\approx_{a,\varphi_1} := \{(w_1, w_2) : w_1 \approx_a w_2 \text{ and } (\mathfrak{M}, w_2) \models_{\mathsf{DEPL}} \varphi_1\}$ and replacing ρ with the updated preference assignment ρ_{φ_1} such that $\rho_{\varphi_1}(a, w_1) = (W_{a,w_1}, \mathcal{F}_{a,w_1}, \preceq_{a,w_1,[\![\varphi_1]\!]^{\mathfrak{M}}_{a,w_1}}, \frown, \times, +, -)$ in which $[\![\varphi_1]\!]^{\mathfrak{M}}_{a,w_1} := \{w_2 \in \mathbf{W} : w_1 \approx_a w_2 \text{ and } (\mathfrak{M}, w_2) \models_{\mathsf{DEPL}} \varphi_1\}$ and, for any $B, C \in \mathcal{F}_{a,w_1}$, $B \preceq_{a,w_1,[\![\varphi_1]\!]^{\mathfrak{M}}_{a,w_1}} C$ iff $[\![\varphi_1]\!]^{\mathfrak{M}}_{a,w_1} \cap B \preceq_{a,w_1} [\![\varphi_1]\!]^{\mathfrak{M}}_{a,w_1} \cap C$.

If $(\mathfrak{M}, w) \models_{\mathsf{DEPL}} \varphi$ for all $w \in \mathbf{W}$, we write $\mathfrak{M} \models_{\mathsf{DEPL}} \varphi$ and say that φ is valid in $\mathfrak{M}$. If φ is valid in all multi-agent Domotor-type structured Kripke models for knowledge and preference, we write $\models_{\mathsf{DEPL}} \varphi$ and say that φ is valid.

It is a nontrivial matter whether or not the updated model satisfies the conditions that the original model satisfied. We must prove that the updated model satisfies such conditions.

Proposition 3 (Original Model and Updated Model) *If* $\mathfrak{M} := (\mathbf{W}, R_{\mathbf{FCP}}, L, \{\approx_a\}_{a\in\mathbf{A}}, V, \rho)$ *satisfies Consistency and World-Dependent Preference, and* $\rho(a, w) := (W_{a,w}, \mathcal{F}_{a,w}, \preceq_{a,w}, \frown, \times, +, -)$ *satisfies Connectedness and Projectivity,* $\mathfrak{M}_\varphi := (\mathbf{W}, R_{\mathbf{FCP}}, L, \{\approx_{a,\varphi}\}_{a\in\mathbf{A}}, V, \rho_\varphi)$ *also satisfies Consistency and World-Dependent Preference, and* $\rho_\varphi := (W_{a,w}, \mathcal{F}_{a,w}, \preceq_{a,w,[\![\varphi]\!]_{a,w}}, \frown, \times, +, -)$ *also satisfies Connectedness and Projectivity.*

4.3 Syntax

We provide DEPL with the following proof system.

Definition 36 (*Proof System*)

• Axioms of DEPL

Besides (A1), (A2), (A3), (A4), (A5), (A6), (A7), (A8), (A9), (A10), (A11), (A12), (A13), (A14), (A15) and (A16), the proof system of DEPL has the following axioms:

(A17) $[\varphi]s \leftrightarrow (\varphi \rightarrow s)$ (**Atomic Permanence**),
(A18) $[\varphi_1]\neg\varphi_2 \leftrightarrow (\varphi_1 \rightarrow \neg[\varphi_1]\varphi_2)$ (**Announcement and Negation**),
(A19) $[\varphi_1](\varphi_2 \& \varphi_3) \leftrightarrow ([\varphi_1]\varphi_2 \& [\varphi_1]\varphi_3)$ (**Announcement and Conjunction**),
(A20) $[\varphi_1]\mathbf{K}_a(\varphi_2) \leftrightarrow (\varphi_1 \rightarrow \mathbf{K}_a([\varphi_1]\varphi_2))$ (**Announcement and Knowledge**),
(A21) $[\varphi_1]\mathbf{WPR}_a(\varphi_2, \varphi_3) \leftrightarrow \mathbf{WPR}_a(\varphi_1 \& \varphi_2, \varphi_1 \& \varphi_3)$ (**Announcement and Weak Preference**),
(A22) $[\varphi_1]\mathbf{FCP}(\varphi_2, \varphi_3, \varphi_4, \varphi_5) \leftrightarrow (\varphi_1 \rightarrow \mathbf{FCP}([\varphi_1]\varphi_2, [\varphi_1]\varphi_3, [\varphi_1]\varphi_4, [\varphi_1]\varphi_5))$ (**Announcement and Four-Fold Cartesian Product**),
(A23) $[\varphi_1][\varphi_2]\varphi_3 \leftrightarrow [\varphi_1 \& [\varphi_1]\varphi_2]\varphi_3$ (**Announcement and Composition**).

• Inference Rules of DEPL

Besides (R1), (R2), (R3) and (R4), the axiom system of DEPL has the following inference rule:

$$\text{(R5)} \quad \frac{\varphi_2}{[\varphi_1]\varphi_2} \quad (\textbf{Announcement Necessitation}).$$

A proof of $\varphi \in \Phi_{\mathcal{L}_{\mathsf{DEPL}}}$ is a finite sequence of $\mathcal{L}_{\mathsf{DEPL}}$-formulae having φ as the last formula such that either each formula is an instance of an axiom, or it can be obtained from formulae that appear earlier in the sequence by applying an inference rule. If there is a proof of φ, we write $\vdash_{\mathsf{DEPL}} \varphi$.

4.4 Soundness and Completeness

We can prove the soundness of DEPL in the usual way.

Theorem 7 (Soundness) *For every $\varphi \in \Phi_{\mathcal{L}_{\mathsf{DEPL}}}$, if $\vdash_{\mathsf{DEPL}} \varphi$, then $\models_{\mathsf{DEPL}} \varphi$.*

In order to prove the completeness of DEPL, we give a translation function $t : \mathcal{L}_{\mathsf{DEPL}} \rightarrow \mathcal{L}_{\mathsf{EPL}}$. Because the completeness of EPL is proved, it suffices to prove that every well-formed formula is equivalent to its translation in DEPL. This method is usual in the literature of dynamic epistemic logic.[7]

Definition 37 (*Translation Function*) A translation function $t : \mathcal{L}_{\mathsf{DEPL}} \rightarrow \mathcal{L}_{\mathsf{EPL}}$ is defined as follows:

[7] For this method, consult Van Ditmarsch et al. (2007, pp. 186–189).

1. $t(s) = s$,
2. $t(\top) = \top$,
3. $t(\neg\varphi) = \neg t(\varphi)$,
4. $t(\varphi_1 \& \varphi_1) = t(\varphi_1) \& t(\varphi_2)$,
5. $t(\mathbf{K}_a(\varphi)) = \mathbf{K}_a(t(\varphi))$,
6. $t(\mathbf{WPR}_a(\varphi_1, \varphi_2)) = \mathbf{WPR}_a(t(\varphi_1), t(\varphi_2))$,
7. $t([\varphi]s) = t(\varphi \rightarrow s)$,
8. $t([\varphi_1]\neg\varphi_2) = t(\varphi_1 \rightarrow \neg[\varphi_1]\varphi_2)$,
9. $t([\varphi_1](\varphi_2 \& \varphi_2)) = t([\varphi_1]\varphi_2 \& [\varphi_1]\varphi_3)$,
10. $t([\varphi_1]\mathbf{K}_a(\varphi_2)) = t(\varphi_1 \rightarrow \mathbf{K}_a([\varphi_1]\varphi_2))$,
11. $t([\varphi_1]\mathbf{WPR}_a(\varphi_2, \varphi_3)) = t(\mathbf{WPR}_a(\varphi_1 \& \varphi_2, \varphi_1 \& \varphi_3))$,
12. $t([\varphi_1]\mathbf{FCP}(\varphi_2, \varphi_3, \varphi_4, \varphi_5)) = t(\varphi_1 \rightarrow \mathbf{FCP}([\varphi_1]\varphi_2, [\varphi_1]\varphi_3, [\varphi_1]\varphi_4, [\varphi_1]\varphi_5))$,
13. $t([\varphi_1][\varphi_2]\varphi_3) = t([\varphi_1 \& [\varphi_1]\varphi_2]\varphi_3)$.

We can prove the following lemma:

Lemma 9 (Translation) *For every* $\varphi \in \Phi_{\mathcal{L}_{\mathsf{DEPL}}}$, $\vdash_{\mathsf{DEPL}} t(\varphi) \leftrightarrow \varphi$.

By virtue of Theorem 5 and Lemma 9, we can prove the completeness of DEPL.

Theorem 8 (Completeness) *For every* $\varphi \in \Phi_{\mathcal{L}_{\mathsf{DEPL}}}$, *if* $\models_{\mathsf{DEPL}} \varphi$, *then* $\vdash_{\mathsf{DEPL}} \varphi$.

5 Conclusion

We have proposed a sound and complete epistemic preference logic (EPL) and extended it to dynamic epistemic preference logic (DEPL) that is also sound and complete. Van Benthem and Liu's DEUL is designed only to deal with the dynamic interactions between knowledge and preferences originating from decision makings under certainty. On the other hand, DEPL is designed to deal with the dynamic interactions between knowledge and preferences originating from decision makings under certainty, risk, uncertainty and ignorance. So DEPL has much wider scope of application than DEUL. Providing DEPL with measurement-theoretic semantics has enabled it to have such wide scope.

This chapter is only a part of a larger measurement-theoretic study. By means of measurement theory, we constructed or are trying to construct such logics as

1. dyadic deontic logic Suzuki (2009b),
2. threshold-utility-maximiser's preference logic Suzuki (2010, 2011c),
3. vague predicate logic Suzuki (2011a, 2011b),
4. interadjective-comparison logic Suzuki (2012a, 2012c),
5. gradable-predicate logic Suzuki (2012b),
6. logic for better questions and answers Suzuki (2012d),
7. doxastic and epistemic logic Suzuki (2013), and
8. multidimensional-predicate-comparison logic Suzuki (2012e).

Acknowledgments This chapter is a largely revised version of Suzuki (2009a). This work was supported by Grant-in-Aid for Scientific Research (C) (23520015). The author would like to thank two anonymous reviewers for their helpful comments.

References

Bolker, E. D. (1966). Functions resembling quotients of measures. *Transactions of the American Mathematical Society*, *124*, 292–312.

Bolker, E. D. (1967). A simultaneous axiomatisation of utility and subjective probability. *Philosophy of Science*, *34*, 333–340.

Boutilier, C. (1994). Toward a logic for qualitative decision theory. In: *Proceedings of the 4th international conference on principles of knowledge representation and reasoning* (KR-94), Bonn (pp. 75–86).

Cantor, G. (1895). Beiträge zur Begründung der Transfiniten Mengenlehre I. *Mathematische Annalen*, *46*, 481–512.

Domotor, Z. (1978). Axiomatisation of Jeffrey utilities. *Synthese*, *39*, 165–210.

Halpern, J. Y. (2003). *Reasoning about uncertainty*. Cambridge, Mass: The MIT Press.

Hansson, S. O. (2001). Preference logic. In D. M. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (2nd ed., Vol. 4, pp. 319–393). Heidelberg: Springer-Verlag.

Hansson, S. O. (2006). Preferences. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*.

Hölder, O. (1901). Die Axiome der Quantität und die Lehre vom Mass. Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig. *Mathematisch-Physikalische Klasse, 53*, 1–64.

Jeffrey, R. (1978). Axiomatising the logic of decision. In: C. A. Hooker et al. (Eds.), *Foundations and applications of decision theory* (Vol. 1, pp. 232–237). Dordrecht: Kluwer. Reprinted in Jeffrey (1992).

Jeffrey, R. (1990). *The logic of decision* (Corrected 2nd ed.). Chicago: University of Chicago Press.

Jeffrey, R. (1992). *Probability and the art of judgment*. Cambridge: Cambridge UP.

Joyce, J. M. (1999). *The foundations of causal decision theory*. Cambridge: Cambridge UP.

Krantz, D. H., et al. (1971). *Foundations of measurement* (Vol. I). New York: Academic Press.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions*. New York: Wiley.

Luce, R. D., et al. (1990). *Foundations of measurement* (Vol. III). San Diego: Academic Press.

Naumov, P. (2006). Logic of subtyping. *Theoretical Computer Science*, *357*, 167–185.

Ramsey, F. P. (1926). Truth and Probability. In D. H. Mellor (Ed.), *Philosophical papers (1990)* (pp. 52–94). Cambridge: Cambridge UP.

Roberts, F. S. (1979). *Measurement theory*. Reading: Addison-Wesley.

Scott, D. (1964). Measurement structures and linear inequalities. *Journal of Mathematical Psychology*, *1*, 233–247.

Segerberg, K. (1971). Qualitative probability in a modal setting. In J. E. Fenstad (Ed.), *Proceedings of the second scandinavian logic symposium* (pp. 341–352). Amsterdam: North-Holland.

Suppes, P., et al. (1989). *Foundations of measurement* (Vol. II). San Diego: Academic Press.

Suzuki, S. (2009a). Prolegomena to dynamic epistemic preference logic. In H. Hattori et al. (Eds.), *New frontiers in artificial intelligence* (pp. 177–192). LNCS 5447. Heidelberg: Springer-Verlag.

Suzuki, S. (2009b). Measurement-theoretic foundation of preference-based dyadic deontic logic. In X. He et al. (eds.), *Proceedings of the second international workshop on logic, rationality, and interaction (LORI-II)* (pp. 278–291). LNCS 5834. Heidelberg: Springer-Verlag.

Suzuki, S. (2010). Prolegomena to threshold utility maximiser's preference logic. In: *Electronic proceedings of the 9th conference on logic and the foundations of game and decision theory* (LOFT 2010). Paper No. 44.

Suzuki, S. (2011a). Prolegomena to salient-similarity-based vague predicate logic. In T. Onada et al. (Eds.), *New frontiers in artificial intelligence* (pp. 75–89). LNCS 6797. Heidelberg: Springer-Verlag.

Suzuki, S. (2011b). Measurement-theoretic foundations of probabilistic model of JND-based vague predicate logic. In H. van Ditmarsch et al. (Eds.), *Proceedings of the Third International Workshop on Logic, Rationality, and Interaction* (LORI-III) (pp. 272–285). LNCS 6953. Heidelberg: Springer-Verlag.

Suzuki, S. (2011c). A measurement-theoretic foundation of threshold utility maximiser's preference logic. *Journal of Applied Ethics and Philosophy*, *3*, 17–25.

Suzuki, S. (2012a). Measurement-theoretic foundations of interadjective-comparison logic. In A. Aguilar-Guevara et al. (Eds.), *Proceedings of Sinn und Bedeutung* 16 (Vol. 2, pp. 571–584). Cambridge, Mass. MIT Working Papers in Linguistics.

Suzuki, S. (2012b). Measurement-theoretic foundations of gradable-predicate logic. In M. Okumura et al. (Eds.) *New frontiers in artificial intelligence* (pp. 82–95). LNCS 7258. Heidelberg: Springer-Verlag.

Suzuki, S. (2012c). Psychophysical foundations of interadjective-comparison logic. In: *Proceedings of the first logic and cognition conference* (L& C 2012) (pp. 87–98).

Suzuki, S. (2012d). Measurement-, information-, and decision-theoretic foundations of logic for better questions and answers. In: *Proceedings of the 10th conference on logic and the foundations of game and decision theory* (LOFT 2012) (pp. 504–514).

Suzuki, S. (2012e). Measurement-theoretic foundations of multidimensional-predicate-comparison logic, mimeo.

Suzuki, S. (2013). Remarks on decision-theoretic foundations of doxastic and epistemic logic. *Studies in Logic*, *6*, 1–12.

Van Benthem, J., et al. (2005). Preference logic, conditionals and solution concepts in games. ILLC Amsterdam: ILLC Prepublication Series PP-2005-28.

Van Benthem, J. & Liu, F. (2007). Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logic*, *17*, 157–182.

Van Ditmarsch, H., et al. (2007). *Dynamic epistemic logic*. Heidelberg: Springer-Verlag.

A Modal Scalar-Presuppositional Analysis of *Only*

Katsuhiko Yabushita

Abstract Ippolito noted that there was a certain feature of *only* that could not be explained by any existing analysis of *any*, i.e., the asymmetry between positive and negative *only* sentences in the cancelability of the prejacent; the former can be canceled, while the latter cannot. She then argued that the feature was best analyzed by incorporating a scalar presupposition in the semantics of *only*; roughly, of all the relevant alternatives to the focused element, the element is the most likely to have the property in question. In the current work, however, by demonstrating that her proposed presupposition is empirically inadequate, we will argue that the scalar presupposition in question is not a presupposition *simpliciter*—i.e., a proposition that must be assumed to be true both by the speaker and the addressee—but one restricted to the speaker. We then propose an alternative, modal scalar-presuppositional analysis of *only* as a modification of the theory of *only* proposed by van Rooij and Schulz.

Keywords Semantics and pragmatics of *only* · Asymmetry between positive and negative *only* sentences in the cancelability of the prejacent · Modal scalar-presupposition

1 Introduction

The meaning of an *only* sentence has been considered to consist of what van Rooij and Schulz (2007) called the *negative* and the *positive* contributions. Consider the following example.

(1) Only $[\text{Mary}]_F$ can sing.

K. Yabushita (✉)
Department of English, Naruto University of Education, Takashima, Naruto 772–8502, Japan
e-mail: yabuchan@naruto-u.ac.jp

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*,
Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_15,

(1) is taken to mean that except for Mary, nobody can sing (negative contribution); and that Mary can sing (positive contribution). As for the negative contribution of an *only* sentence, its status as a genuine part of the semantic meaning of the sentence has been uncontroversially assumed in the literature. On the other hand, the status of the positive contribution has been controversial, with many analyses having been proposed. The existing analyses can be classified into three groups according to their treatment of the positive contribution as (i) entailment (Atlas 1993, 1996), (ii) presupposition (Horn 1969, 1996; Rooth 1985, 1992; Geurts and van der Sandt 2004), or (iii) conversational implicature (McCawley 1981; van Rooij and Schulz 2004, 2006, 2007).

Ippolito (2006, 2008) argued that there is a feature of *only* sentences that no existing analysis can account for. She proposed an alternative analysis in which a scalar presupposition was posited for the semantics of *only*. In the current chapter, we will review the feature of *only*-sentences that Ippolito argued cannot be accounted for by the existing analyses, in order to make sure that this is indeed the case. Then, we will demonstrate that her alternative analysis, which was designed to explain the feature, has apparent empirical problems, although it is on the right track. We will diagnose what went wrong with her analysis and propose an alternative analysis, which is a modification of the theory of *only* proposed by van Rooij and Schulz (2007).

2 Asymmetry Between Positive and Negative Only-Sentences

Ippolito (2006, 2008) took notice of a feature of *only* sentences that previously had been given little attention in the literature. That is, there is an asymmetry between a positive *only* sentence and the corresponding negative *only* sentence with respect to the cancelability of the positive contribution. Before illustrating the asymmetry in question, let us introduce a term to designate what has been referred to as the positive contribution of an *only* sentence. Given an *only* sentence, let us refer to (the proposition denoted by) the sentence resulting from it by removing *only* (as well as the word *not*, if it appears) as the *prejacent*.

Consider the following two sets of sentences:

(2) a. Only Mary can speak French.
b. Mary can speak French (Prejacent).
c. Only Mary can speak French, and maybe not even she can. (Ippolito 2008: (44b))
d. #Only Mary can speak French—in fact, not even she can. (Ippolito 2008: (44a))

(3) a. Not only John can speak French.
b. John can speak French (Prejacent).
c. #Not only John can speak French, and maybe he can't (Ippolito 2008: (38)).

As is indicated by the contrast in felicity between (2c) and (3c), it is generally the case that the prejacent of a positive *only* sentence is cancelable, while that of a negative *only* sentence is not.[1] This is rendered as the following observation:

(4) **Observation** (Asymmetry between Positive and Negative *Only* Sentences Ippolito 2006, 2008):
In general, the prejacent of a positive *only* sentence is cancelable, while that of a negative *only* sentence is not. Here, a word is in order about the cancelability of the prejacent of a positive only sentence. The prejacent cannot be canceled by its flat negation as indicated by (2d), but the cancelation characteristically requires an epistemic possibility operator such as *maybe* as seen in (2c).

In the literature, quite a few analyses have been proposed to treat the prejacent of a positive *only* sentence; however, relatively little has been said about the prejacent of a negative *only* sentence, and even fewer proposals have been made to deal with the asymmetry between the two. The natural question to ask here is if the previous analyses of the prejacent of a positive *only* sentence can be applied or extended to treat the prejacent of a negative *only* sentence as well as the noted asymmetry. Ippolito (2006, 2008) convincingly argued that none of the existing analyses can account for the non-cancelability of the prejacent of a negative *only* sentence, not to mention the asymmetry in question. We will not review the arguments here due to space limitations.

3 Ippolito's Alternative Analysis: Scalar Presupposition

Ippolito observed that a (positive) *only* sentence "*Only* A P" characteristically has an implication which can be paraphrased as something like 'A is the most likely to P (of the contextually relevant alternatives to A)'. In fact, she pushed the observation further to propose that the implication was an inherent part of the meaning of *only*; to be exact, a presupposition thereof. With the presupposition backdropped, she claimed that the asymmetry between positive and negative *only* sentences with regards to the cancelability of the prejacent would follow automatically. In the following subsections, we will critically review her analysis in detail and argue that although it is on the right track, the analysis is by no means adequate, with apparent methodological and empirical problems.

3.1 Scalar Presupposition of Only

Ippolito (2006, 2008) proposed that *only* indeed should be a presupposition trigger; however, the presuppositional content is neither the prejacent itself, as in the

[1] However, the *possibility* of the prejacent is *not* cancelable, as attested by the contrast in felicity between (2c) and (2d), as originally pointed out by Atlas (1991, 1993)]

strong presuppositional analyses, nor is it an existential proposition, as in the weak presuppositional analyses. Here is the presupposition she proposed for *only*:

(5) **(Ippolito 2008: (50)):**
Presupposition triggered by *only* (where ϕ is the sentence in its scope):
If some proposition in C is true, then ϕ is true.

Obviously, a word is in order about *C* in (5). Assuming *only* was a focus-sensitive operator, Ippolito adopted Rooth (1992) approach to focus, in which focus is taken to induce a set of alternative propositions in the following sense. Given a sentence ϕ containing a focused constituent, say *A*, *C* denotes the set of propositions alternative to $[\![\phi]\!]$with $[\![A]\!]$ replaced by some "relevant" alternative value.

To take (1) as an example, suppose that the contextually relevant set of people as to the question who can sing is {Mary, Sue, John, Bill, ...}. Then, the value of *C* in (5) will be (the conjunction closure of) {p : ∃x[p = ^[x can sing] & x ∈{Mary, Sue, John, Bill, ...}]}. With the value of *C* being determined and ϕ being "[Mary]$_F$ can sing", the presupposition of (1) induced by *only* is:

(6) $\forall q \in \{\text{p} : \exists \text{x}[\text{p} =^\wedge [\text{x can sing}] \ \& \ \text{x} \in \{\text{Mary, Sue, John, Bill}, \ldots\}]\}[q(w) = 1 \rightarrow [\![[\text{Mary}]_\text{F} \text{ can sing}]\!]^w = 1]$

In words, if a proposition of the form "x can sing", where x is a contextually relevant alternative to Mary, is true, then Mary can sing; equivalently, if Mary cannot sing, nobody can sing. This implicational presupposition in practice puts Mary at the top of the scale of who is most likely to be able to sing, which is why Ippolito characterized the presupposition induced by *only* as scalar.[2] The scalar nature of the alleged presupposition of *only* indeed is motivated by the following fact. That is, as seen in (2c), a statement canceling the prejacent of a (positive) *only* sentence characteristically is accompanied by adverb *even*, which has been commonly acknowledged as a scalar presupposition trigger since Kartunnen and Peters (1979).

3.2 *Prejacent of an Only Sentence: Ippolito's Story*

Now that her proposed presupposition of *only* has been set forth, it is time to see how Ippolito (2006) claimed to account for the prejacent of a positive *only* sentence and that of a negative one, as well as the asymmetry between them. Briefly, her story runs like this: The prejacent of a positive *only* sentence is a conversational implicature, or more specifically, a scalar implicature derived from the literal meaning by means of Gricean reasoning; thus, it is cancelable. On the other hand, the prejacent of a negative *only* sentence is an entailment derived from the presupposition and the literal meaning; hence, it is not cancelable. In this section, we will critically review

[2] *Only* as a scalar-presupposition trigger was independently argued for by Krasikova and Zhechev (2005), who proposed an analysis of *only* occurring in what von Fintel and Iatridou (2005) called "Sufficiency Modal Construction" (SMC), as in "To get good cheese you only have to go to the North End."

the arguments to point out that in her analysis as its stands, not only the prejacent but also its negation can be derived as a conversational implicature for a positive *only* sentence and that, furthermore, there is a fundamental empirical problem with the presupposition she proposed.

3.2.1 Prejacent of a Positive *Only* Sentence as Conversational Implicature

Ippolito proposed to analyze the prejacent of a positive *only* sentence as a case of conversational implicature based on a scalar implicature (Hirschberg 1985). Before reviewing her argument, let us present the standard view of scalar implicature. Archetypal examples of scalar implicature are the default interpretations of the (b) sentences in the context of the (a) sentences in (7) and (8):

(7) a. How many students liked the course?
b. [Most]$_F$ students did.
c. All/Every $\prec_H$ Most $\prec_H$ Many $\prec_H$ Some

(8) a. How many children does John have?
b. He has [two]$_F$ children.
c. ... $\prec_H$ three $\prec_H$ two $\prec_H$ one $\prec_H$ zero

The default readings of (7b) and (8b) are "Most, but not all students liked the course" and "John has exactly two children", respectively. The neo-Gricean account of the type of implicature in question has been standardly formulated as follows (Gazdar 1979, Levinson 1983). When a sentence and sentences resulting from the sentence by replacing a constituent with some alternative expression form propositions that are linearly ordered in terms of strength, or entailment, the constituent and the alternatives are said to comprise a *Horn scale*. Examples of this scale are (7c) and (8c), where $\prec_H$ represents the "is stronger or more informative than" relation. The utterance of the sentence in question is considered to imply that the sentence with the constituent replaced with any stronger item in the Horn scale is not true, for if it were true the speaker would utter the stronger sentence in accordance with the maxim of quantity (Grice 1975).

Ippolito's computation of the prejacent of (2a), i.e., "Mary can speak French" as a conversational implicature derived from the utterance of (1), whose literal meaning is 'Nobody other than Mary can sing' goes like this. Without referring to a *Horn* scale explicitly, she just considers a competing proposition that asymmetrically entails the (literal meaning) of (2a), in this case, 'Nobody can speak French'. If the speaker could have, she should have uttered the stronger proposition (Maxim of Quantity). That she did not implies that she does not believe that the stronger proposition is the case (Maxim of Quality). In the reasoning for scalar implicature presented above, the utterance would be interpreted to imply the negation of 'Nobody can speak French', i.e., 'Somebody can speak French', which in conjunction with the scalar presupposition, 'If somebody can speak French, then Mary can' implies 'Mary can speak French'. This is how Ippolito claimed the prejacent of (2a), i.e., (2b) was to be derived as a conversational implicature; thus, it was cancelable.

We have seen Ippolito's analysis of the prejacent of a(n) (positive) *only* sentence as a conversational implicature. At first sight, it seems very reasonable, but we will show that the analysis can predict a wrong implicature as well. That is, for the calculation of a scalar implicature of the utterance of (2a), Ippolito assumed 'Nobody can speak French' for a competing stronger statement as it asymmetrically entails the literal meaning of (2a), i.e., 'Nobody other than Mary can speak French', (which does not commit itself to whether Mary can speak French or not). In this vein, however, she might as well have assumed 'Mary but nobody else can speak French' as a competing stronger statement, for it asymmetrically entails (2a). Given 'Mary but nobody else can speak French' as the competing stronger statement of (2a), the negation of the statement will be derived as a scalar implicature, i.e., 'Mary cannot speak French or somebody other than Mary can speak French', which is in effect reduced to 'Mary cannot speak French', as the second conjunct is incompatible with the literal meaning of (2a).

It has been demonstrated that 'Somebody can speak French' and 'Mary can speak French', which is derived from the existential conversational implicature in conjunction with the scalar presupposition, are not the only conversational implicatures that are predicted by Ippolito's scalar implicature-based analysis. But a clearly unwanted conversational implicature, i.e., 'Mary cannot speak French' is also predicted. It should be obvious that any analysis that will predict two contradictory statements as conversational implicatures for a given utterance should be rejected as inadequate.

From the above discussion, we contend, it must be concluded that Ippolito's argument for the prejacent of a positive *only* sentence being a conversational implicature is by no means satisfactory as it stands, unless she puts forth a strong argument that 'Nobody can speak French' is the competing stronger statement for (2a), not 'Mary but nobody else can speak French' on independent grounds.

3.2.2 Prejacent of a Negative *Only* Sentence as Entailment

In the foregoing discussion, we reviewed Ippolito's argument that the prejacent of a positive *only* sentence is a conversational implicature derived from a scalar implicature and found that it is by no means sound. Here, we will review her argument that the prejacent of a negative *only* sentence is entailed from the combination of the scalar presupposition of *only*, as proposed in (5) and the literal meaning of the negative *only* sentence.

Let us take (3a) as an example. Unlike the case of the prejacent of a positive *only* sentence, the existential statement, in this case, 'Somebody (other than John) can speak French' is not a conversational implicature, but the literal content of the utterance, thus, it is not cancelable. Combined with the presupposition for *only* 'If somebody can speak French, John can speak French, or, equivalently, if John cannot speak French, nobody can speak French', the existential statement entails the prejacent, i.e., 'John can speak French'. Unlike the case of a positive *only* sentence, Ippolito's account of the prejacent for a negative *only* sentence seems straightforward.

However, in the following section, we will demonstrate that one of the premises of the account, i.e., the presupposition of *only*, has obvious empirical problems.

3.3 Problems with Ippolito's Analysis: Status of the Alleged Presupposition

We have seen that Ippolito's presupposition of *only* in (5) plays a crucial role in explaining the derivation of the prejacent for an *only* sentence, positive or negative. In this subsection, however, we will demonstrate that the presupposition as currently formulated is empirically inadequate.

Consider the following discourse:

(9) Mary is the least likely to be able to sing. But, only Mary can sing.

If the presupposition for *only* proposed in (5) were indeed the case, the utterance of the second sentence in (9) would require it to hold that Mary is the most likely to be able to sing, at the time of the utterance. However, as can be observed, the second sentence is immediately preceded by a statement that explicitly contradicts the alleged presupposition. Thus, it is predicted that the discourse would be infelicitous; however, the fact is that it is perfectly felicitous.

For another piece of evidence against the alleged presupposition for *only* in (5), consider the following question-answer dialogue:

(10) Q: Who can sing?
A: Only Mary.

Again, if the presupposition in question were real, it would be presupposed at the utterance of (10A) that Mary is the most likely to be able to sing. On the assumption that a presupposition of a sentence is a proposition that must be mutually assumed by the speaker and the addressee at the utterance, the alleged presupposition for (10A), i.e., that Mary is the most likely to be able to sing, would have to be assumed by the questioner as well as the answerer. The fact of the matter is that one can ask question (10Q) felicitously without any assumption whatsoever as to who is most likely to be able to sing.[3]

---

[3] Actually, Ippolito (2008) criticized the strong presupposition analysis based on an exactly parallel question-answer pair ((17) therein):

(i) Q: Who can speak French?
A: Only John_F.

She pointed out that according to the strong presupposition analysis, the question-answer sequence should be infelicitous, for the prejacent, in this case, 'John can speak French' as a presupposition would be assumed to be the case by the questioner as well; however, the fact of the matter is the sequence is perfectly felicitous and the question is normally to be asked by somebody who doesn't know the prejacent is the case. As has been demonstrated, the analogous criticism is applicable to Ippolto's scalar presupposition analysis. That is, the question can be felicitously asked by somebody

4 Our Alternative Analysis: Modal Scalar Presupposition

In the previous section we reviewed Ippolito (2006, 2008) analysis of the asymmetry between positive and negative *only* sentences with respect to the cancelability of the prejacent, and we found that it in fact does not deliver what it promised. Specifically, the scalar implicature-based reasoning that the prejacent of a positive *only* sentence is a conversational implicature has a side effect of predicting an unwanted conversational implicature that is a contradiction to the prejacent. Furthermore, the argument crucially depends on the presupposition proposed in (5) in that the prejacent of an *only* sentence is entailed from the presupposition and a scalar implicature or the literal meaning. However, the proposed presupposition has been shown to be empirically inadequate, making false predictions. In the current section, first, we will see some evidence that something like the scalar proposition in (5) indeed is at work in the interpretation of *only* sentences; however, not as a presupposition *simpliciter* as proposed by Ippolito, but rather as one *modalized to the speaker's knowledge*; thus, we will propose a modal scalar-presupposition for *only*. Then, we will show that, coupled with an existing theory of the semantics and pragmatics of *only*, the proposed modal scalar-presupposition can account for the facts about *only* sentences under consideration, namely the status of the prejacent and the asymmetry between positive and negative *only* sentences, free from the problems associated with Ippolito's analysis.

4.1 Modal Scalar-Presupposition

In Sect. 3.3, we observed that a discourse which would be predicted to be infelicitous by the presupposition for *only* proposed by Ippolito was in fact felicitous. This of course suggests that the presupposition in (5) should be rejected. We should not throw out the baby with the bathwater, though. As we remarked, the characteristic occurrence of *even* in a statement canceling the prejacent of a positive *only* sentence definitely suggests that some scalar proposition like the one in (5) should be involved somehow in the interpretation of an *only* sentence, if not as a presupposition at the matrix level.

For a clue as to the status of the scalar proposition in question in the interpretation of an *only* sentence, we present the following example for consideration.

(11) Mary is the person who I'm least certain can sing. #But in fact only Mary/she can sing.

Note that (11) is similar to (9) in that the scalar proposition that Mary is the least likely to be able to sing is involved; however, the two examples are different in that the proposition is asserted at the matrix level in (9) while it is predicated of the

who doesn't assume that Mary is the most likely to be able to speak French, contrary to Ippolito's analysis.

speaker's knowledge in (11). The fact that (9) is felicitous while (11) is infelicitous strongly suggests that the relevant domain for the scalar presupposition in question should be the speaker's knowledge, or epistemic state, not the common ground.[4] That being the case, we propose that the relevant presupposition for *only* should be a scalar presupposition modalized to the speaker's epistemic state; for instance, in the case of (1), it is something like 'The speaker knows that Mary is the most likely to be able to sing, or Mary is the person who the speaker is most certain can sing'. In order to present a formal definition of the presupposition in question, we first go over a framework in which the current analysis will be couched, actually, an existing analysis of *only* proposed by van Rooij and Schulz (2007).

4.2 Van Rooij and Schulz's Semantic Analysis of Only

van Rooij and Schulz (2007) proposed the following as the semantic meaning of *only*:

(12) **Definition** (*The semantic meaning of* '*only*')
Let ψ be a sentence of the form 'only ϕ where F is the semantic meaning of the focus in ϕ and B the semantic meaning of its background. We define the semantic meaning [only]($\langle$F, B$\rangle$) of ψ as the following proposition:

$$[only](\langle F, B\rangle) =$$
$$\{w \in W: \exists v \in W[F(v)(B(v)) \,\&\, \neg\exists u \in W[F(u)(B(u)) \,\&\, u <_B v] \,\&\, w \leq_B v]\}.$$

They adopted the so-called *structured-meaning* approach to focus (von Stechow 1982; Jacobs 1983; Krifka 1991). The structured-meaning approach assumes that, given a phrase with a focused constituent, focus induces a partitioned meaning consisting of the ordinary meaning of the focused constituent and the lambda abstract of the ordinary meaning of the phrase with respect to the variable or the trace left by the focused constituent, denoted $\langle$F, B$\rangle$, where F and B are referred to as the focus and the background meaning, respectively. Within this framework, a focus-sensitive operator like *only* takes as its argument, the structured meaning of the phrase within its scope. For the sake of simplicity, van Rooij and Schulz took (every instance of) *only* to be a sentential operator; thus, the argument is a structured proposition, which we follow here. Now, the semantic representation of, e.g., (1) is something like the following:

(13) $[only](\langle \lambda P.P(m), \lambda x.\text{can-sing}'(x)\rangle)$.

[4] A reviewer questioned the validity of the claimed difference in naturalness of continuity between (9) and (11). I have checked the data with three native speakers of English and they all have attested to the observation that (9) is perfectly felicitous while (11) is infelicitous being inconsistent.

Next, the explanation of '$\leq_B$' is in order. It is a partial order over the set of possible worlds of a model M with respect to the extension of the background meaning, B[5]:

(14) **Definition of $\leq_B$**:

For any two possible worlds in W,v and w, v $\leq_B$ w iff v is exactly like w except that the extension of B in v is a subset of that in w; i.e., B(v) $\subseteq$ B(w).

Now that '⟨F, B⟩'and '$\leq_B$' have been explained, let us see what the formula means. When *only* has within its scope a sentence, ϕ, whose structured proposition is '⟨F, B⟩', '[only](⟨F, B⟩)' is true in such possible worlds that ϕ, i.e., F(B), is true in them and furthermore, they are minimal with respect to '$\leq_B$' or that they are smaller than the minimal worlds with respect to '$\leq_B$'. Applied to, for instance, (1) whose semantic representation is (13), (1) is specified to be true in a possible world if and only if the extension of 'λx.*can-sing*′(x)', i.e., the set of (relevant) people who can sing in the world is {Mary} or ∅.

What is to be noted here is that from the semantic meaning of *only* in (12), the prejacent of neither a positive nor a negative *only* sentence is entailed. In the case of a positive *only* sentence, it is obvious from the fact that the sentence is true in possible worlds where the extension of the background meaning is the empty set. In the case of a negative *only* sentence, the denotation is the complement of the set in (12), which contains possible worlds in which the prejacent is not true but some "alternatives" are true. For instance, the denotation of "Not only Mary can sing" contains possible worlds in which Bill and Sue, but not Mary, can sing. From the denotation, it is not entailed that Mary can sing.

The question is how van Rooij and Schulz have treated the prejacent of the *only* sentence. They analyzed the prejacent of a positive *only* sentence as a conversational implicature, but in a different way than Ippolito. As for the case of a negative *only* sentence, they were silent about it, or it seemed that they were not aware of the asymmetry problem in the first place. In the next subsection, we will review van Rooij and Schulz's pragmatic analysis of the prejacent.

4.3 Van Rooij and Schulz's Pragmatic Analysis of Only

Van Rooij and Schulz proposed to analyze the prejacent of a(n) (positive) *only* sentence to be a conversational implicature; specifically, an instance of exhaustive interpretation. It is well known that the default interpretation of (10A) as an answer to (10Q) is 'Mary and nobody else can sing'—the exhaustive reading. Groenendijk (1984) attributed the exhaustiveness to the semantic meaning of an interrogative sentence, claiming the possible answers to a question should themselves be exhaustive-

[5] Van Rooij and Schulz were not the first to introduce this ordering relation among possible worlds; for example, Yabushita (1993, 2003) introduced it in the analyses of the exhaustive interpretation of answers, especially, of the disjunctive form and the multiple-sentence form.

reading propositions. However, as has been commonly observed, many instances of answer sentences are not interpreted exhaustively, being accompanied by expressions defying the exhaustive interpretation, like "at least" and "for example"; thus, being construed with the so-called "mention-some" interpretation. The exhaustive reading of other instances is canceled by the utterances that follow them. So it is now commonly assumed that exhaustive interpretation is a pragmatic phenomenon; specifically, a conversational implicature. As to how it should be analyzed—although it had been widely assumed to be analyzed in terms of Gricean maxims, especially the maxim of quantity—van Rooij and Schulz (2006, 2007) were among the first to make a formally precise proposal.

We subscribe to van Rooij and Schulz's analysis of the prejacent of a positive *only* sentence as a conversational implicature based on exhaustive interpretation. Van Rooij and Schulz take the exhaustive interpretation to be a consequence of assuming the speaker (i) to comply with the maxim of quality and the first sub-clause of the maxim of quantity and furthermore, (ii) to be maximally competent (knowledgeable, or informed) on the issue/matter/question under discussion. Before reviewing their analysis of the case under consideration, let us illustrate their approach by going over the exhaustive interpretation of (15A) as an answer to (15Q).

(15) Q: Who will smoke?
A: $[\text{John}]_F$ (will).

(16) a. Only $[\text{John}]_F$ will smoke.
b. Not only $[\text{John}]_F$ will smoke.

In the environment of question (15Q), the utterance of (15A) will be interpreted as meaning that, of the relevant propositions of the form "X will smoke", "John will smoke" is all the speaker knows to be true, according to assumption (i) above. However, in the speaker's knowledge state characterized as above, it is possible for other "relevant" people other than John to smoke. It is reasonable to further assume that the answerer knows of all the relevant people whether they smoke or not; this is consistent with assumption (ii), which in effect results in concluding that by default, the speaker knows that all the rest of the people will not smoke. Finally, by the veridicality of knowledge, it is concluded that John and nobody else will smoke.[6]

Now that van Rooij and Schulz's mechanism of exhaustive interpretation has been illustrated using (15) as an example, we are ready to review their analysis of the prejacent of a positive *only* sentence with (16a) as an example. In fact, van Rooij and Schulz analyzed the prejacent of an *only* sentence as a conversational implicature; specifically, as an instance of exhaustive interpretation. In the case of (15A) as an answer to (15Q), the exhaustive interpretation of (15A) was calculated as to the speaker's knowledge with regards to the relevant sentences of the form "X will smoke". The relevant sentences were determined based on the focus-background meaning of (15A), i.e., $\langle \lambda P.P(j), \lambda x.\text{will-smoke}'(x)\rangle$; the relevant sentences result

[6] This is just an informal synopsis of van Rooij and Schulz's analysis. For the full version of it, see van Rooij and Schulz (2004, 2006, 2007).

from applying the background meaning to (the name of) each relevant individual. As (16a) has the same focus-background meaning, i.e., [*only*] (⟨λP.P(j), λx.will-smoke′ (x)⟩), it is expected that the exhaustive interpretation of (16a) will be calculated with regards to the sentences of the form “X will smoke” as well. However, when the background-predicate occurs in a downward-entailing context as is the case with *only* sentences, which is evidenced in (17), the exhaustive interpretation is calculated with respect to the complement of the background meaning, i.e., ‘λx. ¬will-smoke′(x)’, following Stechow and Zimmermann (1984) and van Rooij and Schulz (2004), which in turn means that we are concerned with the speaker’s knowledge as to the truth of the sentences of the form “X will not smoke” this time.

(17) a. Only $[\text{John}]_F$ has *any money* left.
b. *Only $[\text{John}]_F$ has *some money* left.

With the sentences relevant to the calculation of the exhaustive interpretation for (16a) being fixed, let us proceed. The result of applying assumption (i) to (16a)—whose semantic meaning is that everybody except for John will not smoke—is that the speaker is considered to know the semantic meaning is the case, but no more about the truth of the relevant sentences. This means that, in her knowledge at this stage, it is only possible that John will smoke. Assumption (ii), however, strengthens the inference to that the speaker knows the people other than John are the all and the only elements of the extension of ‘λx.¬will-smoke′ (x)’, which means that John is an (sole) element of the extension of ‘λx.will-smoke′ (x)’, or simply John will smoke. Furthermore, van Rooij and Schulz stipulated that assumption (ii) is highly context-dependent and therefore easily cancelable, while assumption (i) in terms of Gricean maxims is much more robust and therefore hardly cancelable. It should be obvious that in the above pragmatic reasoning of the truth of the prejacent of a positive *only* sentence, the truth of the prejacent will be canceled as assumption (ii) is canceled, which accounts for the cancelability of the prejacent of a positive *only* sentence.

4.4 Comparison with Ippolito’s Analysis

We have seen that from the semantic meaning for *only* proposed by van Rooij and Schulz (2007) and the pragmatic mechanism of exhaustive interpretation as a conversational implicature proposed by van Rooij and Schulz (2006), the truth of the prejacent of a positive *only* sentence can in fact be analyzed as a conversational implicature. Note that van Rooij and Schulz’s analysis captures the prejacent of a(n) (positive) *only* sentence as a conversational implicature as Ippolito’s does; however, they are different in that the former is based on exhaustive interpretation while the latter is based on scalar implicature. From which there are consequent differences between them. That is, in relation to Ippolito’s scalar implicature-based analysis, we argued that as it stands, there is nothing to prevent it from predicting the negation of the prejacent as a conversational implicature as well as the prejacent, which is due to the indeterminacy of selecting the competing stronger statement. On the other hand, van Rooij and Schulz's exhaustive interpretation-based analysis is exempt from the problem, for it deterministically calculates the exact extension of the complement of the predicate phrase. Another difference is manifested in the treatment of "disjunctive" *only* sentences like the following:

(18) Only Mary or Sue voted for John.

In fact, (18) is an example Ippolito (2008): fn. 28 mentioned, but simply left as problematic without going into the details. Examples like (18) are problematic to Ippolito's analysis of *only*. That is, her proposed definition of the truth conditions for an *only* sentence, which says that of the relevant alternative propositions to the prejacent, no alternatives that are not entailed by the prejacent are true (Ippolito 2008: (51)), will predict clearly wrong truth conditions for (18). Despite the question of exactly what the relevant alternative propositions to the prejacent are, it is reasonable to assume that propositions of the form "X voted for John", where X is the name of a relevant individual including Mary and Sue, are among the relevant alternative propositions. Given that and the proposed definition of the truth conditions, the truth conditions predicted for (18) will be that nobody voted for John, which is clearly wrong. With no stronger competing proposition around, there will be no calculation of a scalar implicature fired, either. On the other hand, van Rooij and Schulz (2007) semantic and pragmatic analysis of *only* illustrated above predicts a much more sensible interpretation for (18). First, the truth conditions, or the proposition specified for (18) by (12) is the set of possible worlds such that nobody voted for John in them, Mary was the sole voter for John in them, or Sue was the sole voter for John in them; in other words, the proposition that nobody or at most one of Mary and Sue voted for John. As for the conversational implicature, the speaker is taken to know for sure that everybody except for Mary and Sue did not vote for John by virtue of assumption (i), namely, the maxims of quality and the first sub-clause of the maxim of quantity. Can the inference be strengthened to that the speaker knows that Mary and Sue were the only voters for John by virtue of assumption (ii), namely, the speaker's maximal competence? The answer is no. In the speaker's epistemic state characterized by the predicted truth conditions, it is not possible for both Mary and Sue to have voted for John; thus, the conversational implicature predicted by van Rooij and Schulz's analysis is that either Mary or Sue not both voted for John and the rest of the people didn't. That in fact is an intuitively correct interpreation of (18).

What we would like to know now is what van Rooij and Schulz's analysis can say about the truth of the prejacent of a negative *only* sentence. As a matter of fact, the authors commented in passing that the prejacent of a negative *only* sentence is typically inferred as in the case of its positive counterpart; they might have assumed that the negative case was also an instance of conversational implicature and would be as susceptible to the pragmatic reasoning as was the positive one. However, as we have ascertained with Ippolito, there is an asymmetry between the two cases: The prejacent of a positive *only* sentence is cancelable, while that of a negative *only* sentence is not, which indicates that the two are different animals. In the following, we will propose to revise van Rooij and Schulz's analysis of *only* sentences by adding the modal scalar presupposition proposed in Sect. 4.1. It will be seen that in the revised analysis, the prejacent of a negative *only* sentence is entailed from the result of applying the assumption of (i) alone to the utterance of the negative *only* sentence in conjunction with the modal scalar-presupposition.

4.5 Current Analysis: Semantic Meaning of Only Retrofitted with Modal Scalar-Presupposition

First, we would like to propose a novel analysis of the semantic meaning of *only,* which is a revised version of van Rooij and Schulz (2007) as presented in (12) retrofitted with the modal scalar-presupposition:

(19) **Definition** (*Interpretation of ONLY*($\langle F, B\rangle$))

- (presupposition)
  $[\![\text{ONLY}(\langle \text{F}, \text{B}\rangle)]\!]^{\text{M,w}}$ is defined only if
  $\forall \text{v} \in \text{W:wRv}[\exists \text{x}[[\![\text{B}]\!]^{\text{M,v}}(\text{x}) = 1] \rightarrow [\![\text{F(B)}]\!]^{\text{M,v}} = 1]$, and
- (propositional content)
  $[\![\text{ONLY}(\langle \text{F}, \text{B}\rangle)]\!]^{\text{M}} =$
  $\{\text{w} \in \text{W}: \exists \text{v} \in \text{W}[[\![\text{F(B)}]\!]^{\text{M,v}} = 1 \ \& \ \neg\exists \text{u} \in \text{W}[[\![\text{F(B)}]\!]^{\text{M,u}} = 1 \ \& \ \text{u} <_{\text{B}} \text{v}] \ \& \ \text{w} \leq_{\text{B}} \text{v}]\}$.

In (19), ‘ONLY’ is a logical translation of *only*, ‘$\langle$F, B$\rangle$’is as specified as in (12), and ‘R’ is the epistemic accessibility relation for the speaker. The presupposition says that the speaker knows that if the extension of ‘B’ is not empty, then ‘F(B)’ is true, which, we contend, expresses the speaker’s certainty of F’s being most likely to B among the relevant alternatives to F. The presupposition in question for (16a) will be that the speaker knows that if someone smokes, then John will smoke. The propositional content is exactly the same as in (12).

Let us see what consequences the current definition of the semantic meaning of *only* will have for the status of the prejacent of a negative *only* sentence. In the case of (16b), the propositional content is the complement of that for (16a), i.e., the set of possible worlds in which there is at least one person other than John who will smoke, with or without John’s smoking. Here, let us assume that the Gricean maxim of quality is at work; that is, the speaker knows the proposition is true, from which, along with the presupposition that the speaker knows that if someone will smoke, John will smoke, follows that the speaker knows that John will smoke. Finally, by the veridicality of knowledge, it follows that John will smoke.

Note that in the above derivation of the prejacent of a negative *only* sentence, only the maxim of quality was assumed, in contrast to the case of that of a positive *only* sentence, in which the speaker’s maximal competence was assumed in addition to the maxim of quality and the first subclause of the maxim of quantity. Remember that van Rooij and Schulz attributed the cancelability of the prejecent of a positive *only* sentence to the stipulation that the assumption of the speaker's maximal competence is highly context-dependent and therefore is easy to revoke, while that of the Gricean maxims is much more robust and therefore is hard to revoke. Consequently, from the current analysis of the prejacent of a negative *only* sentence, in which only the maxim of quality is resorted to as a pragmatic principle, it is predicted that the prejacent of a negative *only* sentence is much harder[7] to cancel than that of a positive one. The prediction is in fact borne out by the asymmetry we have assumed throughout this paper. In Appendix is a table summarizing the features of the respective analyses.

5 Conclusion

To the myriad semantic and pragmatic properties of *only* (sentences), Ippolito (2006, 2008) added yet another one to be accounted for, i.e., the asymmetry between positive and negative *only* sentences in the cancelability of the prejacent. Ippolito convincingly argued that none of the previous analyses of *only* was adequate to treat the phenomenon, and she proposed an analysis of the semantic meaning of *only* in which a scalar presupposition was alleged to play an crucial role in explaining the asymmetry in question. However, we have found that her account of the prejacent of a positive *only* sentence as a conversational implicature based on scalar implicature was questionable at best. Furthermore, the scalar presupposition has turned out to be empirically inadequate, as it makes false predictions. Despite the apparent inadequacy, there is some evidence that a scalar presupposition is involved in the semantics of *only*. That being the case, we have argued that the scalar presupposition in question should be reformulated as one modalized to the speaker's knowledge. Accordingly we proposed a modal version of the scalar presupposition, which has, in fact, turned out to be empirically motivated. As an alternative framework to Ippolito's basic theory of *only* that has obvious problems with the derivation of conversational implicatures and the truth conditions of some type of *only* sentences, we proposed to adopt the semantic and pragmatic theory of *only* developed by van Rooij and Schulz (2007). However, the theory was shown to be inadequate in the treatment of the prejacent as a conversational implicature, not being able to account for the very asymmetry problem. Then, we have proposed to retrofit van Rooij and Shulz's theory with the modal scalar presupposition. The resulting revised theory has proved to be very effective in accounting for the asymmetry.

Acknowledgments I would like to thank the audience of LENLS 4 for helpful questions and comments. And my sincere gratitude goes to two anonymous reviewers of the present volume and the series editor for their critical questions, comments, and suggestions, which tremendously helped me to identify inadequacies in the argument and remedy them for the thesis to be presented more cogently. However, all remaining errors and shortcomings are mine. Last, but not least, I am very grateful to Kathy Hoyt for her thorough proofreading.

[7] In fact, the cancellation is impossible as long as the speaker is truthful.

Appendix

Table 1 Summary of the features of the respective analyses of *Only* sentences

	Ippolito (2006, 2008)	van Rooij and Schulz (2007)	Current analysis
Truth conditions of an *only* sentence	Entailment-based: It specifies the truth conditions as that all the relevant alternative propositions that are not entailed by the prejacent are false	Based on the idea of minimal worlds with respect to the extension of the predicate phrase (see (12))	
	Problem: It predicts clearly wrong truth conditions for *only* sentences of disjunctive form like (18)	No problem	
Treatment of the pejacent	Conversational implicature based on scalar implicature	Conversational implicature based on exhaustive interpretation	
	Problem: It predicts not only the prejacent but also its negation as conversational implicatures	Problem: It cannot deal with the asymmetry in cancelability of the prejacent between a positive *only* sentence and its corresponding negative one	No problem: It is equipped with the modal scalar presupposition

References

Atlas, J. D. (1991). Topic/comment, presupposition, logical form and focus stress implicatures: The case of focal particles only and also. *Journal of Semantics*, *8*, 127–147.

Atlas, J. D. (1993). The importance of being ‘only’: Testing the neo-Gricean versus neo-entailment paradigms. *Journal of Semantics*, *10*, 301–318.

Atlas, J. D. (1996). ‘Only’ noun phrases, pseudo-negative generalized quantifiers, negative polarity items, and monotonicity. *Journal of Semantics*, *13*, 265–328.

Gazdar, G. (1979). *Pragmatics: implicature, presupposition, and logical form*. New York: Academic Press.

Geurts, B., & van der Sandt, R. (2004). Interpreting focus. *Theoretical Linguistics*, *30*, 1–44.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics volume 3: Speech acts* (pp. 41–58). New York: Academic Press.

Groenendijk, J., & Stokhof, M. (1984). Studies on the semantics of questions and the pragmatics of answers. *Dissertation*, University of Amsterdam.

Hirschberg, J. B. (1985). A theory of scalar implicature. *Dissertation*, University of Pennsylvania.

Horn, L. (1969). A presuppositional analysis of 'only' and 'even'. In *Papers from the Fifth Regional Meeting of the Chicago Linguistics Society* (pp. 98–107). CLS. Chicago, IL.

Horn, L. (1996). Exclusive company: 'Only' and the dynamics of vertical inference. *Journal of Semantics*, *13*, 1–40.

Ippolito, M. (2006). Remarks on only. [Electronic version.] *Proceedings of Semantics and Linguistic Theory XIV* (SALT 16). Available at:http://research.nii.ac.jp/salt16/proceedings.html

Ippolito, M. (2008). On the meaning of only. *Journal of Semantics 25*, 45–91. Advance Access. Publication, 12, 2007.

Jacobs, J. (1983). *Fokus und Skalen. Zur Syntax und Semantik von Gradpartikeln im Deutschen*. Tübingen:Niemeyer

Kartunnen, L., & Peters, S. (1979). Conventional implicature. In Ch. K. Oh & P. A. Dinneen (Eds.), *Syntax and semantics, volume 11: Presuppositions* (pp. 1–56). New York: Academic Press.

Krasikova, S., & Zhechev, V. (2005). You only need a scalar only. In: C. Ebert & C. Endriss (Eds.) *Proceedings of the sinn und Bedeutung 10* (pp. 199–210). Berlin: Zentrum für allgemeine Sprachwissenschaft (ZAS).

Krifka, M. (1991). A compositional semantics for multiple focus constructions. In S. K. Moore & A. Z. Wyner (Eds.), *Proceedings of the first conference on semantics and linguistic theory (SALT 1)* (pp. 127–158). Ithaca: Cornell.

Levinson, S. R. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

McCawley, J. (1981) *Everything that linguists have always wanted to kow about logic but were ashamed to ask*. Chicago:University of Chicago Press.

Rooth, M. (1985). Association with focus. *Ph.D Dissertation*, University of Massachusetts, Amherst.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*, 75–116.

van Rooij, R., & Schulz, K. (2004). Exhaustive interpretation of complex sentences. *Journal of Logic, Language, and Information*, *13*, 491–519.

van Rooij, R., & Schulz, K. (2006). Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy*, *29*, 205–250.

van Rooij, R., & Schulz, K. (2007). Only: Meaning and implicattures. In M. Aloni, A. Butler & P. Dekker (Eds.), *Questions in dynamic semantics* (pp. 193–224). Oxford: Elsevier.

von Fintel, K., & Iatridou, S. (2005) Anatomy of a modal. In J. Gajewski et al. (Eds.), *New work on modality* (pp. 63–121). MIT Working Papers in Linguistics 51. Cambridge, Mass.: MIT, Department of Linguistics and Philosophy, MITWPL.

von Stechow, A. (1982). Structured Propositions, Arbeitspapier 59 des Sonderforschungsberichs 99, Universität Konstanz.

von Stechow, A., & Zimmermann, T. E. (1984). Term answers and contextual change. *Linguistics*, *22*, 3–40.

Yabushita, K. (1993). *Exhaustiveness as an operation on information states*. Paper presented at the Conference on Logic and Linguistics, The Ohio State University, July 31, 1993.

Yabushita, K. (2003). Topicality in the semantics and pragmatics of questions and answers: Evidence for a file-like structure of information states. In I. Kruijff-Korbayová & C. Kosny (Eds.), *Proceedings of the 7th workshop on the semantics and pragmatics of dialogue (DiaBruck 2003)* (pp. 147–154). Saarbrücken: Universität des Saarlandes.

Floating Quantifiers in Japanese as Adverbial Anaphora

Kei Yoshimoto and Masahiro Kobayashi

Abstract A surface-based analysis of so-called floating quantifiers (FQs) in Japanese is proposed based on Head-Driven Phrase Structure Grammar and Minimal Recursion Semantics. We hypothesize that sentences with FQs, as other sentences, are processed incrementally from left to right and that an FQ has an independent semantic representation as NP that is anaphorically related to its antecedent. On these assumptions, we account for the asymmetry between the subject and object in terms of the quantification by a non-adjacent FQ. We also address a kind of FQ construction that has hitherto escaped the researchers' attention, that with a whole-part relationship between the FQ and its host, and show that it is explained by the framework we bring forward.

Keywords Floating quantifier · Japanese · Head-driven phrase structure grammar · Minimal recursion semantics · Information structure

1 Introduction

The so-called floating quantifier in Japanese has attracted a great number of researchers with its Janus-like nature: while it is syntactically an adverbial phrase, it semantically *quantifies* a remote nominal. Furthermore, its status is complicated considerably by the fact that there exists another means of representing the quantity of nominals, the prenominal quantifier, in one and the same language.

K. Yoshimoto (✉)
Center for the Advancement of Higher Education and Graduate School of International Cultural Studies, Tohoku University, Sendai, Japan
e-mail: kei@compling.jp

M. Kobayashi
Education Center, Tottori University, Tottori, Japan

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*,
Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_16,

Most studies on this issue have adopted syntactic approaches using transformation (see among others Miyagawa 1989). As extensively discussed by Takami (1998a, b, c), however, solutions based solely on syntax are confronted with the observation that the speakers' judgement on the grammaticality of sentences with the same syntactic structure differs noticeably depending on the context. Furthermore, no investigation has ever been made of a construction in which a floating quantifier (FQ hereafter) does not simply quantify its host, but stands in a whole-part relationship to it, which urges the necessity of a more elaborate semantic study of the FQ phenomenon.

In this chapter, we propose a surface-based analysis of FQs in Japanese as a kind of verbal adjunct, not as a form moved from its original prenominal location as they have been dealt with in the transformational tradition. We thus suggest giving it a semantic interpretation independent of that of its host. Strictly speaking, the FQ is not even a quantifier according to our proposal: it has an independent meaning as a quantified NP, standing in an anaphoric relation to its host NP.[1]

Based on this approach, we provide an account from a new perspective of the long-debated asymmetry between the subject and object in terms of quantification with another case-marked NP intervening between an FQ and its host. We also show that the above-mentioned usage of FQs, a construction in which an FQ and its host stand in a whole-part relationship, is given an explanation consistent with the majority of FQs. Furthermore, our investigation leads to an answer to a rarely raised, but nonetheless crucial question: *Since prenominal quantifiers are already available, what purpose then do FQs serve?*

We set up an incremental, left-to-right sentence processing model that takes into consideration the information structure of the sentence, as formalized on the basis of Head-Driven Phrase Structure Grammar (HPSG; Pollard and Sag 1994) the semantics of which is represented using Minimal Recursion Semantics (MRS; Copestake et al. 1999).

2 FQs in Japanese

Japanese has two means of expressing the quantity of a nominal: one is a prenominal quantifier attached to the quantified noun by an adnominal postposition *no* as in example (1a), and the other is an FQ that is most often taken to have been *displaced* from its host NP to a position closer to the predicate, as exemplified by (1b).[2]

[1] Nevertheless, we continue to use both the terms *floating quantifier* and its *host*, observing the conventions to enhance readability. They should not be understood as precisely reflecting our approach toward the subject of our study.

[2] Throughout this paper, an FQ in an example sentence is given in bold, and its (intended) host NP in italics.

(1) a. **San-nin** **no** *gakusei* ga ki-ta.
three-CLASS$_{person}$ ADNOM student SBJ come-PAST
'Three students came.'

b. *Gakusei* *ga* **san-nin** ki-ta.
student SBJ three-CLASS$_{person}$ come-PAST
'Three students came.'

The quantifier consists of a numeral and a classifier chosen in accordance with the semantic characteristics of the quantified noun. For example, -*nin*/-*ri* is used with humans, -*satsu* with books, magazines, or notebooks, and -*tō* with large animals. There is also a semantically unmarked classifier for inanimate objects, -*ko*. See Miyagawa (1989, p. 20) for details.

Floating is allowed only from subject or object NPs. Between the two syntactic cases, however, asymmetry has been observed in terms of the hosting of an FQ (Kuroda 1983; Saito 1985; Miyagawa 1989). Whereas an FQ *floated* from the subject of the sentence must be adjacent to its host (see (2a) below) thus making a sentence like (2b) in which the object intervenes between the subject and FQ ungrammatical, the host object and the FQ can occur either next to each other or with the subject appearing between them, as in (2c) and (2d).

(2) a. *Gakusei* *ga* **go-nin** hon wo kat-ta.
student SBJ five-CLASS$_{person}$ book OBJ buy-PAST
'Five students bought books.'

b. **Gakusei* *ga* hon wo **go-nin** kat-ta.
student SBJ book OBJ five-CLASS$_{person}$ buy-PAST
'Five students bought books.'

c. Gakusei ga *hon* *wo* **san-satsu** kat-ta.
student SBJ book OBJ three-CLASS$_{book}$ buy-PAST
'Students bought three books.'

d. *Hon* *wo* gakusei ga **san-satsu** kat-ta.
book OBJ student SBJ three-CLASS$_{book}$ buy-PAST
'Students bought three books.'

This subject-object asymmetry in the FQ position has been heatedly discussed in relation to configurationality in Japanese sentences. Note also that the same holds true for sentences in which the FQ is dislocated to the sentence-initial position:

(3) a. ***Go-nin** hon wo *gakusei* *ga* kat-ta.
five-CLASS$_{person}$ book OBJ student SBJ buy-PAST
'Five students bought books.'

b. **San-satsu** gakusei ga *hon* *wo* kat-ta.
three-CLASS$_{book}$ student SBJ book OBJ buy-PAST
'Students bought three books.'

3 Previous Studies

3.1 Syntactic Approaches

The earlier mainstream studies on FQs in Japanese have obtained the construction by moving an FQ from the prenominal location quantifying a nominal, as represented by Kuno (1978). These studies have argued that the host of the FQ must be either a syntactic subject or object.

Shibatani (1977) offers a counterargument to this widely accepted account, emphasizing the role the surface case marking of the antecedent NP plays. He relies on two tests for subjecthood in terms of reflexivization and honorification. In Japanese, not only the nominative, but also the dative case can mark the syntactic subject in reflexivized and subject-honorific sentences. See the examples taken from Shibatani (1977):

(4) a. *Sensei* *ni* (*wa*) jibun ga wakara-nai.
teacher DAT TOP self NOM understand-NEG
'The teacher does not understand himself.'

b. *Sensei* *ni* (*wa*) eigo ga o-wakari-ni-naru.
teacher DAT TOP English NOM understand-HON
'The teacher understands English.'

Despite the fact that *sensei ni* (*wa*) in (4a,b) is syntactically assigned the grammatical function of subject, floating an FQ from the same NP is illicit as in (5a) below. On the other hand, sentence (5b) whose subject is marked by the nominative case marker *ga* is grammatical.

(5) a. ∗*Korera-no* *kodomo-tachi* *ni* **san-nin** eigo
these child-PLUR NOM three-CLASS$_{person}$ English
ga wakaru.
NOM understand
'These three children understand English.'

b. *Korera-no* *kodomo-tachi* *ga* **san-nin** eigo
these child-PLUR NOM three-CLASS$_{person}$ English
ga wakaru.
NOM understand
'These three children understand English.'

Upon this observation, Shibatani bases his claim that it is the morphological case marking, the nominative (*ga*) and accusative (*wo*), that licenses quantifier floating.

However, a closer inspection reveals that even the case marking by *ni* triggers floating:

(6) Tom ga *yūmeina kyōju ni* **san-nin**
NAME NOM famous professor DAT three-CLASS$_{person}$
at-ta.
see-PAST
'Tom met three famous professors.'

In this sentence, despite the marking by *ni*, the complement subcategorized for by the verb *at-ta* (see-PAST), *yūmeina kyōju ni* (famous professors-DAT), can be quantified by the FQ. Therefore, Shibatani's surface case-based account of quantifier floating must be abandoned and the reason for the inappropriateness of (5a) should be looked for elsewhere (see Miyagawa (1989) for an argument that the dative subject is a PP). A discussion along these lines is conducted by Inoue (1978), who concludes that only the subcategorized-for NPs, or arguments in distinction from adjuncts, can host FQs. Miyagawa's (1989) theory, a syntactic account that develops Inoue's (1978) ideas, is reviewed later in this subsection and Sect. 3.2.

Yatabe (1990) assumes a hierarchical order among θ-roles, like Jackendoff (1972), as below:

(7) ⟨agent ⟨recipient ⟨instrumental ⟨locative ⟨theme(predicate)⟩⟩⟩⟩⟩

On the basis of this hierarchy and a binary syntactic tree, Yatabe puts forward his hypothesis on quantifier floating.

(8) A floated quantifier can be associated only with the thematically lowest argument slot of the predicate that it combines with.

This explains the appropriateness of the sentences (1b) and (2a, c, d). The list of θ-roles associated with the predicate (verb or VP) combined with the FQ in each sentence is shown in angular brackets below each predicate.[3] It also explains why (2b) is not allowed; whereas the theme is the lowest θ-role in the list associated with the predicate *kat-ta* (buy-PAST) to be combined with the FQ *go-nin*, the FQ's only semantically possible antecedent is the agent.

(1) b. [[*Gakusei ga*] [[**san-nin**] [KI-TA]]]
student SBJ three-CLASS$_{person}$ come-PAST
⟨ag⟩
'Three students came.'

(2) a. [[*Gakusei ga*] [[**go-nin**] [[HON WO] [KAT-TA]]]]
student SBJ five-CLASS$_{person}$ book OBJ buy-PAST
⟨ag⟩
'Five students bought books.'

b. ∗[[*Gakusei ga*] [[*hon wo*] [[**go-nin**] [KAT-TA]]]]
student SBJ book OBJ five-CLASS$_{person}$ buy-PAST
⟨ag, th⟩
'Five students bought books.'

[3] The verb or VP that is combined with the FQ is given in small capitals in these examples.

c.	[[Gakusei	ga]	[[*hon*	*wo*]	[[**san-satsu**]	[KAT-TA]]]]
	student	SBJ	book	OBJ	three-CLASS$_{book}$	buy-PAST
						⟨ag, th⟩

‘Students bought three books.’

d.	[[*Hon*	*wo*]	[[gakusei	ga]	[[**san-satsu**]	[KAT-TA]]]]
	book	OBJ	student	SBJ	three-CLASS$_{book}$	buy-PAST
						⟨ag, th⟩

‘Students bought three books.’

However, Yatabe fails to provide an appropriate interpretation for examples like (9) below, since he predicts that the FQ is only allowed to quantify the NP with the theme role, *hon wo* (book-OBJ); actually, however, this is a grammatical sentence with the FQ *san-nin* floated from the subject NP *gakusei ga* (student-SBJ).

(9)	[[*Hon*	*wo*]	[[gakusei	*ga*]	[[**san-nin**]	[KAT-TA]]]]
	book	OBJ	student	SBJ	three-CLASS$_{person}$	buy-PAST
						⟨ag, th⟩

‘Three Students bought books.’

In fact, in his awareness that his prediction is not borne out by this kind of sentence, Yatabe (1990) claims that the FQ and its host NP together form an NP in sentence (9), while the following sentence, in which the FQ separated from its host by the adverbial phrase is evidently an independent phrase, is an instance of Right Node Raising.

(10)	Hon	wo	*gakusei*	*ga*	KORE MADE NI	**san-nin**	kat-ta.
	book	OBJ	student	SBJ	so far	three-CLASS$_{person}$	buy-PAST

‘So far, three students have bought books.’

Yatabe’s explication is not adopted in this paper, however. First of all, the examples with an adjacent host and FQ hitherto cited, (1b), (2a), (2c), (5b), and (6), do not change their grammaticality by inserting an adverbial phrase. Given this evidence, the unity of the host and FQ as a single NP in (9) seems most unlikely. Furthermore, the application of Right Node Raising to (10) is ad hoc, lacking a clear criterion to limit its application.

Miyagawa’s (1989) account of FQs in Japanese is given on the basis of syntactic configuration and benefits from extensive data coverage. He proposes that, in order for an FQ to quantify its host NP, both must c-command each other. The definition of c-command is given below, which Miyagawa attributes to Reinhart (1979).[4]

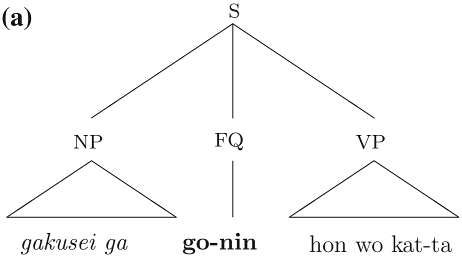
(11) A c-commands B if neither A nor B dominates the other and the first branching node dominating A also dominates B.

[4] In fact, Reinhart (1979) does not mention how c-command is defined. (11) is an incorrect citation of Langacker’s (1966) definition of ‘command’ rather than ‘c-command’. Reinhart (1981, 1983), for instance, provides definitions of the latter term in more widely accepted forms. However, we follow Miyagawa’s conceptualization throughout this paper.

This ‘mutual c-command requirement’ explains the difference in grammaticality between (2a) and (2b): while in the former sentence both the subject NP *gakusei ga* (student-SBJ) and the FQ *go-nin* c-command each other, in (2b) the subject c-commands the FQ that is embedded in a VP but is not c-commanded by the FQ.

(2)

(a)



The acceptability of (2c) and (2d) is also accounted for by the c-command relationship between the FQ and its host. Specifically in (2d), the FQ *san-satsu* does not directly c-command its host *hon wo*, but its trace in the position from which the host moved, thus scrambling the word order.

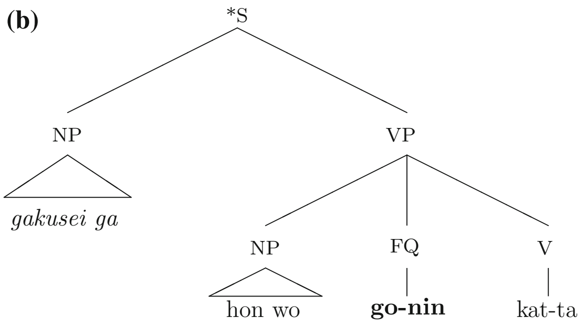
(2) c. Gakusei ga [$_{VP}$ [$_{NP}$ *hon* *wo*] [$_{FQ}$ **san-satsu**] [$_{V}$ kat-ta]]
student SBJ book OBJ three-CLASS$_{book}$ buy-PAST
‘Students bought three books.’

d. *Hon*$_i$ *wo* gakusei ga [$_{VP}$ t_i [$_{FQ}$ **san-satsu**] [$_{V}$ kat-ta]]
book OBJ *student* SBJ TRACE three-CLASS$_{book}$ buy-PAST
‘Students bought three books.’

Also explained within the same framework is why only the subject or object of the sentence can host an FQ: Miyagawa obtains subject and object NPs by cliticizing case markers *ga* and *wo*, while other case particles combine with an NP to form a PP. Thus, for example, in the case of *uchi kara* (house-SOURCE), the postposition *kara* embeds the NP *uchi*:

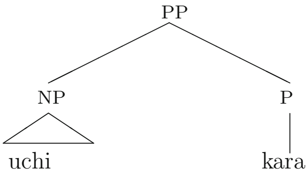
(12)

(b)



In distinction from subject and object NPs, this embedding of an NP prohibits an NP from c-commanding the FQ, since there exists a node PP between the NP and FQ:

(13)

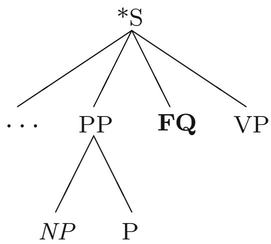


Miyagawa's mutual c-command requirement should not hold between the FQ and the subject in sentences (14a, b) below, since the FQ occurs within a VP, as illustrated in the tree diagram (14c); nevertheless, these are accepted sentences.

(14) a. [[$_{NP}$ *Gakusei* *ga*] [$_{VP}$ [$_{PP}$ ofisu ni] [$_{FQ}$ **futa-ri**]
student SBJ office GOAL two-CLASS$_{person}$
[$_V$ *ki-ta*]]]
come-PAST
'Two students came to the office.'

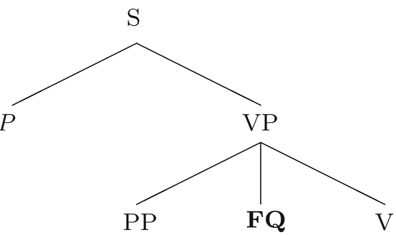
b. [[$_{NP}$ *Otoko* *ga*] [$_{VP}$ [$_{PP}$ bā ni] [$_{FQ}$ **futa-ri**]
man SBJ bar GOAL two-CLASS$_{person}$
[$_V$ hait-ta]]]
enter-PAST
'Two men entered the bar.'

c.



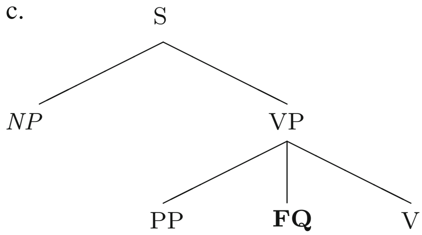
Miyagawa avoids this difficulty by assuming that the subject of unaccusative verbs like those in the above sentences originates in the object position and is moved to the subject position by Move-α. For example, the following is the syntactic structure shared by (14a, b) resulting from the movement:

c.



(14)c'.

c.



Between the trace left by the subject and the FQ, the mutual c-command requirement holds.

By contrast, unergative verbs, which according to Miyagawa do not allow quantifier floating from the subject, do not undergo the movement of the subject and are therefore not affected by a remaining trace:

(15)	?**Gakusei*	*ga*	jibun	no	kane	de	**futa-ri**
	student	SBJ	self	ADNOM	money	with	two-CLASS$_{person}$

denwashi-ta.
telephone-PAST
'Two students telephoned using their own money.'

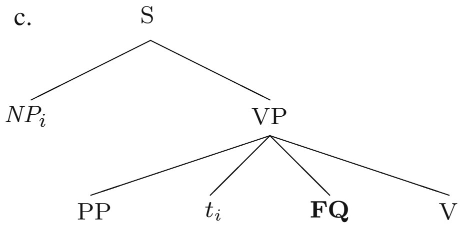
A syntactic approach from an alternative perspective is made by Fukushima (1991, 1993), who proposes a surface-based analysis and semantic interpretation of FQs in Japanese on the basis of HPSG and Montague Semantics. In a similar fashion to Miyagawa (1989), he develops a theory based on syntactic configuration. In his theory, therefore, the same kind of difficulty as seen in Miyagawa is inevitable, in that it is not possible to account for the context-dependent difference in grammaticality of FQ sentences.

3.2 An Information Structure-Based Approach

An entirely new perspective on the issue based on the information structure of sentences is provided by Takami (1998a, b, c). He begins his discussions by criticizing Miyagawa's (1989) mutual c-command requirement, which can be applied to many of the other syntactic approaches. As illustrated in the previous subsection, Miyagawa explains the ungrammaticality of sentences in which an antecedent subject and an FQ *floated* from it is separated by an object NP, as in sentence (2b), by the lack of c-command by the FQ on the subject NP:

(2) b.

c.



However, Takami points out that there exist sentences with the same syntactic structure as above that are nevertheless acceptable:

(16) A: Kono shinkan-zasshi, urete-masu-ka?
this newly-issued magazine sell well-POL-Q
'Does this newly-issued magazine sell well?'

B: Ē, kesa mo *gakusei-san* *ga* SORE WO
yes this morning also student-HON SBJ DEM- PRON OBJ
go-nin katte-iki -mashi-ta yo.
five-CLASS$_{person}$ buy- POL-PAST MOD- PARTCL
'Yes, this morning five students came to buy it.'

As in example (16B) cited by Takami (1998a), the ⟨NP$_{sbj}$-NP$_{obj}$-FQ⟩ construction in which the FQ quantifies the subject is appropriate if the object NP is a pronoun or another type of definitely marked NP (like *sono zasshi wo* 'that magazine-OBJ') that has already been established in the context.

Takami also gives other acceptable sentences with the same syntactic structure:

(17) a. *Gakusei* *ga* repōto wo **san-nin** DAKE
student SBJ report OBJ three-CLASS$_{person}$ only
teishutsushi-ta.
hand in-PAST
'Only three students handed in a report.'

b. *Gakusei* *ga* boku no jugyō wo tochū-de
student SBJ I ADNOM class OBJ halfway
san-nin MO yame-mashi-ta.
three-CLASS$_{person}$ as many as quit-POL-PAST
'As many as three students quit my class.'

c. *Gakusei* *ga* watashi no hon wo
student SBJ I ADNOM book OBJ
futa-ri SHIKA kawa-nakat-ta.
two-CLASS$_{person}$ only buy-NEG-PAST
'Only two students bought my book.'

d. *Wagako* ga kokyō wo **san-nin**
my children SBJ home town OBJ three-CLASS$_{person}$

TOMO dete-itte- shimat-ta.
all leave COMPL-PAST
'All of my three children left our home town.'

When FQs are suffixed by the so-called adverbial particles (fuku-joshi) such as *dake* (only), *mo* (as many as), *shika* (only), and *tomo* (all), the sentences are also grammatical.

Takami also cites counterexamples to Miyagawa's arugument based on the syntactic distinction between subject and object as complement NPs on the one hand and adjunct PPs on the other.

(18) a. Boku wa *yūmeina gakusha ni* **san-nin**
I TOP famous scholar DAT three-CLASS$_{person}$
at-ta.
meet-PAST
'I met three famous scholars.'

b. *Tarō ga Hanako wo *eigo no sensei ni*
NAME SBJ NAME OBJ English ADNOM teacher DAT
san-nin shōkaishi-ta.
three-CLASS$_{person}$ introduce-PAST
'Taro introduced Hanako to three English teachers.'

Miyagawa attributes, following Inoue (1978), the grammaticality of (18a) to the argument (complement) status of the dative phrase *yūmeina gakusha ni* (famous scholars-DAT). Takami maintains that if this discussion is on the right track, the dative NP *eigo no sensei ni* (English teacher-DAT) in (18b), evidently subcategorized for by the verb *shōkaishi-ta* (introduce-PAST), should be able to host an FQ. In fact, however, sentence (18b) is inappropriate. This implies that what is crucial with the floating of FQs is the grammatical function (subject, direct object, indirect object, etc.) that a noun phrase bears, rather than the NP/PP distinction.

Takami also provides counterevidence to Miyagawa's claim that unergative verbs, unlike unaccusative verbs, do not have a trace left at the object position and consequently do not allow an FQ within a VP:

(19) a. *Gakusei ga* [$_{VP}$ kyōshitsu de **yo-nin**
student SBJ classroom LOC four-CLASS$_{person}$
ABARE-MAWATTE- I-TA]
act violently PROG-PAST
'Four students were acting violently in the classroom.'

b. *Dōryō* ga [$_{VP}$ Yamada-kun no teian ni
colleague SBJ NAME ADNOM *proposal* IOBJ
go-nin SANSEISHI-TA]
five-CLASS$_{person}$ agree-PAST
'Five colleagues of mine agreed to Mr. Yamada's proposal.'

These sentences also suggest that Miyagawa's c-command-based account of the acceptability of quantifier floating is untenable.[5]

On the basis of the above-mentioned criticism of Miyagawa's (1989) syntactic approach, Takami (1998a, b, c) attempts to develop a functional account of the issue. Takami's discussion consists of two main points: the preverbal position as the place for the newest information and parallelism of the FQ construction to topicalization.

First, Takami (1998a, b, c) attempts to found his discussion on Kuno's (1978) hypothesis that in Japanese the constituent immediately before the verb provides the newest information in the sentence. See the following sentences:

(20) a. Gakusei ga *hon* *wo* kat-ta.
student SBJ book OBJ buy-PAST
'The student bought a book.'

b. Gakusei ga *sore/* *sono* *hon* *wo* kat-ta.
student SBJ DEM-PRON DET book OBJ buy-PAST
'A student bought it/the book.'

In (20a), both the subject and object are without any explicit marker for definiteness.[6] Under unmarked conditions, the object NP *hon wo* (book-OBJ), which occurs immediately before the verb, tends to be interpreted as conveying the newest information. Thus the sentence most likely means 'It is a book/books that the student bought.' By contrast, sentence (20b) with a definitely marked object NP is most often construed as giving information on the person that bought the book(s).

On the basis of this observation, Takami goes on to analyze the following sentences with FQs from the information structure perspective:

(21) a. **Gakusei* ga hon wo **yo-nin** kat-ta.
student SBJ book OBJ four-CLASS$_{person}$ buy-PAST
'The student bought a book.'

b. *Gakusei* ga SORE/ SONO HON WO **yo-nin**
student SBJ DEM-PRON DET book OBJ four-CLASS$_{person}$
kat-ta.
buy-PAST
'A student bought it/the book.'

Takami (1998c, p. 100) writes,

[5] Furthermore, Takami quotes examples with unaccusative verbs that have an FQ within the VP but are nonetheless unacceptable. However, since his discussion is based on a subtle interpretation of grammaticality, we do not deal with those sentences in this chapter.

[6] This does not necessarily mean that they stand for indefinite NPs in all contexts. *Bare* NPs in Japanese may be associated with definiteness, or more specifically with specificity, depending on the context. Note that it is hard to translate the delicateness into English concisely. The same holds for plurality.

> In [(21a)], the FQ *yo-nin* placed immediately before the verb is interpreted as the most important information, whereas the object *hon wo*, being an indefinite NP, is also construed as valuable information. Therefore it is impossible to judge whether this sentence is a statement on how many students bought the book or what those students bought. That is why most people feel that it is an unnatural sentence.

On the other hand, according to Takami, in (21b) the object as definite NP is informationally less important and accordingly the FQ is the most important information without any competitor. Thus the sentence is appropriate.

However, it is impossible to reconstruct how these sentences are processed successfully or otherwise based solely on what Takami writes. Most crucial is the lack of reason for the indefinite object NP *hon wo* being interpreted as competing with the FQ for being the most important information, if, as Kuno (1978) argues, the importance depends solely on word order. To compensate for what is missing, we need to hypothesize that the object NP in a construction with a transitive verb, even if not located immediately before the predicate, provides the newest information or focus in the sentence by default. We also propose that an FQ conveys the newest information in the sentence and that a conflict between the object NP and the FQ concerning the focus placement is the cause of unacceptable FQ sentences. This we discuss at length in the following section.

The second, even more vague, argument in Takami's proposal seems to involve the information structure of Japanese sentences. We address this issue in the next section and thereafter within a formal framework, without entering into details of Takami's discussions, which are suggestive but hard to follow on many points.

Lastly, let us point out two types of sentences that are beyond the scope of Takami's explanation.

First, look at the following sentence cited by Takami(1998a, p. 91):

(22) *Nadakō* *no* *seito* *wa*, mai-toshi Tōdai
Nada Highschool ADNOM student TOP every year Tokyo Univ.
wo **hachijū-nin** **ijō** jukensuru.
OBJ eighty-CLASS$_{person}$ more than take an exam
'As for Nada Highschool students, more than eighty of them take an exam for Tokyo University every year.'

One of the difficulties with this sentence is that the host NP *Nadakō no seito wa* (the students of Nada Highschool-TOP) is topicalized and as such definite. Furthermore, the meaning of this sentence is not simply that more than eighty students of the highschool take an entrance exam for Tokyo University every year, but that the group of students has a certain property (e.g., being successfully crammed to win the competition) implied by the fact that more than eighty of them take an exam for the university every year. Therefore, the relationship between the topic NP and the FQ providing new information still remains to be explained.

The second kind of sentence Takami does not cover is the following:

(23) *Mishiranu* *gakusei* *ga* hon wo **go-nin** katte-it-ta.
unfamiliar student SBJ book OBJ five-CLASS$_{person}$ buy-go-PAST
'Five unfamiliar students came to buy books.'

While (2b) is unacceptable, sentence (23) with the same syntactic structure is appropriate, given that its subject is modified by the adjectival word *mishiranu* (unfamiliar), which marks indefiniteness. Since Takami deals only with sentences whose subjects are associated with the unmarked information structure, this example is also beyond the framework of his theory.

On the whole, Takami's approach based on information structure seems potentially to cover more linguistic data than syntactic approaches have ever done. However, since his theory lacks some essential components, an overall reconstruction of the theory is the only way to give proper embodiment to his ideas.

3.3 A Semantic Approach

An alternative solution to the subject-object asymmetry in terms of quantifier floating within the framework of semantics is proposed by Gunji and Hasida (1999). They assume that an FQ ('a measure phrase' or 'MP' in their terminology), whose original function is to modify a predicate rather than a nominal, can exert its influence only on an incremental theme (Dowty 1991).[7] On the basis of this assumption, they count costs in processing sentences with FQs in the following way:

(A) Measurement of events in terms of participants other than the incremental theme, including the agent, is also possible but cost-burdensome.
(B) An adverbially measurable NP intervening in an NP-MP pair tends to be associated with the MP and reduces acceptability.

Gunji and Hasida (1999) maintains that (A) and (B) raise costs of processing. If both of them hold in one and the same sentence with an FQ, they cause fatal unacceptability. A typical case is sentence (2b), in which an object NP intervenes between an FQ and an agent NP (the subject) quantified. It also explains why (21b) with the same syntactic structure as (2b) is licit—since in this sentence the object NP has a fixed denotation, no more measurement is allowed.

(2) b. **Gakusei* *ga* hon wo **go-nin** kat-ta.
student SBJ book OBJ five-CLASS$_{person}$ buy-PAST
'Five students bought books.'

(21) b. *Gakusei* ga SORE/ SONO HON WO **yo-nin**
student SBJ DEM- PRON DET book OBJ four-CLASS$_{person}$
kat-ta.
buy-PAST
'A student bought it/the book.'

[7] According to Gunji and Hasida's (1999) definition derived from Dowty (1991), if Y is the incremental theme of X, then Y plays the theme role of X and the quantity of X is a homomorphic image of Y. In the case of telic predicates, the 'part-of' relation is preserved. To put it intuitionally, the event of eating three apples is three times as large as that of eating an apple.

In arguing for (A) above, however, Gunji and Hasida (1999) do not provide sufficient grounds. It is hard to understand why the 'measurement' of an incremental theme is primary and cost-free, while the quantification of other thematic roles is secondary and costly. Their second point (B) is not persuasive, either. According to them, the cause of the processing cost is the 'ambiguity' produced by the intervention. But in a great number of cases, there are no grounds for such ambiguity, given the disambiguating nature of the classifiers. For example in (2b), the FQ can never be associated with its neighbor, the object NP, owing to the classifier *-nin* that is always used with humans. It may be possible to save a part of their discussion by abandoning semantic ambiguity and reinterpreting the cost of FQ intervention in the context of left-to-right incremental processing, but then their theory will look much closer to ours.

Some essential examples concerning the context-dependence of FQ sentences are cited by Gunji and Hasida (1999).

(24) a. *Gakusei ga kōhī* WO **san-nin** chūmonshi-ta.
student SBJ coffee OBJ three-CLASS$_{person}$ order-PAST
'Three students ordered coffee.'

b. *Gakusei ga* sake WO DANSHI DE WA **san-nin**
student SBJ sake OBJ boy among-TOP three-CLASS$_{person}$
non-da.
drink-PAST
'Three students, among boys, drank sake.'

As they observe, the acceptability of sentences with an object intervening between the subject and an FQ is remedied remarkably in these sentences. Whereas they maintain that in (24a) *kōhī* (coffee) has a predetermined amount and therefore is free from measurement by the FQ, according to our view coffee is a routine order at a coffee shop and as such evades focusing. We agree with Gunji and Hasida (1999) that (24b) is acceptable because the contrastive phrase *danshi de wa* (among boys) focuses the FQ. While, however, they do not try to analyze the sentence any further, in our framework the contrast leads to defocusing of the object NP. The very existence of this kind of sentence testifies to the significance of context and, more specifically, information structure.

Gunji and Hasida's (1999) ideas have been developed in a more intelligible manner by Nakanishi (2007). Her analysis of Japanese FQs relies on Schwarzschild's (2002, 2006) monotonicity constraint. According to Nakanishi, in the ⟨NP-Quantifier-Postposition⟩ construction, the quantifier measures the number of individuals in the extension of the host NP, observing the monotonicity constraint in the nominal domain. In other words, the measure function represented by the quantifier must be monotonic to the part-whole structure given by the meaning of the NP. By contrast, an FQ, which Nakanishi analyzes as an adverbial, must be monotonic with respect to the individuals mapped from events denoted by the verbal phrase it adjoins to. This homomorphism obtaining between events and the meaning of the FQ, as Nakanishi argues, accounts for the constrained meanings of the FQ construction in contrast to the non-FQ one.

Nakanishi's explication seems to be applicable to a good part of linguistic data in Japanese. For instance, the appropriateness of sentence (21b) can be explained as denoting events, which are repeated four times, of buying a copy of the book. In this manner, her account may be revised to deal with the same kind of context-dependent data that we have done in this paper.

However, her theory is not without difficulties. First, since Nakanishi (2007) assigns syntactic factors as the cause of the subject/object asymmetry, the problem pointed out in Sect. 3.1 remains unresolved. This may be overcome by extending the theory in a further semantics-based manner to cover context-dependent cases as mentioned above.

The second difficulty that Nakanishi (2007) faces is more serious. As Tam (2011) points out, her concept of *event* that plays a central role in her paper is obscure and is sometimes too much overgeneralized. See the following sentence, a minimally modified citation from Tam (2011):

(25)	Asai-shi	no	kodomo	wo	**san-nin**	nokoshi-ta.
	the Asai family	ADNOM	child	OBJ	three-CLASS$_{person}$	leave-PAST
	'The lives of three children of the Asai family were saved.'					

The standard interpretation of this sentence is a single event of saving three children, rather than three events each denoting saving of a child as argued by Nakanishi. Following an information-based theory, we can understand why this is so—the sentence emphasizes that there exist children, in fact three, of the Asai family whose lives were saved, in contrast to a case where all the family members might have been massacred.

To conclude this subsection, the issue of FQs in Japanese lies beyond the scope of intrasentential semantics. We admit, however, that an incremental theme or monotonicity with respect to the event is prototypically associated with the focus of the sentence (and an agent with the non-focus), and this is where semantics interacts with context. Accordingly, it may be possible to enrich our theory in the future by taking into account the semantic factors that Gunji and Hasida (1999) and Nakanishi (2007) emphasize.

4 Linear Analysis of FQs

In Sect. 3.2 we saw that to make full use of the information structure indicated by Takami (1998a, b, c), asymmetric informational roles performed by the subject and object need to be introduced. However, these are not sufficient to account for the difference between (2a) and (2b) in acceptability.

(2)	a.	*Gakusei*	*ga*	**go-nin**	hon	wo	kat-ta.
		student	SBJ	five-CLASS$_{person}$	book	OBJ	buy-PAST
		'Five students bought books.'					

b. **Gakusei ga* hon wo **go-nin** kat-ta.
student SBJ book OBJ five-CLASS$_{person}$ buy-PAST
'Five students bought books.'

Instead of drawing on embedding within a VP and c-command as in Miyagawa (1989), we adopt a new criterion of processing order; in (2a) the FQ immediately follows its host, while in (2b) the FQ is fed in only after the object NP has been processed. We found our theory on the assumption that a sentence is processed incrementally, which is shared by a great number of psycholinguistic studies (see e.g. Mazuka and Itoh 1995 and Poesio and Rieser 2011). We develop this into the principle held throughout this paper that the interpretation on information structure of a sentence is made incrementally as the sentence is processed from left to right in real time.

Based on this theory, we put forward the following three hypotheses:

(26) 1. An FQ provides an independent nominal meaning, which is anaphorically related with that of the host NP.
2. The meaning of an FQ must be the focus of the sentence.
3. A sentence with a transitive verb is given an interpretation ⟨Non-Focus(Sbj) + Focus(Obj)⟩ by default (i.e., unless no explicit expression contradictory to it is given) as soon as the object NP has been recognized. No later correction is possible.

By *focus* we mean the most important new information in the sentence. *Non-focus* denotes such constituents as are not included in focus. Note that non-focus includes constituents that represent new information outside the focus, namely specific information, i.e., that known to the speaker but not to the hearer. It has been empirically shown that the subject and object behave differently in terms of susceptibility to definiteness or focus. For instance, the centering theory (Grosz et al. 1995) took advantage of this fact to model local anaphora within discourse units. For Japanese discourse, Walker et al. (1994) propose the following forward center (Cf) ranking:

(GRAMMATICAL OR ZERO)TOPIC > EMPATHY > SUBJECT > OBJECT2 > OBJECT > OTHERS

This is a formalization of the fact that, for example, the *ga*-marked NP is more likely to be the antecedent of a pronoun than the *wo*-marked NP.

Figure 1 illustrates how sentences (2a, b) and (21b) are processed according to the approach we propose. We assume that the processing is performed on the basis of accentual phrase (AP), a basic processing unit for Japanese sentences.

As shown in Fig. 1a, when the first AP of the sentence, the subject NP *gakusei ga* (student-SBJ) is read in, its information status is not yet fixed, as indicated by the feature [FOCUS *bin*(*ary*)]. The next input, the FQ *go-nin* (five-CLASS$_{person}$), stipulates that its host is the focus of the sentence; thus the FOCUS value of the subject NP is set to +, resulting in a correct interpretation of the sentence.

Figure 1b explains how the processing of example (2b) crashes. As soon as the second AP is input, the hypothesis (26.3) is applied to the sequence *gakusei ga hon wo* (student-SBJ book-OBJ); given no contradictory information, this is interpreted as

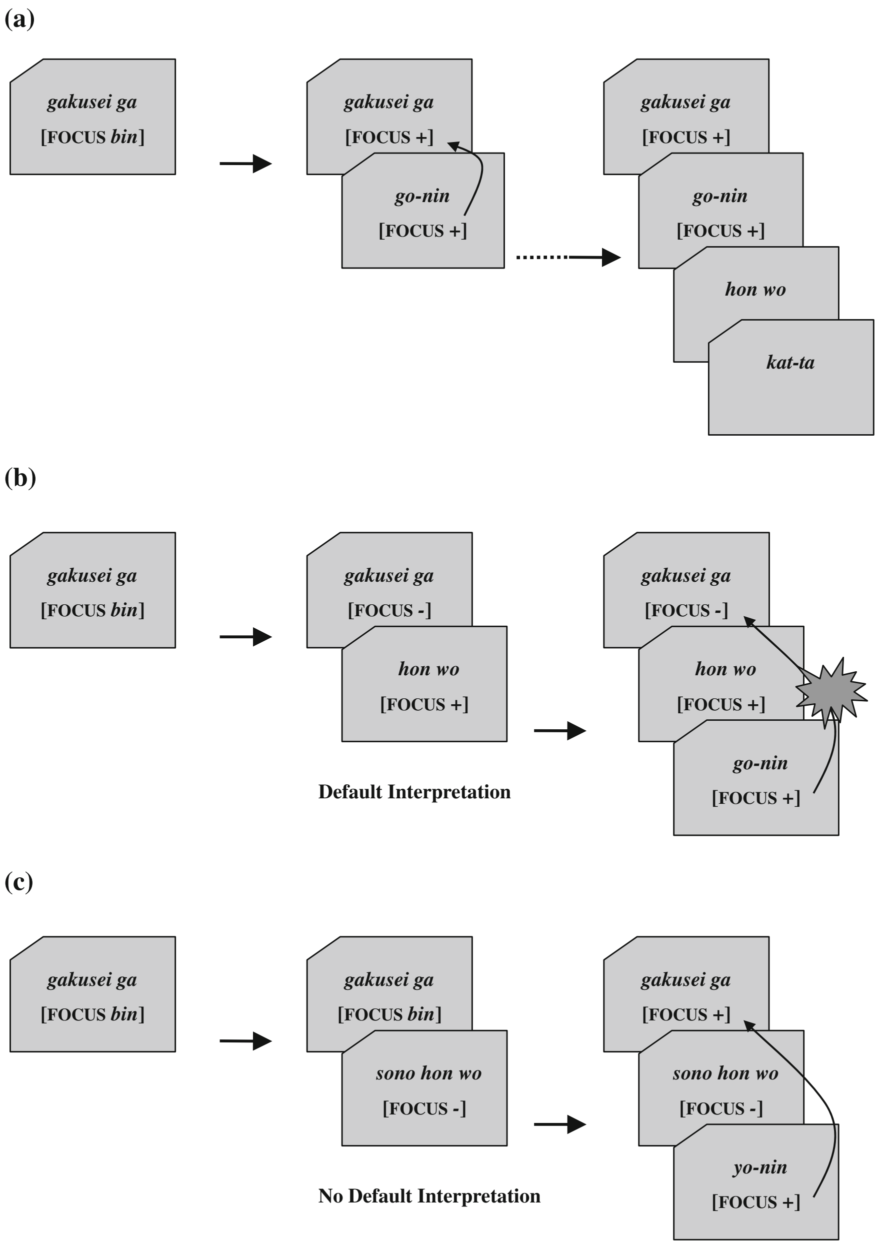


Fig. 1 Linear Processing of Sentences (2a, b) and (21b). **a** 2a. Gakusei ga go-nin hon wo kat-ta (student-SBJ FQ book-OBJ buy-PAST). **b** 2b. *Gakusei ga hon wo go-nin kat-ta (student-SBJ FQ book-OBJ buy-PAST). **c** 21b. Gakusei ga sono hon wo go-nin kat-ta (student-SBJ FQ book-OBJ buy-PAST)

⟨Non-Focus (Sbj) + Focus (Obj)⟩. The third input *go-nin* (five-CLASS$_{person}$), however, says that its host, the subject NP, must be the focus. A contradiction is incurred between the two specifications, and there is no way to remedy it. In the following section, we formalize the inappropriateness of this kind of sentence not directly as unification failure, but as a pragmatic inconsistency within an obtained feature structure.

In Fig. 1c, by contrast, the default interpretation in (26.3) does not apply, since the object NP *sono hon wo* (DET book OBJ) is definitely marked. Accordingly, the subject is felicitously quantified by the FQ *yo-nin* (four-CLASS$_{person}$).

The appropriateness of sentences (17a–d) with adverbial particles attached to FQs can be accounted for along the same lines. These particles, whose correlates in English have been analyzed as 'associating with focus' by Jackendoff (1972), are used in a context and often invoke one in which a certain part of the sentence is a presupposition. The object NPs in (17b–d) are definite. Furthermore, the object *repōto wo* ('report-OBJ') in (17a) does not seem to be a part of the focus, since in the most probable setting for the utterance, the speaker's attention is fixed on number of the students who have handed in a report.

In Sect. 3.2 we pointed out that the following sentence is not accounted for by Takami (1998a, b, c):

(23)	*Mishiranu*	*gakusei*	*ga*	hon	wo	**go-nin**	katte-it-ta.
	unfamiliar	student	SBJ	book	OBJ	five-CLASS$_{person}$	buy-go-PAST

'Five unfamiliar students bought books.'

Since the subject is explicitly modified by the adjectival word *mishiranu* (unfamiliar), which introduces new information, the default interpretation (26.3) does not apply. Therefore, the quantification of the subject NP by the FQ is not hindered here.

Our approach proposed in this paper, a functional one combined with incremental sentence processing, can thus account for the context-dependent difference in acceptability of FQ constructions. What Miyagawa's (1989) syntactic approach covers is only typical cases in which the subject and object are not explicitly marked to be inconsistent with the ⟨Non-Focus (Sbj) + Focus (Obj)⟩ interpretation; it constitutes but a part of the phenomena that we deal with. According to our framework, the asymmetry between the subject and object is not directly attributed to the configurational structure. It is rather explained by the fact that the transitive construction exhibits a strong tendency to undergo the ⟨Non-Focus (Sbj) + Focus (Obj)⟩ construal. FQ constructions are judged to be inappropriate when this default is rebutted by an input at a later stage of the sentence processing. They are similar to garden-path sentences in this respect.

Let us examine further whether or not other examples cited by Takami (1998a, b, c) to develop his argument can be accounted for by our framework. As illustrated in Sect. 3.2, Takami argues that sentences with an unergative verb embedding an FQ within its VP may be acceptable, contrary to Miyagawa's (1989) claim.

(15)	?∗*Gakusei*	*ga*	jibun	no	kane	de	**futa-ri**
	student	SBJ	self	ADNOM	money	with	two-CLASS$_{person}$

denwashi-ta.

telephone-PAST
'Two students telephoned using their own money.'

(19) a. *Gakusei* *ga* [$_{VP}$ kyōshitsu de **yo-nin**
student SBJ classroom LOC four-CLASS$_{person}$
ABARE- MAWATTE- I-TA]
act violently PROG-PAST
'Four students were acting violently in the classroom.'

b. *Dōryō* ga [$_{VP}$ Yamada-kun no teian ni
colleague SBJ NAME ADNOM proposal IOBJ
go-nin sanseishi-ta]
five-CLASS$_{person}$ agree-PAST
'Five colleagues of mine agreed to Mr. Yamada' s proposal.'

Takami implies that the difference in appropriateness between (15) and (19a, b) is caused by the difference in their information structures: he associates the adverbial phrases (AdvP) *jibun no kane de* (with their own money) in (15) with important information and *kyōshitsu de* (in the classroom) in (19a) with old information. However, how these remarks are related to Kuno's (1978) principle on word order, on which he founds his theory, is completely obscure.

These sentences also are accounted for by our approach. We assign by default different information roles to AdvPs depending on their semantic types. AdvPs for means and manner are treated as focus by default just as the object with a transitive verb in (26.3). This default interpretation on the information structure turns out to be inconsistent with an FQ fed in at a later step, causing the breakdown of the processing of sentence (15) in much the same manner as that of the transitive sentence (2b). In this context, note that the following is an acceptable sentence:

(27) *Gakusei* *ga* **futa-ri** jibun no kane de
student SBJ two-CLASS$_{person}$ self ADNOM money with
denwashi-ta.
telephone-PAST
'Two students telephoned using their own money.'

The contrast between (27) and (15), both with the means AdvP, is parallel to that between (2a) and (2b) with the direct object NP.

On the other hand, AdvPs for place or time as in (19a) are interpreted as non-focus or even old information by default. Thus the possibility of contradiction with an FQ is excluded.

The appropriateness of (19b), which Takami leaves unmentioned, can be attributed to the typical situation that this sentence brings to mind: given the context of the utterance, people's attention would be most likely to gravitate to the number of participants agreeing to Mr. Yamada's proposal. In such a context, only the FQ is the focus and the other parts of the sentence including the AdvP become the presupposition. In this case it is the situation evoked by the sentence, not the AdvP's semantic type, that prevents the AdvP from conflicting with the FQ.

Next, let us see how sentences (3a, b) are accounted for without relying on configurational structures.

(3) a. ∗**Go-nin** hon wo *gakusei ga* kat-ta.
five-CLASS$_{person}$ book OBJ student SBJ buy-PAST
'Five students bought books.'

b. **San-satsu** gakusei ga *hon wo* kat-ta.
three-CLASS$_{book}$ student SBJ book OBJ buy-PAST
'Students bought three books.'

We take an FQ at the initial position of the sentence as marked focus. Assuming that the default interpretation in (26.3) is applied to *gakusei ga hon wo* in (3b), the object NP is associated with the focus, which is compatible with the interpretation of the FQ *san-satsu* as focus. By contrast, no ready-made explanation is available for (3a). However, suppose that the object NP is construed as the focus by default, although this falls outside the scope of the default rule (26.3), the subject NP *gakusei ga* is again given inconsistent interpretations, non-focus and focus simultaneously.

5 Incremental Processing of FQ Sentences

5.1 Bottom-up Incremental Processing

In this section, we illustrate how the ideas we have proposed so far can be incorporated into a formal sentence processing system. In order to simulate an incremental, left-to-right sentence processing, we assume a kind of bottom-up, 'shift-reduce' parser (see Aho and Ullman 1972) that tries to combine lexical feature structures into larger ones, rather than words into phrase structure constituents, according to the HPSG grammar and lexicon.

In the main part of the algorithm, either *Shift* or *Reduce* applies (see Fig. 2). At the *Shift* step, the parser looks at the first word in the input string and pushes the feature structure of the word in the lexicon onto the top of the stack *L*. The parser repeats *Shift* steps until the right-hand side of a rewriting rule (more specifically, an HPSG schema) appears. Then, at the *Reduce* step, this sequence of element(s) is substituted for by the feature structure for the left-hand-side category.

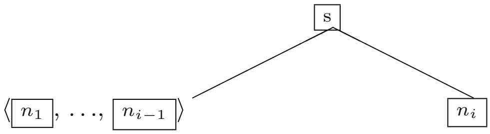
The basic scheme illustrated above is both impractical as a parsing algorithm and imperfect as a psychological model for human sentence processing. Readers interested in the latter aspect of the issue are referred to Joshi (1990), Kempson et al. (2001), and Poesio and Rieser (2011). Specifically, PTT by Poesio and Rieser, a dialogue interpretation theory modeled in terms of default logic, can be developed to deal with the context-dependent type of data discussed in this chapter. However, the simple parser given above should suffice for the purpose of this chapter.

1. Initialize the stack $L = \langle\ \rangle$
2. Either *Shift*:

- Consume the next word in the input string.
- Push the feature structure of the word in the lexicon onto the top of the stack.

3. Or *Reduce*:

- If $L = \langle \boxed{1}, \ldots, \boxed{n_1}, \ldots, \boxed{n_{i-1}}, \boxed{n_i} \rangle$ and there exists a schema



— Pop $\langle \boxed{n_1}, \ldots, \boxed{n_{i-1}}, \boxed{n_i} \rangle$ from L.

— Push $\boxed{\text{s}}$ onto the top of L.

4. If there are no more words in the sentence, then

- If $L = \langle \textit{sentence} \rangle$, then done.

5. Go to step 2.

Fig. 2 Shift-Reduce Incremental Processing

5.2 HPSG Specifications

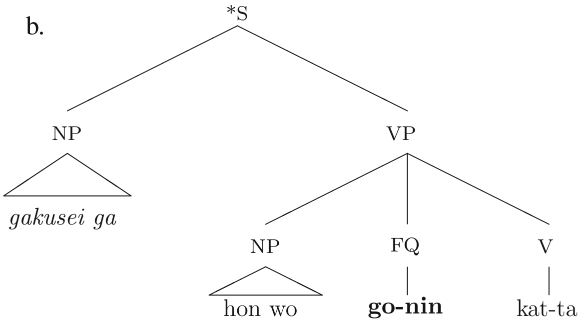
Our approach is similar to Fukushima (1992) in that it treats the FQ as a non-floating, VP-projection endocentric AdvP. While it is syntactically natural at least from the non-transformational point of view to take the FQ as AdvP on the one hand,[8] we must account for the semantic relationship between the FQ and its host NP on the other. We solve this difficulty without using transformation.

We propose a LOCAL feature FLOATING-QUANTIFIER, abbreviated as FQ, to relate the FQ to its host NP. This is lexically introduced by the FQ as shown in (31) below and discharged when the host NP is combined with the syntactic tree as illustrated in (32). Complete sentences must have a void FQ value. This feature is percolated up trees by the FQ *Feature Propagation Principle* in a similar way to the NONLOCAL features, but its application is confined within a single clause, since FQ is a *local* feature. The FQ feature prevents multiple FQs from being hosted by the same NP. The association between the FQ and its host is further constrained by the

[8] A base-generated approach to the syntax of FQs in English dates back to Dowty and Brodie (1984).

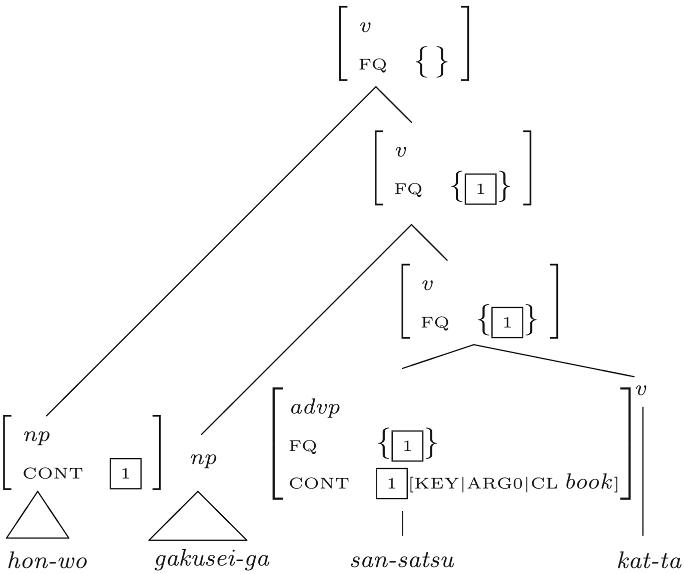
CLASSIFIER feature within ARG0 (the latter feature is called INSTANCE in the older version of MRS). This indicates the semantic category of the host NP (e.g., *book* for the classifier *satsu*, *person* for *nin*). The tree below illustrates how the FQ feature is introduced and discharged as sentence (2d) is processed.

(28)



We adopt Minimal Recursion Semantics (MRS; Copestake, Flickinger, Sag and Pollard 1999) as our semantic framework, which makes possible underspecified representations in terms of quantifier scope ambiguity. Below is the semantic information (the value of SYNSEM|LOC) for example (2d) using MRS:

(29)



The most remarkable feature with MRS is its non-recursive semantic representation, allowing as it does underspecification in terms of scope. The core of the specification is given as a list value of the attribute REL(ATION)S, called LISZT in the earlier version of the theory. In (29) above, the scopes of the quantifiers *n*(*u*)*m*(*era*)*l_fq_rel* and *ex*(*i*)*st*(*ential*) *_q*(*uan*)*t*(*ifier*)*_rel* remain underspecified, as shown by the fact that the values $\boxed{6}$ and $\boxed{13}$ of the BODY feature within the two quantifiers do not unify with any other feature structures. There exist two ways of disambiguating this: $\boxed{6} = \boxed{10} \wedge \boxed{13} = \boxed{1}$ for the reading according to which *nml_fq_rel* outscopes *exst_qt_rel* and $\boxed{6} = \boxed{1} \wedge \boxed{13} = \boxed{3}$ for the reversed scope interpretation. Note, however, that there is no substantial difference in this case.

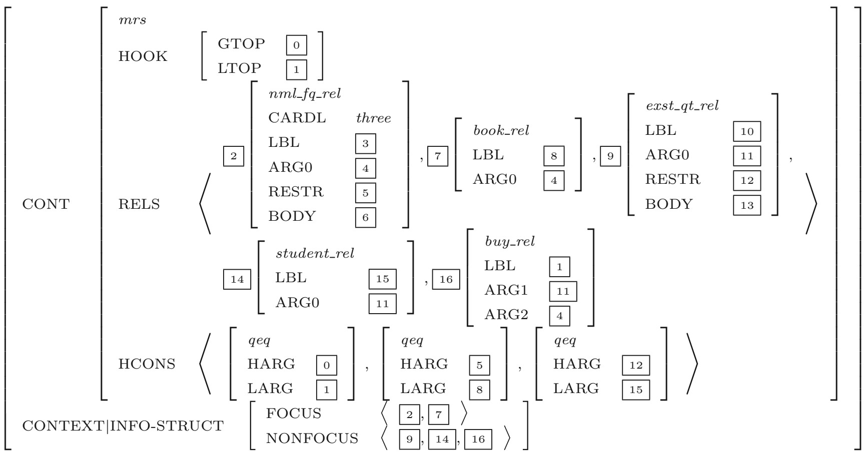
A *qeq* constraint stands for 'equality modulo quantifiers' (Copestake et al. 1999). Relating a handle in an argument position (the value of RESTR in case of quantifiers) to a label (the value of LBL), it indicates that the handle argument *h* must be identical with the label *l*, or more quantifiers 'float in' between *h* and *l*.

The HOOK feature is used to specify the parts of an MRS that are 'visible to semantic functors' (Copestake et al. 1999). LTOP stands for the semantic head of each partial MRS. GTOP is the highest LTOP of the sentence.

The specification of the information structure at the bottom of (29) is added to the MRS formalism after the HPSG version of the information packaging theory proposed by Engdahl and Vallduví (1996). The portions of the sentence other than the focus, i.e. link and tail, are merged into NONFOCUS in this paper. The values of FOCUS and NONFOCUS are lists consisting of *rel*(*ation*)s rather than *sign*s as formulated in Engdahl and Vallduví (1996). The FOCUS and NONFOCUS values are provided by the topic of the sentence, the FQ specification, and a default interpretation of the transitive construction as proposed later in this section. A more comprehensive framework to assign information structure to sentences, one specifically taking into account De Kuthy's (2002) extension of the theory applicable to a wider range of syntactic constructions, is left as a task to be carried out in future.

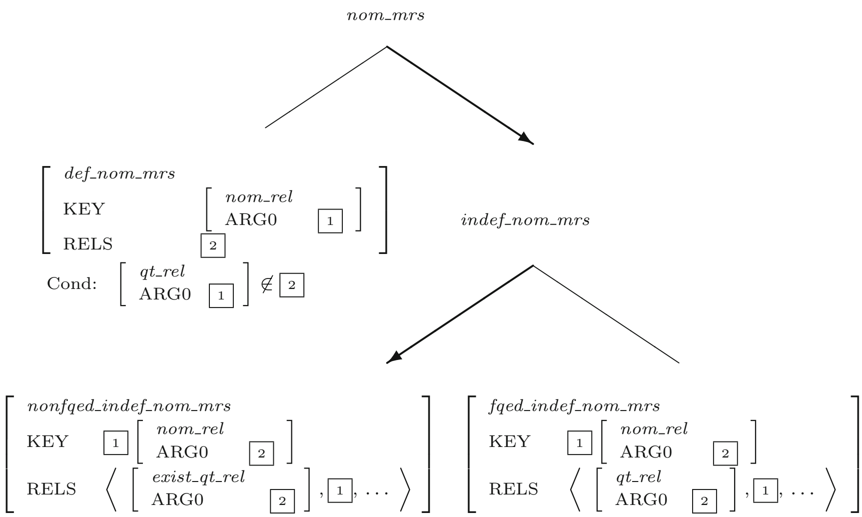
Since Japanese nominals are not obligatorily accompanied by articles, it is essential for our study to specify semantic information of nominals on the basis of a default inference mechanism. As illustrated in (30), all nominals are given the type *nom_mrs* at the top node of this type hierarchy, following the Minimal Recursion Semantics formalism. From this, *indef_nom_mrs* is inferred by default. The default inferences are shown by thick arrows in this diagram. When specified explicitly, *def_nom_mrs*, whose RELS contains no *relation* corresponding to a quantifier, is inferred from this. The default subtype of *indef_nom_mrs* is *nonfqed_indef_nom_mrs*. Its *nom_rel* is quantified by an existential quantifier. From *indef_nom_mrs*, when followed by an FQ, *fqed_indef_nom_mrs* is inferred. The substantial information on its quantifier is provided externally by an FQ. The feature KEY is adopted following the older version of MRS to specify the semantic head or the *relation* whose LBL value unifies with that of HOOK|LTOP.

(30)



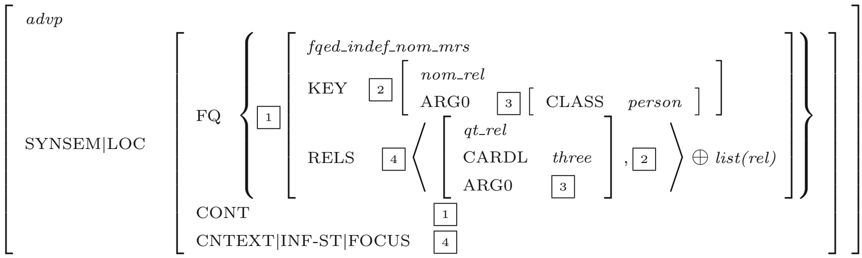
Let us explain further how the FQ and its host are associated with each other by the introduction and discharge of the FQ feature (see also the tree in (28)). (31) is the feature specification for the FQ *san-nin* (three-CLASS$_{person}$), which is constructed by combining the specifications for the numeral *san* and classifier *nin*.

(31)



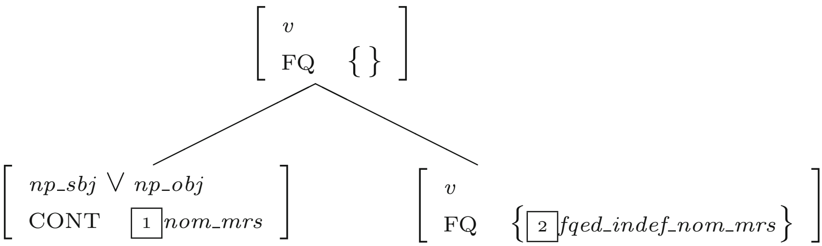
Thus it introduces the FQ feature. The FQ feature, introduced in this manner, is percolated up the tree by the FQ Feature Propagation Principle, which propagates the FQ value between the daughters and the mother. Note that this principle applies, unlike the NONLOCAL Feature Principle (Pollard and Sag 1994), only within a single sentence. The FQ feature is bound off by the FQ Cancellation Principle shown as (32) when the host, either a subject or object, is combined with the head phrase.

(32) FQ Feature Cancellation Principle



As formulated here, the mother node's FQ value is made empty when the result obtained by applying the function *fq_anaphoric* to the CONTENT value of the complement daughter unifies with the element within the head's FQ set value. The definition of this function is given below:

(33)



Condition: *fq_anaphoric*(1) = 2

The first half of this covers typical cases in which the FQ quantifies its indefinite host NP. Then the value of the function is the same as its input. Other cases like sentence (22), in which a whole-part relationship holds between the FQ and its definite host, are dealt with by specifying the superset-subset relationship between them as a condition. We interpret these cases as what has been treated under the rubric of 'substitution', a subclass of anaphora within the traditional discourse study (see Halliday and Hasan 1976), where the host NP and the FQ, referring to different denotations, share the same concept.

1. Initialize the stack $L = \langle\ \rangle$
2. Either *Shift*:

- Consume the next word in the input string.
- Push the feature structure of the word in the lexicon onto the top of the stack.
- If the new input is

$$\boxed{1}\begin{bmatrix} NP_{\textbf{wo}} & \\ \text{CONT|RELS} & \boxed{2} \\ \text{CONTEXT|INFO-STRUCT|NONFOCUS} & \boxed{3} \end{bmatrix} \text{ where } \boxed{2} \notin \boxed{3}$$

$$\text{AND } L \ni \boxed{4}\begin{bmatrix} NP_{\textbf{ga}} & \\ \text{CONT|RELS} & \boxed{5} \\ \text{CONTEXT|INFO-STRUCT|FOCUS} & \boxed{6} \end{bmatrix} \text{ where } \boxed{5} \notin \boxed{6},$$

then

$$\boxed{1} \longrightarrow \begin{bmatrix} \text{CONTEXT|INFO-STRUCT|FOCUS} & \boxed{2} \end{bmatrix}$$

$$\& \ \boxed{2} \longrightarrow \begin{bmatrix} \text{CONTEXT|INFO-STRUCT|NONFOCUS} & \boxed{5} \end{bmatrix}$$

3. Or *Reduce*:

⋮

4. If there are no more words in the sentence, then

- If $L = \langle \textit{sentence} \rangle$, then done.

5. Go to step 2.

Fig. 3 Shift-reduce incremental processing (Extended Version)

5.3 Linear Processing

Here we formalize how sentences with FQs are parsed by means of the typed feature-based shift-reduce incremental processing introduced in Sect. 5.1.

When the *Shift* operation applies to a new input that is an object NP not explicitly marked as non-focus, if there already exists a subject within the stack which is not marked as focus explicitly, then non-focus and focus are by default assigned to the subject and object, respectively, as specified in Fig. 3. This is not specified as a default rule commonly used in HPSG and constraint-based formalisms in general (see, for example, Lascarides et al. 1966), since it applies to only cases in which neither the subject is explicit focus nor the object explicit non-focus. Accordingly, sentences like (20b) with a definite object NP and (23) with a new information-marked subject NP are excluded.

The tree diagram in (34) illustrates how sentence (2b) is proven to be inappropriate by our framework. As shown in Fig. 1b in Sect. 4, during the first two steps of sentence processing, the sequence *gakusei ga hon wo* ('student-SBJ book-OBJ') is given the default information structure 'non-focus followed by focus' by the rule in Fig. 3. When the sentence processing has been completed, the feature SYNSEM|LOC|CONTEXT|INFO-STRUCT|FOCUS contains the information corresponding to both of the object NP *hon wo*, as provided by the specification in Fig. 3, and the FQ/subject NP, *go-nin* and *gakusei ga*, deriving from the specification in (31).

(34)

$$\text{If } \boxed{1} = \mathit{fqed\_indef\_nom\_mrs},$$
$$\text{then } \mathit{fq\_anaphoric}(\boxed{1}) = \boxed{1};$$
$$\text{otherwise } = \begin{bmatrix} \mathit{fqed\_indef\_nom\_mrs} & \\ \text{KEY}|\text{ARG0} & \boxed{2} \end{bmatrix}.$$
$$\text{Condition: } \mathit{subset\_relationship}(\boxed{3}, \boxed{2}) \wedge \boxed{1} = \begin{bmatrix} \text{KEY}|\text{ARG0} & \boxed{3} \end{bmatrix}$$

On the other hand, the feature NONFOCUS is composed of a list deriving from the semantics of the subject NP, following the default constraint in Fig. 3. This information, indicated by $\boxed{2}$ in the tree diagram, occurs both within FOCUS and NONFOCUS. The contradictory processing result—the same piece of information analyzed as focus and non-focus simultaneously—explains the observation that (2b) is an inappropriate sentence. Note that the inappropriateness does not directly correspond to a failure in unification.

6 Non-Standard FQ Sentences

The last problem left to us is how to account for sentences like (22):

(22) *Nadakō* *no* *seito* *wa*, mai-toshi Tōdai
Nada Highschool ADNOM student TOP every year Tokyo Univ.
wo **hachijū-nin** **ijō** jukensuru.
OBJ eighty-CLASS$_{person}$ more than take an exam
'As for Nada Highschool students, more than eighty of them take an exam for Tokyo University every year.'

We repeat here the difficulties that we pointed out in Sect. 3.2 as being presented by this kind of sentence. First, the subject, as a topicalized definite NP, is incompatible with the constraint by the FQ that it should be the focus. Second, the subject is not an appropriate host for the FQ even semantically; it refers to the whole set of the high school students, of which eighty students who take an entrance exam for Tokyo University is just a subset. In this respect, the sentence is similar to the following double-nominative sentence in which a whole-part relationship also holds between the *big* and *small* subjects (see Nakamura and Mori 2004 for the syntactic and semantic properties of double-nominative sentences in Japanese):

(35) *Nadakō* *no* *seito* *wa*, mai-toshi **hachijū-nin**
Nada Highschool ADNOM student TOP every year eighty-CLASS$_{person}$
ijō **ga** Tōdai wo jukensuru.
more than SBJ Tokyo Univ. OBJ take an exam
'As for Nada Highschool students, more than eighty of them take an exam for Tokyo University every year.'

The FQ construction undergoes this kind of interpretation under very limited conditions. See the following sentences with relative clauses.[9]

(36) a. John wa [*Mary* *ga* *mui-ta*] *ringo* *wo* **san-ko**
NAME TOP NAME SBJ peel-PAST apple OBJ three-CLASS$_{inanim}$
tabe-ta.
eat-PAST
'John ate three of the apples that Mary peeled.'

b. [Kinō **san-ko** tabe-ta] *ringo* *wa* jikka
yesterday three-CLASS$_{inanim}$ eat-PAST apple TOP home
kara okutte ki-ta mono da.
SOURCE send-PAST FORM- NOUN COPL
'The apples three of which I ate yesterday were sent from my home.'

We assume that contexts are provided by constructions such as topicalization and relativization in which the *host* NP is interpreted as definite and a whole-part reading of the host/FQ relationship is forced accordingly.

In Sect. 5.2 we assigned a semantic specification for a whole NP (see (31)) to an FQ and related this to the semantics of its host by the function *fq_anaphoric* as illustrated in (32) and (33). In default cases in which the 'reference' relationship holds between the FQ and its host, *fq_anaphoric* identifies the semantic information of both constituents. When a topic or relative constructions are supported by some specific context, the FQ's denotation is just a subset of that of its 'host'. As pointed out in Sect. 5.2, we treat these cases under the rubric of 'substitution', another subclass of anaphora. How this constraint and, above all, the contexts that trigger the latter interpretation should be defined, is left as an open question. In any case, the treatment of the FQ as anaphora with the semantics of an entire NP lays the foundation for its interpretation depending on contexts.

7 Conclusions

We have in this paper provided a new perspective on FQs in Japanese on the basis of real-time, incremental sentence processing and information structure. Syntactically,

[9] Sentence (36a) is cited from Kempson et al. (2001, p. 141).

the FQ is analyzed without transformation or using long-distance dependency like SLASH in HPSG; the FQ and its host are matched by feature percolation within a single sentence. Semantically, the FQ is given a whole piece of information as an independent NP, and this stands in an anaphoric relation to its host. This framework accounts for not only typical FQ constructions in which the FQ is semantically identical to the host, but also the kind of FQ sentence that have a partitive relationship with the host. We have shown that the approach we propose covers more linguistic data than hitherto.

The frameworks we have been relying on — HPSG and MRS — have been adopted as formalisms with established general applicability upon which we can formalize our proposal. The essence of our arguments repeated above could be given embodiment based on another formal theory as long as we can simulate with it incremental, left-to-right processing with default interpretation while integrating information from the lexicon, syntax, semantics, and context. One candidate is Dynamic Syntax (Kempson et al. (2001)), which directly makes linear processing possible.

Now we are ready to answer the question we posed at the beginning of this paper. Having throughout this paper developed the hypothesis that the FQ is the syntactic position on which (a part of) the sentential focus falls, we have shown that it is tenable. By contrast, the prenominal quantifier is informationally unmarked, in that any portion of the quantified NP may or may not carry focus. Thus the co-existence of two parallel means of quantifier expression in Japanese is explained by the two different informational functions imposed on the two constructions.

Acknowledgments We wish to express our gratitude to the reviewers and participants at LENLS 2005 for suggestions and comments on our earlier work whose outline appeared as Yoshimoto, Kobayashi, Nakamura and Mori (2006). Thanks are also due to the editor and two anonymous reviewers of SLAP for valuable advice.

References

Aho, A. V., & Ullman, J. D. (1972). *The theory of parsing, translation, and compiling* (Vol. 1 Parsing). Englewood Cliffs, NJ: Prentice-Hall.

Copestake, A., Flickinger, D., Sag, I. A., & Pollard, C. (1999). Minimal recursion semantics: An introduction, Stanford: Ms., Center for the Study of Language and Information, Stanford University.

De Kuthy, K. (2002). *Discontinuous NPs in German - A case study of the interaction of syntax, semantics and pragmatics*. Stanford, CA: CSLI Publications.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–619.

Dowty, D., & Brodie, B. (1984). The semantics of “floated" quantifiers in a transformationless grammar. In *Proceedings of the Third West Coast Conference on Formal Linguistics*, (pp. 75–90).

Engdahl, E., & Vallduví, E. (1996). Information Packaging in HPSG. In: C. Grover & E. Vallduví (Eds.), *Edinburgh Working Papers in Cognitive Science*, vol. 12, Studies in HPSG, (pp. 1–31).

Fukushima, K. (1991). Phrase structure grammar, Montague semantics, and floating quantifiers in Japanese. *Linguistics and Philosophy*, *14*, 581–628.

Fukushima, K. (1993). Model theoretic semantics for Japanese floating quantifiers and their scope properties. *Journal of East Asian Linguistics*, *2*(3), 213–228.
Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, *21*(2), 203–225.
Gunji, T., & Hasida, K. (1999). Measurement and quantification. In T. Gunji & K. Hasida (Eds.), *Topics in constraint-based grammar of Japanese* (pp. 39–79). Dordrecht: Kluwer Academic.
Halliday, M. A. K. (1976). *Cohesion in English*. London: Longman.
Inoue, K. (1978). *Nihongo no Bunpo Kisoku (Grammatical Rules in Japanese)*. Tokyo: Taishukan.
Jackendoff, R., (1972). *Semantics in generative grammar*. Cambridge, MASS: MIT Press.
Joshi, A. (1990). Processing crossed and nested dependencies: An automation perspective on the psycholinguistic results. *Language and Cognitive Processes*, *5*(1), 1–27.
Kempson, R., Meyer-Viol, W., & Gabbay, D. (2001). *Dynamic syntax: The flow of language understanding*. Oxford: Blackwell.
Kuno, S. (1978). Theoretical perspectives on Japanese linguistics. In J. Hinds & I. Howard (Eds.), *Problems in Japanese syntax and semantics*. Tokyo: Kaitakusha.
Kuroda, S-Y. (1983). What can Japanese say about government and binding? In *Proceedings of the Second West Coast Conference on Formal Linguistics*, (pp. 153–164).
Langacker, R. (1966). On pronominalization and the chain of command. In D. A. Reibel & S. A. Schane (Eds.), *Modern studies in English: Readings in transformational grammar*. Englewood Cliffs, NJ: Prentice-Hall.
Lascarides, A., Briscoe, E., Asher, N., & Copestake, A. (1996). Order independent and persistent typed default unification. *Linguistics and Philosophy*, *19*(1), 1–89.
Mazuka, R., & Itoh, K. (1995). Can Japanese speakers be led down the garden path? In R. Mazuka & N. Nagai (Eds.), *Japanese sentence processing* (pp. 295–329). Hillsdale, NJ: Lawrence Erlbaum.
Miyagawa, S. (1989). *Structure and case marking in Japanese. Syntax and semantics* (Vol. 22). San Diego: Academic.
Nakamura, H., & Mori, Y. (2004). Relational nouns as anaphors. In: H. Masuichi, T. Ohkuma, K. Ishikawa, Y. Harada & K. Yoshimoto(Eds.), *Proceedings of the 18th Pacific Asia Conference on Language, Information and Computation*, (pp. 71–80), Logico-Linguistic Society of Japan.
Nakanishi, K. (2007). *Formal properties of measurement constructions*. Berlin: Walter de Gruyter.
Poesio, M., & Rieser, H. (2011). An incremental model of anaphora and reference resolution based on resource situations. *Dialogue and Discourse*, *2*(1), 235277.
Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago, IL: University of Chicago Press.
Reinhart, T. (1979). Syntactic domains for semantic rules. In F. Guenthner & S. J. Schmidt (Eds.), *Formal semantics and pragmatics for natural languages* (pp. 107–130). Dordrecht: Reidel.
Reinhart, T. (1981). Definite NP anaphora and c-command domains. *Linguistic Inquiry*, *12*(4), 605–635.
Reinhart, T. (1983). *Anaphora and semantic interpretation*. Chicago: The University of Chicago Press.
Saito, M. (1985). *Some asymmetries in Japanese and their theoretical implications*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge.
Schwarzschild, R. (2002). The grammar of measurement. In *The Proceedings of 12th Semantics and Linguistic Theory* (SALT12), (pp. 225–245).
Schwarzschild, R. (2006). The role of dimensions in the syntax of noun phrases. *Syntax*, *9*, 67–110.
Shibatani, M. (1977). Grammatical relations and surface case. *Language*, *53*(4), 789–809.
Takami, K. (1998a). Nihongo no Sūryōshi Yūri ni tsuite: Kinōron-teki Bunseki [Jō] (On quantifier floating in Japanese: A functional analysis, Part 1). *Gengo*, *27*(1), 86–95.
Takami, K. (1998b). Nihongo no Sūryōshi Yūri ni tsuite: Kinōron-teki Bunseki [Chū] (On quantifier floating in Japanese: A functional analysis, Part 2). *Gengo*, *27*(2), 86–95.
Takami, K. (1998c). Nihongo no Sūryōshi Yūri ni tsuite: Kinōron-teki Bunseki [Ge] (On quantifier floating in Japanese: A functional analysis, Part 3). *Gengo*, *27*(3), 98–107.

Tam, W. L. (2011). *Scope, distributivity and categorization in classifier constructions*. Ph. D. Thesis, Graduate School of Interdisciplinary Information Studies, The University of Tokyo, Tokyo.

Walker, M., Iida, M., & Cote, S. (1994). Japanese discourse and the process of centering. *Computational Linguistics*, *20*(2), 193–232.

Yatabe, S. (1990). Quantifier floating in Japanese and the θ-hierarchy. In: M. Ziolkowski, M. Noske & K. Deaton (Eds.), CSL 26. Vol. 1: The Main Session, Chicago Linguistics Society, (pp. 141–163).

Yoshimoto, K., Kobayashi, M., Nakamura, H., & Mori Y. (2006). Processing of Information Structure and Floating Quantifiers in Japanese. In: T. Washio, et al. (Eds.), *New Frontiers in Artificial Intelligence: Joint JSAI 2005 Workshop Post-Proceedings*, (pp. 103–110). *Lecture Notes in Artificial Intelligence 4012*, Berlin: Springer.

Correction to: Formal Approaches to Semantics and Pragmatics

Elin McCready, Katsuhiko Yabushita and Kei Yoshimoto

Correction to:
E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*, Studies in Linguistics and Philosophy, DOI 10.1007/978-94-017-8813-7

This book was inadvertently published with the incorrect author name. "Eric McCready" has been corrected as "Elin McCready". This has now been amended throughout the book.

The updated version of the book can be found at
DOI 10.1007/978-94-017-8813-7

E. McCready et al. (eds.), *Formal Approaches to Semantics and Pragmatics*,
Studies in Linguistics and Philosophy, DOI: 10.1007/978-94-017-8813-7_17,