Manabu Okumura
Daisuke Bekki
Ken Satoh (Eds.)

LNAI 7258

New Frontiers in Artificial Intelligence

JSAI-isAI 2011 Workshops, LENLS, JURISIN, ALSIP, MiMI
Takamatsu, Japan, December 2011
Revised Selected Papers

Springer

Lecture Notes in Artificial Intelligence 7258

Subseries of Lecture Notes in Computer Science

Manabu Okumura Daisuke Bekki
Ken Satoh (Eds.)

New Frontiers in Artificial Intelligence

JSAI-isAI 2011 Workshops
LENLS, JURISIN, ALSIP, MiMI
Takamatsu, Japan, December 1-2, 2011
Revised Selected Papers

Springer

Volume Editors

Manabu Okumura
Tokyo Institute of Technology
Precision and Intelligence Laboratory
Yokohama, Japan
E-mail: oku@lr.pi.titech.ac.jp

Daisuke Bekki
Ochanomizu University
Faculty of Science
Department of Information Science
Tokyo, Japan
E-mail: bekki@is.ocha.ac.jp

Ken Satoh
National Institute of Informatics/Sokendai
Tokyo, Japan
E-mail: ksatoh@nii.ac.jp

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-32089-7 e-ISBN 978-3-642-32090-3
DOI 10.1007/978-3-642-32090-3
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: Applied for

CR Subject Classification (1998): I.2.4, I.2.6, I.2.7, I.2, H.3, H.4, H.2.8, F.4.3, J.1, F.2, G.2.2

LNCS Sublibrary: SL 7 – Artificial Intelligence

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

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

Preface

JSAI (The Japanese Society for Artificial Intelligence) is a premier academic society that focuses on artificial intelligence in Japan and was established in 1986. JSAI publishes journals and bimonthly transactions, and hosts 18 special interest groups. The JSAI annual conference attracts about 700 attendees each year. JSAI-isAI (JSAI International Symposia on Artificial Intelligence) 2011 was the Third International Symposium, which hosted four co-located international workshops that had been selected by the JSAI-isAI 2011 Organizing Committee. This was in succession to the international workshops co-located with the JSAI annual conferences since 2001. JSAI-isAI 2011 was successfully held during December 1–2 in Takamatsu, Japan; 110 people from 14 countries participated in JSAI-isAI 2011. This volume of "New Frontiers in Artificial Intelligence: JSAI-isAI 2011 Workshops" is the proceedings of JSAI-isAI 2011. The organizers of the four workshops, LENLS, JURISIN, ALSIP, and MiMI, hosted by JSAI-isAI 2011, selected 21 papers in total. The acceptance rate was about 48%. This has resulted in the excellent selection of papers that are representative of some of the topics of AI research both in Japan and in other parts of the world. LENLS (Logic and Engineering of Natural Language Semantics) is an annual international workshop on formal semantics and pragmatics. LENLS hosted by JSAI-isAI 2011 was the eighth event in the series. LENLS focuses on the formal and theoretical aspects of natural language, which demonstrates one of the strengths of Japanese AI studies. The Workshop Chair was Alastair Butler (Japan Science and Technology Agency, PRESTO). JURISIN (Juris-Informatics) was the fifth event focusing on juris-informatics among people from various backgrounds such as law, social science, information and intelligent technology, logic and philosophy, including the conventional "AI and law" area. The Workshop Chairs were Shozo Ota (University of Tokyo) and Ken Satoh (National Institute of Informatics). ALSIP (Algorithms for Large-Scale Information Processing in Knowledge Discovery) was the second event focusing on large-scale data processing in problems such as data mining, clustering, machine learning, statistical analysis, and other computational aspects of knowledge discovery problems. The Workshop Chairs were Koji Tsuda (National Institute of Advanced Industrial Science and Technology) and Shin-ichi Minato (Hokkaido University). MiMI (Multimodality in Multispace Interaction) was held for the first time. In this workshop, how multispace is managed in socially, temporally, and sequentially complex environments was discussed. The Workshop Chairs were Mayumi Bono

(National Institute of Informatics) and Nobuhiro Furuyama (National Institute of Informatics). It is our great pleasure to be able to share parts of the outcome of these fascinating workshops through this volume. We hope the readers of this book find a way to grasp the state-of-the art research outcomes of JSAI-isAI 2011, and may be motivated to participate in future JSAI-isAI events.

April 2012

Manabu Okumura
Daisuke Bekki
Ken Satoh

Table of Contents

LENLS

JURISIN

ALSIP

MiMI

Logic and Engineering of Natural Language Semantics (LENLS) 8

Alastair Butler

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

1 The Workshop

On December 1–2, 2011 the Eighth International Workshop of Logic and Engineering of Natural Language Semantics (LENLS 8) took place at Sunport Hall Takamatsu. This was held as a workshop of the third JSAI International Symposia on AI (JSAI-isAI 2011), sponsored by The Japan Society for Artificial Intelligence (JSAI).

LENLS is an annual international workshop focusing on topics in formal semantics, formal pragmatics, and related fields. This year the workshop featured invited talks by Frank Veltman, on imperatives, and Kentaro Inui, on knowledge acquisition and large-scale lexical semantic resources. In addition there were 16 presentations of talks selected by the program committee (see Acknowledgements) from the abstracts submitted for presentation.

As always with the LENLS workshops, the content of the presented papers was rich and varied, at times exhibiting extreme focus on formal accounts of specific empirical phenomena, and at other times tackling broader theoretical issues. Topics represented included, on the empirical side, sentence-final stress, expressions of sensory, evidential and epistemological meanings, the counterfactual future in the past, quantification items, polarity items, discourse markers and, more generally issues of language acquisition, cooperativity and unawareness and on the theoretical side, formal results capturing linguistic constraints restricting extractability and considerations of Sheaf-Theory as a framework for pursuing semantic modelling.

In addition, following the workshop on December 3rd, two Tutorial Lectures were held at the same venue in Takamatsu. The first, by Eric McCready, presented theories of evidentials. The second lecture by Frank Veltman consisted of a tutorial introducing theories of counterfactuals.

The remainder of this introduction will briefly indicate the content of the papers selected to appear in the present volume.

2 Papers

The submitted papers in the LENLS part of the present volume fall into two classes. The first are papers addressing pragmatic phenomena from a formal perspective. In ‘Begging Questions, Getting Answers and Basic Cooperativity’,

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 1–2, 2012.

Nicholas Asher and Jason Quinley consider game-theoretic rationales for minimal cooperativity. In his ‘Players who don’t know how to play’, Mauricio Hernandes provides a Haskell implementation of an analysis of unawareness in a game context. In their paper ‘Linking probabilistic accounts: polarity items and discourse markers’, Margot Colinet and Grégoire Winterstein link previously unrelated probabilistic approaches. In ‘The Japanese particle yo in declaratives’, David Oshima presents an analysis of declaratives with *yo* with a non-rising contour as carrying out the illocutionary acts of assertion and blaming.

The second class of papers apply either functional programming or logic-based mechanisms with properties to shine light on structural semantic phenomena. In ‘Conjoined nominal expressions in Japanese: Interpretation through monad’, J.-R. Hayashishita and Daisuke Bekki capture properties of nominal expressions with an interpretation through monad. Christina Unger proposes a formulation of the basic ideas of dynamic semantics in terms of the state monad in her ‘Dynamic semantics as monadic computation’. Satoru Suzuki’s paper ‘Measurement-Theoretic Foundations of Gradable-Predicate Logic’ gives a logical system designed for the analysis of gradable adjectives. Finally Alastair Butler and Kei Yoshimoto, in ‘Towards a self-selective and self-healing evaluation’, propose an operation for automating a regulation of binding dependencies based on evaluation state and expression content.

Acknowledgements. Let me acknowledge some of those who helped with the workshop. The program committee and organisers, in addition to myself, were Daisuke Bekki, Eric McCready, Yoshiki Mori, Yasuo Nakayama, Katsuhiko Yabushita, Tomoyuki Yamada, Shunsuke Yatabe and Kei Yoshimoto. Daisuke Bekki also was liaison with JSAI and together with Kei Yoshimoto organised and mentored many aspects of the workshop. Natsuha Katakura was the vital cause for things going well. Finally, the organisers would like to thank the JST PRESTO program ‘Synthesis of Knowledge for Information Oriented Society’ for financial support and JSAI for giving us the opportunity to hold the workshop.

Begging Questions, Their Answers and Basic Cooperativity

Nicholas Asher[1] and Jason Quinley[2]

[1] CNRS, Institut de Recherche en Informatique de Toulouse et Université Paul Sabatier
[2] University of Tübingen

Abstract. We consider game-theoretic rationales for minimal cooperativity, in particular responses to questions or requests for help with false answers. Lying enables preservation of property and face for both speaker and hearer and constitutes a Pareto-optimal outcome. Rationales for this behavior include expectations of reciprocity, other-regarding, and maintenance of face.

1 Introduction

Cooperativity is a hallmark of almost every philosophical and linguistic theory of conversation. Much of this is due to the influence of Grice (1975) but also of Lewis (1969) and his work on signalling games. Typically, cooperativity, when made formally explicit in theories of conversation, entails an alignment of interests and sincere and honest cooperation among conversational participants—for instance, in Gricean cooperative conversation people should answer questions with what they believe to be true answers. We call this strong form of cooperativity *Grice cooperativity* (Asher & Lascarides, 2011). Actual examination of conversations, however, shows that cooperativity is a more nuanced affair. Conversation, like all rational interaction, is driven by preferences or interests. Agents may have only partially aligned interests but nevertheless engage in basic forms of linguistic cooperativity: people tend to reciprocate greetings to strangers, ask questions when they aren't interested in the answers, and answer questions, even when they don't answer truthfully. These forms of interaction form the basis of rhetorical coherence in dialogue (Asher & Lascarides, 2003). The question we want to look at is, what rational principle underlies speakers' commitment to rhetorical coherence?

To sharpen our analysis, we concentrate on one of these basic forms of linguistic exchange. What is the rationale for deciding to answer a question, and whether to answer it truthfully or not? Consider the following conversation:

(1) a. **Beggar:** Do you have any money?
b. **Passer-by:** No I don't.

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 3–12, 2012.

The passer-by may in fact have money; in this case it appears as though the passer-by is basic cooperative but not Grice cooperative. He is providing an answer, but not a truthful one. One might argue that this is not the case, because the beggar is not asking the question literally meant by the interrogative.[1] The beggar's question, so the argument goes, is shorthand for:

(2) Do you have any money that you can spare?

However, we think that trying to fold the implicatures of the speech act into the question is not a useful strategy. The sense of *can* here is still underspecified. To be sure, if the passer-by is a well-to-do academic, then in some sense of *can* he has money he can spare. In either case, we think it is relatively clear that the passer-by in our situation has not answered truthfully.

There are other conversational options that the passer-by could have used. He could simply have not answered the question and walked away (which sometimes happens). He could have answered truthfully–

(3) I have some change, but I don't want to give you any.

2 Social Dimensions of Questions and Answers

When considering the replique in (3), most people think this is not a likely or an appropriate response. But why would the interchange not go this way? The answer lies in a facet of language and linguistic usage that is not directly related to truth conditional content. According to Brown and Levinson's (1978) 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/ needs/ autonomy.

The problem with many speech acts is that they are inherently face-threatening. Brown and Levinson term these face-threatening acts *FTA's*, and argue that the second capability of language helps us to manage situations where the face of one or more of the conversants is in jeopardy. Speakers must therefore derive an appropriate strategy by weighing their own preferences, those of the hearer, and the potential of an FTA. Politeness theory has so far concentrated on alternative dialogue actions for a given conversational turn—-for example, discussing the differences in face threatening level of a question, an indirect speech act, a deferential request, or a bald command. We will apply the notion of an FTA to complex discourse acts and alternative discourse structures, by exploiting work on discourse structure.

---

[1] We are grateful to an anonymous reviewer for this point.

Asking a question such as the one we are considering involves a possible loss of face by the questioner, at least insofar as he places himself at risk of rejection. On the other hand, this risk is balanced by the possibility of a financial reward. This emphasis on strategic interaction makes politeness an inviting ground for testing the merits of game-theoretical models. It also provides a justification for our assignments of utilities in the game scenario for questioning and answering.

3 The Model

We provide here a formal model of making a request through the mechanism of asymmetric bargaining and exchange games. An exchange game can be thought of as a formal model of two or more agents sending goods to one another. The asymmetric component is crucial here, as we place our fate in the hands of the hearer when making a request. To model this, we incorporate the literature on the burgeoning field of trust games.

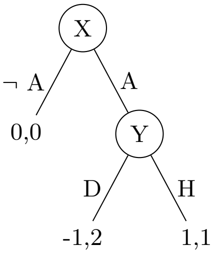


Fig. 1. Extensive Form

		Player Y	
		H	D
Player X	A	1;1	-1;2
	$\neg A$	0;0	0;0

Fig. 2. Normal Form

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

Trust games depict a scenario where Player X has an initial option to defer to Player Y for a potentially larger payoff for both. However, similar to the Prisoner's Dilemma (e.g. Rand et al. 2009), Player Y could defect on Player X and keep more money for himself. For a one-shot game, this act of deference will not occur for a rational Player X. Hence, a signal granting yes-no power to a hearer would not be rational; i.e. a kind response does not emerge as an optimal strategy in a single dialogue or in a single round of trust game play. To account for the emergence of polite language, Quinley (2011) examines trust game decision and payoff structure, the motivations for using them, and the optimal strategy for one round of play and of the effects of repetition. Building on Rand et al. (2009), he shows that reputation effects provide a rational basis for actions based on politeness. The crucial thing to observe about conversations is that conversations can be prolonged indefinitely. We can consider them as games thus in which reputation effects matter.[2] This game, as seen in Figure 1, models a situation of possible cooperation under conflicting interests like the Prisoner's Dilemma and other public goods games, but the inherent asymmetry in the social situations like those involving begging for money makes it a starting point for modeling requests. Where our model diverges is in the multitude of information states and options for cooperative and defective behavior.

We take the exchange between the Beggar (B) and the Passer-by (P) in (1) to be an extensive game of incomplete information. Unlike van Rooij's (2003) analysis, we are not supposing a standard signaling game environment. Indeed, we think that this gets the essential nature of politeness and basic cooperativity wrong.

To elaborate on why, one must consider that the essential benefit of signaling games is to arrive at a (hopefully)non-trivial equilibrium (i.e. a convention) whose communication strategies are somehow efficient, whether they be Nash equilibria or evolutionary stable strategies. Van Rooij (2003) posits that politeness is an instance of the Zahavian Handicap Principle, the means by which costly signals are preserved in a population via sexual selection as honestly communicating fitness. This depicts politeness then as an emergent phenomenon like the *peacock's tail*, a showy feature designed to signal a presumably fit speaker and rests on the notion that politeness formulae typically add length to an utterance. The argument is that just as peahens select showy peacocks for mating based on their show of fitness, so politeness is also selected for. Our example with the beggar is a direct affront to this analysis. Although the beggar and passer-by have different preferences, much like the peacock and peahen, the reply to the beggar is far from honest in the Gricean sense of cooperativity and yet it preserves the face of both conversants. Further, there is no informative *signaling system* that would emerge, as the passer-by has little incentive to give the beggar money regardless of his personal bankroll.

We suppose with Farrell (1993) that linguistic messages have determinate meanings, and following Asher & Lascarides (2012) we extend Farrell's assumption and make it more precise by saying that linguistic messages not only have a stable compositional semantics but also a discourse meaning that includes their rhetorical functions. Thus, we take P's response to have the discourse meaning that it is a direct answer to B's question.

B decides to ask P for money based on his expectation of return. While the question invites perhaps an infinite number of responses each with a different shade of politeness, we simplify the game: the Passer-by can **Walk** or **Talk.** Should he Talk, he can be honest about his state or not. The utilities and states in the diagram of figure 1 stem from the following reasoning.

B cannot deduce if P has money or not, and this is common knowledge. Suppose that we further assume that the interests of B and P diverge, in that B prefers to have P give money and P does not prefer to do so. Furthermore, we assume that P has no preference to have a conversation with B. While P

---

[2] This work provides a game-theoretic reconstruction of Traum and Allen's (1994) of conversational obligations, to which we are very sympathetic.

could walk past B, keeping his money, this threatens B's face. B has put himself in a position of risk by asking for help, but in asking a question he has also put a certain burden on B by opening up a certain conversational structure which has linguistically encoded continuations. These continuations stem from the question's discourse semantics. To sketch very quickly the outlines of the view, a question's discourse semantics is defined in terms of its effect on an input information state. For our purposes here, we take the input information state for a question to be 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 introduced by this structuring of the information set are three: eliminate some of the subsets by providing a direct answer or indirect answer (which on the basis of certain defeasible inferences provides a direct answer), respond with another question, which may add further structure to the information set, or 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. The third option is not a plausible one for the addressee: it invites a continuation of the conversation by B with a possible, unpleasant exchange (or even violent reprisal). On the other hand, if P responds with another question, he puts himself at risk needlessly and invites a further exchange with the beggar, which is not desired. Similarly, and interestingly, if P answers *Yes*, the conversation can be continued with further negotiations, further conversation and maybe a potential donation, which is also undesirable. If P answers *No*, he avoids the potential face-threat and incurs no further negotations. In the rest of this paper, we concentrate on the answering moves to B's question.

The point about the*Yes* and *No* answers is crucial and perhaps not so obvious. We elaborate on it by appealing to our underlying linguistic discourse model, SDRT (Asher & Lascarides 2003) and its semantics of questions and the discourse relations they enter into. In principle, any conversation may always be continued with further moves. But these moves have costs; they always induce commitments by the speaker, in particular in our case commitments to exposure to negative face or threats to the other's face. Thus, 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 new move by participant j at $m + 1$, effectively ending the conversation.

The *Yes* direct answer is not such a move. It provides a natural continuation for further questions or actions. One such continuation is a further question like:

(4) Can you give some of it to me?

is an instance of Asher & Lascarides (2003)'s *Question-Elaboration*, or Q-elab discourse move. A Q-elab move is a question β that attaches to another discourse move α in such a way such that at least a part of a conventionally associated *speech act related goal* (SARG) behind i's making α is met or established as unmeetable by a direct answer to β. When someone asks a question, we defeasibly infer as part of our linguistic competence that one of the SARGs behind the question is the goal of knowing an answer to the question. If we have Q-elab(α, β) where α is a question, the answer to the second question determines a strict subset of the answers of the first. In this case, a Q-elab move would in effect give us a specification of the SARG behind B's question, which is to get money. A continuation by B in this case would attach not only to the first question but also with a elaborating or result discourse relation to its answer. In Asher & Lascarides (2003)'s terminology, we would *QAP*(1*a*, *yes*) ∧ *Q-elab*(1*a*,4) ∧ *Result*(yes,4). A picture of this discourse structure that results from (1a), the *Yes* response and the continuation (4) is given in figure 3.

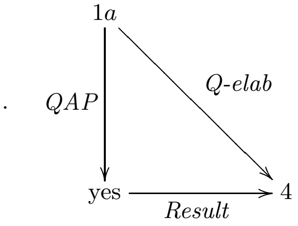


Fig. 3. The Yes response and likely continuation

The question is why these moves are not costly. With the beggar's first question we conventionally associate a SARG, that of getting an answer to his question, but also an ancillary goal of getting money. This incurs a certain cost for the beggar, but once he has committed to the SARG, he incurs little further cost by developing the conversation to see whether the SARG can be met. The *Yes* answer invites a Q-elab elaboration as its response, and enables a first step of B's SARG to get money. We thus postulate that costs of turns that continue to develop a speaker's SARG, once such a development is started but not completed, are low, constituting a single, natural unit of discourse structure, dominated by the original question as in the figure above. Conversely, once a participant's conversational goal has been met or it has been established that it cannot be met, the local discourse structure is closed off.

If P answers *No*, he avoids the potential face-threat and incurs no further negotations. A *No* answer to the question simplifies the structure and does not lead to any further Q-elab moves by B. Given the No answer, there is no way of specifying further any part of the plan to get money. If we assume that a SARG development provides the grounds of a single dialogue structure, then the *No* answer means the termination of that discourse structure, since the SARG underlying the original question can no longer be continued.

There are possible continuations to be sure to the *No* answer. But they range over the whole previous structure. Consider the response by B in (5)

(5) So you can't give me any money

Here is the structure for (1a), the *No* response and the continuation in (5) on the whole structure:

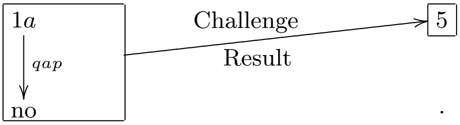


Fig. 4. The No response and challenge

(5) acts as an implicit correction or challenge, which is face threatening and ups the potential costs to B. B could also question P's response with a restatement—he could continue with, *are you sure you don't have any money?* or *Oh come on, I'm sure you have some.* All such moves are also clearly face-threatening, hence more costly, and hence not preferred. These factors in our model are reflected in P's payoffs in Figure 5.

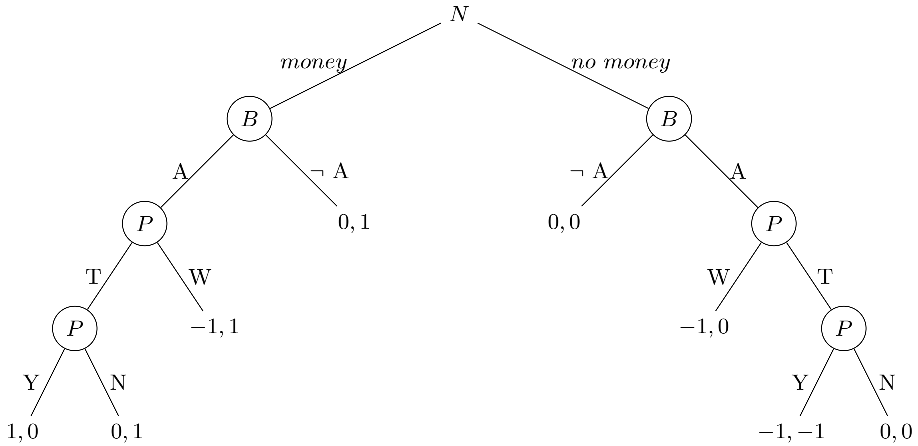


Fig. 5. Game Tree

This game is sub-tree compatible with a sequential trust game (McCabe et al. 2003), as is its solution concept. Whereas a trust game presents a naked model of a request, this game is a step toward the multitude of ways in which agents can defect on or cooperate with each other. While Walk and No are equally rational for P, Walk puts B at a disadvantage. This disadvantage may lead to an unpleasant conversational turn, stemming from a perceived social slight. Underlying this is a generous tit-for-tat model (Rand et al. 2009) that addresses repeated reciprocal interactions under the threat of punishment. The gist is that the incentive to not walk away emerges under repeated interactions as the threat of punishment or probablility of role reversal increases. In a one-shot game, the rationales for P to answer *No* lie in a preference for payoff-dominant outcomes, but, as with other such public goods games, this scales up to repeated games with reputation.

Answering No also has a counterpart in Batesian mimicry (similar situations with incomplete information and the non-aligned preferences of predator and prey), the practice of certain species to mimic the signals of other species that predators know to avoid. As in biology, a Passer-by has incentive to mimic the signals of those without money even in the case that he does.

Our model and argument above depend on a hidden reputation effect. Conversational games are not closed, and so to speak of a one-shot question-answer pair is not really sensible. It is always open to a conversational participant to continue the conversation, to continue the game, or rather to engage in a new game. It is the possible continuations that constrain the addressee to adhere to a basic form of cooperativity, even if he or she is not Grice cooperative, as is the case in the scenario we are considering. In effect conversations essentially enfold the notion of reputation and repeated games because of their open-ended nature. On the other hand, the possibility of "repetition" in conversation occurs only at a very abstract level. We do not have to assume that conversational participants need to repeat the same exchange in order for reputation effects to have their bite. And indeed we should not.

4 Generalization

We can generalize from our example and refine the analysis based on politeness theory. Following conversational analysis or theories of discourse structure, we claim that certain conversational moves introduce an incomplete structure into conversation, and for the conversation to be minimally coherent, the structure must be completed. Questions are such moves, and their natural completions involve some sort of a response. But greetings are also similarly conversation opening moves that can have continuations with little to no content.

(6) a. A: What's up?
b. B: Nothin. You?
c. A: Nothin.
d. OK Cool.

Perhaps certain non–linguistic situations are similar, demanding a linguistic move; as an example of such a situation, consider the "need" to make conversation with the barber when one gets a haircut.[3] A conversation opening move puts a burden on the addressee of the sort we have outlined.

On the other hand, there are discourse structure continuing moves to signal that a certain discourse structure is still open, which allows participants to continue to attempt to satisfy their SARGS. Such moves invite another conversational move from the addressee. In cases like the "Yes" response to the Beggar's question that we considered above, such moves can have potentially deleterious effects for the agent making them. The follow up question *You?* in (6b) is a discourse continuing move as well.

---

[3] Thanks to Robert van Rooij for this point.

Finally, there are moves that provide for a closure of a local (or global) discourse structure. The “No” answer we considered in the Beggar-Passerby game was one example of this. But there are others—acknowledgments followed by some sort of minimal positive evaluation such as in (6d) provide another example. These three kinds of moves, conversation or discourse structure opening moves, continuation moves and structure closing moves, provide another classification of dialogue moves that provides a higher level typology of the more detailed move descriptions of a theory like SDRT.

5 Conclusions

Being minimally cooperative by responding to a question reflects utilities dependent on principles of politeness but its rationale is interestingly different from the rational justification for general politeness conventions abstracted away from considerations of language. As discussed in Quinley (2011), repetition and observation seem important to establishing general politeness conventions. But we do not need to assume the fact that repetitions or observation occurs to justify the rational for minimal linguistic cooperativity. The semantics of dialogue moves together with the open-ended nature of conversation provide the basis for a rational reconstruction of basic cooperativity, even in the absence of explicit repetition or observation.

Basic cooperativity also brings to the fore the notion of safe strategy. A safe strategy minimizes risk to the responder by offering discourse structure closing moves, which are costly for the conversational partner to re-open or add to. These become conventionalized, as part of an expected continuation of a discourse structure that is currently in play. The notion of a safe strategy gives us a heuristic to follow because of the infeasibility of computing analytic solutions to large games with partial information, which almost all conversational games are.

This parallel between discourse structure and politeness strategy as being constrained by processes of mutual recognition of the autonomy of others we find fruitful. In addition, the application of game-theoretic treatments to such phenomena opens up avenues for application and analysis on a more transparent level.

References

Asher, N., Lascarides, A.: Logics of Conversation. Cambridge Univeristy Press (2003)

Asher, N., Lascarides, A.: Strategic Conversation (manuscript, 2012)

Brown, P., Levinson, S.: Politeness: Some Universals in Language Use. Cambrige University Press (1978)

Farrell, J.: Meaning and credibility in cheap talk games. Games and Economic Behavior 5 (1993)

Fehr, E., Fischbacher, U.: The Nature of Human Altruism. Nature 425 (2003)

Homans, G.: Social Behavior as Exchange. The American Journal of Sociology 63(6) (1958)

Jağer, G.: Game-Theoretical Pragmatics. In: Handbook of Logic and Language, 2nd edn. Elsevier (2008)

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

Pinker, Nowak, Lee: The Logic of Indirect Speech. PNAS (2007)

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

Rand, D., Ohtsuki, H., Nowak, M.: Direct reciprocity with costly punishment: Generous tit-for-tat prevails. Journal of Theoretical Biology (2009)

Sacks, H.: Lectures on Conversation. Blackwell Publishers (1992)

Van Rooij, R.: Being polite is a handicap: Towards a game theoretical analysis of polite linguistic behavior. TARK 9 (2003)

Traum, D., Allen, J.: Discourse Obligations in Dialogue Processing. In: Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL 1994), Las Cruces, New Mexico (1994)

Players Who Are Learning How to Play: A Haskell Implementation of Awareness Dynamics

Mauricio Hernandes

Sokendai,
Department of Informatics
`mauricio@nii.ac.jp`

Abstract. Unawareness is analyzed under a framework due to Michael Franke. When a player is facing a game which she is unaware I propose a solution to choose an action to be performed. I discuss a game update raising a need for awareness dynamics, finally a Haskell implementation is proposed helping to sharpen our intuition about this topic. *abstract* environment.

Keywords: Unawareness, Game Theoretic Pragmatics, Haskell.

1 Introduction

We are always playing games, either a traditional game – such as chess or blackjack – or what modern scholars call a game – like dating, buying assets or trusting someone. It's not rare that we catch ourselves in situations where we are not aware of all the rules of a given game (whether by ignorance or imperfect memory). By rules I mean feasible actions that are in line with that game. (Waltzing is not a reasonable action for a chess game, but might be for a date.) To help clarify the notion of awareness consider the following example:

Example 1. Imagine a situation where you had traveled abroad and brought an expensive linguistics book. Now you are coming back to your country and the customs officer wants to charge you the taxes for the item, then he gives you two options; (1) Pay the tax; or (2) Leave the book at the checkpoint. Realizing that you don't have enough money, you leave the book there. A while after, you are telling this to a friend and he asks you "Why didn't you bribe the officer?". The answer would be "I wasn't aware of this possibility".

Unawareness was introduced by [2] in an attempt to solve the traditional epistemic logic omniscience problem (in a few words: assuming that when an agent gets to know a proposition all its consequences are known as well). We introduce unawareness of a contingency when an agent "*...has no mental representation of it [contingency] and consequently lacks all explicit belief about it*" [3].

But when a player is facing a situation with unawareness how does she choose which action to perform?

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 13–26, 2012.

Example 2 (Unawareness version of "Battle of the Sexes"). In the classical example of The Battle of the Sexes, He wants to go to a soccer match and She wants to watch a movie, and both want to do something together rather than disagree and stay at home. However, for the sake of unawareness, imagine that He, secretly from Her, reads on the newspaper that on the same night Madonna is performing in town. He knows that both of them are unconditional fans of the diva and would prefer to attend her concert rather than any other plans. What should He do? How would the game look like now?

In this paper I propose a method to formally decide which action the agents should choose. I use as a framework dynamic games, and on top of that I add a structure that encodes the unawareness. This framework was proposed by [3].

Consider the following example, inspired by an example in [3]:

Example 3 (Prisoner's Dilemma with Super Defection). Imagine now that after the last episode at the checkpoint of your country you were confused with a thief and were arrested, at the same time the police arrested another suspect whom they assumed to be your accomplice. Although you think it is an awful experience, you consider it a good opportunity to practice your skills in game theory. The police interrogates you and your alleged accomplice in separate rooms and make the following offer: (1) if both you and the other suspect testify against each other, you will each receive a three-years sentence (Defect); (2) if both you and the other suspect remain silent, each would receive a two-years sentence (Cooperate); and if one "Defects" and the other "Cooperates", the Defector gets one year in jail and the one who Cooperated gets ten. Given this possibilities you remember the customs office episode and bribe the officer this time and go free on the same day because your unknown accomplice didn't think about this possibility. However, for the sake of our example, let's imagine a situation where you and the same accomplice are arrested everyday and receive the same proposal on and on. When does your accomplice become aware of the bribe action?

In this paper I propose a naive way to perform the update of a game with awareness, and then we can say that the players are learning how to play.

Finally, the last part of this paper is a Haskell[1] implementation for those games, calculating a natural move for the players and updating the model while the players are learning the rules of the game.

In Section 2, I establish notations by introducing the idea proposed by [3]. In Section 3, I take a step forward with those definitions and propose what I will call a simple solution for a game, and I provide a formal method for updating for games with awareness. Section 4 is the Haskell code of this framework; it is divided in two parts in accordance with section 2 and 3, respectively. The code in this paper is self contained, so even if you don't fully understand some parts of it you can copy it and compile; just do not forget the appendix.

---

[1] Haskell is a purely functional programming language.

2 Dynamic Games and Awareness Structures

Unawareness (according to [3]) is a structure with its own properties alongside a game, which we could (ideally) use with any kind of game[2] where we can incorporate the notion of awareness. Whether such an approach is fruitful is still unknown. Following [3], we will work with Dynamic Games which prove to be a good source of examples and applications. All definitions in this section are taken from [3].

Definition 4 ([3] Dynamic Game with Imperfect Information). A dynamic game with imperfect information is a structure

$G = \langle H, <, N, \{A_i\}_{i \in N}, P, A, \{u_i\}_{i \in N}, Pr, \{V_i\}_{i \in N} \rangle$

that consists of (i) a game tree, (ii) a collection of labels, and (iii) a labeling assigning labels to elements of the game tree. In particular:

- $\langle H, < \rangle$ is a game tree:
  - H is a (finite) set of *histories*, or decisions nodes
  - $<$ is a partial order on H such that:
    - there is a unique $<$-minimal element h_0, called the *root*
    - every history $h \neq h_0$ has exactly one *predecessor*, namely the unique $<$-maximal element of the set $\{h' \in H | h' < h\}$
  - elements in $Z = \{h \in H | \neg \exists h' \in H(h < h')\}$ are called *terminal histories*
  - every non-terminal history h has a non-empty set of *sucessors* $H(h)$, defined as the set of all $<$-minimal elements in $\{h' \in H | h < h'\}$
- $\langle N, \{A_i\}_{i \in N} \rangle$ are labels:
  - $N = \{1, 2, ...n\}$ is a set of players with designed player n as Nature
  - A_i is a set of actions for player i

  $\langle P, A, \{u\}_{i \in N} \rangle$ is the *labeling*:
  - $P : H \setminus Z \to N$ is a *player function*
  - $A : H \times H \to \bigcup_{i \in N} A_i$ assigns action labels to each choice as follows:
    - $A(h, h')$ is only defined if $h' \in H(h)$
    - $A(h, \cdot) \mapsto A_{P(h)}$ is an injection
  - $u_i : Z \to \mathbb{R}$ is a *utility function* for all $i \in N, i < n$
    - we write $u_i(z)$ for the i-th component of $u(z)$
  - Pr is a function that gives for each nature move $h \in P^{-1}(n)$ a *probability distribution* $Pr_h \in \Delta(H(h))$ over successors of h
  - $V_i \subseteq \wp(H \setminus Z)$ is a *set of information states* of player i such that:
    - $\bigcup_{i \in N} V_i$ is a partition of $H \setminus Z$
    - if $v \in V_i$ and $h, h' \in v$, then $P(h) = i$ and, moreover, $A(h, \cdot)$ and $A(h', \cdot)$ have the same image sets
    - notice that V_n is a set of singletons and can be treated as $\bigcup V_n$

Say that an action $a_i \in A_i$ is available to player i at choice point v if there is an $h \in v$ and an $h' \in H(h)$ such that $a_i = A(h, h')$. A *behavioral strategy* σ_i for player i is a function from V_i to a probability distribution over actions available to i at each $v \in V_i$. A behavioral strategy profile for game G is a tuple $\sigma = \langle \sigma_1, ..., \sigma_{n-1} \rangle$ of behavioral strategies for each player.

---

[2] Games with perfect information, Bayse Games or signaling games are examples of games.

2.1 Subjective Games (Pruning)

As Game Modelers we are aware of the whole picture (model). The players on the other hand, every time they are unaware of some part of the game they have only a partial view of it. Formally, a partial view of the game is a pruned version of the entire game. Informally, a pruned game G' is a pruned version of G ($G' \subset G$) if G' is obtained from G by removing nodes and keeping labels unchanged as much as possible.

Definition 5 ([3] Pruning). Take two dynamic games with imperfect information (and non-redundant labeling):

$G = \langle H, <, N, \{A_i\}_{i \in N}, P, A, \{u_i\}_{i \in N}, Pr, \{V_i\}_{i \in N} \rangle$

$G' = \langle H', <', N', \{A'_i\}_{i \in N'}, P', A', \{u'_i\}_{i \in N'}, Pr', \{V'_i\}_{i \in N'} \rangle$

G' is a *pruning* of G ($G' \sqsubseteq G$) if the following conditions hold:

- $H' \subseteq H$
- $<' = <_{\upharpoonright H'}$
- $Z' \subseteq Z$
- $N' \subseteq N$, and $A'_i \subseteq A_i$
- $A'(h, h') = A(h, h^*)$, where h^* is the unique element such that $h^* \in H(h)$ and $h^* \leq h'$
- $P' = P_{\upharpoonright (H \setminus Z)}$
- $u'_i = u_{i \upharpoonright Z'}$
- $Pr'_h = Pr_h(\cdot | H'(h))$ for all $h \in P'^{-1}(n)$
- $V'_i = \{v \cap H' | v \in V_i\} \setminus \{\emptyset\}$

2.2 Awareness Structure

To represent players' awareness we will consider a set of worlds, each "containing" a game, and for each player a relation over the set of worlds. We will say that a player is unaware of a game if she can't reach its respective world through her relation.

Definition 6 ([3]Awareness Structure). Let G be a dynamic game with imperfect information with information states $V = \cup_{i<n} V_i$ of all non-nature players. An awareness structure based on G is a tuple $\mathcal{A}_G = \langle W, w_0, \{R_v\}_{v_V}, L \rangle$ such that:

- W is a set of possible worlds,
- w_0 is the actual world (specifying the modeller's view),
- $R_v \subseteq W \times W$ is an accessibility relation for the viewpoint $v \in V$,
- $L : W \to \mathcal{G}$ assigns to each world w a game $L(w)$.

Moreover, $\mathcal{A}_G$ needs to satisfy the following constraints:

Centering: $L(w_0) = G$,
Reduction: if $w R_v w'$, then $L(w') \sqsubseteq L(w)$,

Existence: if v is an information state in game $L(w)$, then there is a world w' such that wR_vw',

Relevance: whenever wR_vw' them v is an information state in $L(w)$, then there is a world w' such that wR_vw',

Introspection: for all v the relation R_v are transitive and Euclidian.[3]

2.3 Modal Game Model

In this section, following what is proposed by [3], we add the element of unawareness to Dynamic games.

Definition 7 ([3]Modal Game Model). Fix an awareness Structure $\mathcal{A}_G$. (For convenience, assume that it is image-finite.) A game model for $\mathcal{A}_G$ is a structure:

$$\mathcal{M}_{\mathcal{A}_G} = \langle W, w_0, \{R_v, P_v\}_{v \in V}, L, \{\sigma_w\}_{w \in W} \rangle$$

such that $\langle W, w_0, \{R_v\}_{v \in V}, L \rangle$ is an awareness structure equivalent to $\mathcal{A}_G$, $P_v : \wp(W) \to \mathbb{R}^{\geq 0}$ is an additive measure function for each v which can be viewed as a credence function over the set of worlds for each v, and σ_w is a behavioral strategy profile for game $L(w)$ such that for all $w' \in R_v(w) : \sigma_w(v) = \sigma_{w'}(v)$.

And define the probabilistic belief of type $\langle w, v \rangle$ as:

$$\pi_{\langle w,v \rangle}(w') = \frac{P_v(R_v(w) \cap \{w'\})}{P_v(R_v(w))} \tag{1}$$

And for each belief type $\langle w, v \rangle$ define behavioral strategy profile $\sigma_{\langle w,v \rangle}$ by:

$\sigma_{\langle w,v \rangle}(v', a) = \sum_{w' \in W} \pi_{\langle w,v \rangle}(w') \times \sigma_{w'}(v', a)$

We say that a type $\langle w, v \rangle$ is rational in $\mathcal{M}$ iff $\sigma_w(v)$ is a best response in game $L(w)$

3 Selecting an Action and Learning the Game

Now, with all the tools of Dynamic games and awareness in mind I want to start building the new ideas of this paper. In the first part of this section I propose a way of choosing actions for our unaware players and give some examples where it seems natural. In the second part I propose a way to make players update their unawareness model between iterations of plays.

I will assume a game model $\mathcal{M}_{\mathcal{A}_G} = \langle W, w_0, \{R_v, P_v\}_{v \in V}, L, \{\sigma_w\}_{w \in W} \rangle$ to be given, and I am not paying attention to how the mixed strategies σ_w were obtained. A sound way to compute the strategy profiles of a game with unawareness is by itself an interesting problem but is out of the scope of this work.

For each belief type $\langle w, v \rangle$ define her belief that v' is playing a as:

$$\theta_{\langle w,v \rangle}(v', a) = \sum_{w' \in W} \pi_{\langle w,v \rangle}(w') \times \sigma_{\langle w',v' \rangle}(v', a) \tag{2}$$

---

[3] The relation R_v be Euclidian means: $w_1 R_v w_2$ and $w_1 R_v w_3$ implies $w_2 R_v w_3$.

What θ says is that if I am a player my belief of what the other player will do is given by what I believe how the other player sees the game.

We define the *expected utility* of action a^* for a type $\langle w, v\rangle$, with $v \in V_i$ by the formula:

$$EU_{\langle w,v\rangle}(\theta_{\langle w_0,v\rangle}, a^*) = \sum_{a\in A}\Big(\prod_{v'\in V\setminus\{v\}} \theta_{\langle w,v\rangle}(v', a)\Big) \cdot U_i(a, a^*) \tag{3}$$

For a player i, we call *simple solution* an action a^* if it maximizes $EU_{\langle w_0,v\rangle}(\theta_{\langle w_0,v\rangle}, x)$, with $v \in V_i$.

Let's work an example making use of this simple solution:

Example 8. Consider once again the example 2 of "Battle of the sexes". For the sake of our example, consider that $\{\sigma_w\}_{w\in W}$ is the Nash Equilibrium of each game in its respective world, i.e., let v_1 and v_2 stand for She and He, respectively. Then $\sigma_{w_0} = (\sigma_{w_0}(v_1), \sigma_{w_0}(v_2))$ where $\sigma_{w_0}(v_1) = \sigma_{w_0}(v_2) = (0, 0, 1)$ and $\sigma_{w_1} = (\sigma_{w_1}(v_1), \sigma_{w_1}(v_2))$ where $\sigma_{w_0}(v_1) = (2/3, 1/3), \sigma_{w_0}(v_2) = (1/3, 2/3)$. With these parameters, for v_1 the set of actions that maximizes equation 3 is $\{C, S\}$, and the same holds for v_2. Note that this idea makes the second player (He) ignore the choice of going to the concert, since it would be in vain to try to attend the event alone because She doesn't know about it. Of course we still need a method to choose between the actions in the solution set, but to establish a sound choice is out of the scope of this work, so when needed I will assume a mixed choice of $\sigma_{\langle w_0,v_i\rangle}(v_j, \cdot)$ for each player.

Fact 9. When applied to games without unawareness the expected utility coincides with standard solutions. To see that, consider a game with no unawareness, for example with $W = \{w_0\}$. And let σ_{w_0} be the Nash Equilibrium for $L(w_0)$. Note that the set of simple solutions for this game are the ones in the support of σ_{w_0}. And this enforces the idea that this solution is natural, because whatever equilibrium concept you impose on a game with unawareness it should coincide with a game where all the players are aware of the whole picture.

To finish this section consider a situation where you are supposed to play the unawareness version of "Battle of the sexes" game every day, for some days in a row. Supposing that He performs the action which She is not aware of. How does She react in the following play? Is the way she reacts common knowledge?

3.1 Unawareness Dynamics

I want to address the idea of how to make players aware, but before that I want to restrict my analysis to specifics games, I want the set of possible actions for each agent to be the same ($A_i = A_j, \forall i, j$). It might seem to be too restricted but you could be more flexible and just ask for a function $f_{ij} : A_i \to A_j$.[4]

What happens with our players when they find out that they were not entirely aware of the game? Consider the following, quite famous, example:

---

[4] For instance, consider a signaling game where given a message "m" sent by the Speaker the Hearer could interpret it as "$\neg m$".

Example 10 (CryWolf). In the fable "The Boy Who Cried Wolf" the main character, the boy, loses credibility after lying about a false wolf that would be eating the sheep of his village, and then when he actually spotted a wolf trying to eat those animals nobody believed him; the flock was doomed. Now imagine a signaling game modeling the situation before the lie, the villagers were unaware that the boy would be playing a game where the Sender would be lying. Furthermore, after lying, the villagers were unaware that the boy would be playing a game where the Sender would be telling the truth.

Before analyzing example 10 observe that naively one could update the game in the following way: given a modal game model $\mathcal{M}_{\mathcal{A}_G} = \langle W, w_0, \{R_v, P_v\}_{v \in V}, L, \{\sigma_w\}_{w \in W} \rangle$ and actions (a, b) (in our example lying and believe respectively) played on the first round, build the second round in such a way that every game $L(w)$ has the actions a and b as possibles actions. A way to perform it could be, for each $L(w)$ to make $A'_1 = A_1 \cup \{b\}$ and $A'_2 = A_2 \cup \{a\}$, what could generate redundancy.

Now back to the CryWolf example, which game was the boy playing? Our first lesson from the lost sheep is that after the first update they weren't playing the same game (otherwise the boy would foresee the city's behavior and lie, making them believing him). The game in which the boy is aware is a game where the mixed strategies don't match reality (an awareness of $\sigma_{\langle w', boy \rangle}(city, believe) = 2/3$ while $\sigma_{w_0}(city, believe) = 1/4$ would fit the description from the boy's point of view). From the city's point of view we could not use a framework of unawareness once they realize that the boy is able to lie, and in that case the mixed strategy adopted would be the one associated to w_0. To avoid the signaling game structure we can assume that Nature is playing and choosing sender's type, i.e., choosing the probability of whether the wolf exists or not).

To try to motivate an unawareness perspective to this problem I want to modify a little the CryWolf problem for a Cry n-Wolf problem. Suppose that the boy's role in this situation now is to say how many wolves exist (maybe without considering an upper bound for the possible number of wolves). For n wolves, the extensive form of the game (without unawareness) Nature chooses from 0 to n, the boy and the city too, thereby increasing the complexity. Although the use of awareness that are in this problem makes the description incomplete, it is meant to prune options less likely to happen instead of cutting the tree arbitrarily. However, how to do that precisely is left as an open question here.

From now on we will consider the computational aspects of implementing this framework.

4 Code

I see computational implementation of unawareness as having a double role in this topic. First, it helps justify the need for formalizing this ideas. Second, it helps the modeler sharpen her intuitions on the topic. Another consequence of it is the possibility of areas within Artificial Intelligence to take advantage of studies in awareness.

Haskell was chosen for being a convenient language for treatating game theoretic concepts and the choice was also motivated by the book [1] which gives a good introduction to Haskell applying it to formal semantics of natural language.

In my program the main module is called DynamicGame and I import a module of set and statistical operations given in the appendix.

```
module DynamicGame  where
import SetsStatistics(nub,Dist,intersect,h,listMinus,vProd,
    sublist)
```

As mentioned earlier a dynamic game with Imperfect Information has the following form: $G = \langle H, <, N, \{A_i\}_{i \in N}, P, A, \{u_i\}_{i \in N}, Pr, \{V_i\}_{i \in N} \rangle$.

4.1 Dynamic Game

The data set in our module for dynamic games is given by:

```
data Player = C Int | P Int deriving (Show, Eq, Ord)
type Players = [Player] -- N
data Action = A String deriving (Show, Eq, Ord)
type Actions = [Action] -- A
data Info = V Int Int deriving (Show, Eq, Ord)
type InfoSet = [Info] -- V
type Pr = Node -> Node -> Rational -- Ideally Node -> (Dist
    Node)
data Node = NonTerminal Players Actions Info [Node]
          | Terminal Players Actions [Rational]
          deriving (Show, Eq, Ord)
type Game = Node
```

A Player is either a person (distinguished by an integer) or the Chance (Nature) and in this case the Integer in `C Int` is used for two things: (1)Positive numbers represents how likely it is for Chance to perform an action that leads to that node; and (2) 0 (Zero) is used to mark Chance's turn. An action is simply a type constructor followed by a String (`A String`) which stands for the name of the action. Information states are denoted by InfoSet, while `V Int Int` encloses the number for its information state and the second Integer for the player in it. A distribution over Chance's plays is given by `Pr`, a type that given a node on which Chance moves, we get back a distribution over the nodes that can be reached by Chance from that point.

A node is either a terminal node or a non-terminal node containing all the information on the game which is requested by the code as it is performed on the following way: (1)`Players` denote the set of players that perform an action in that node; (2) `Actions` is the set of actions performed by the previous player to reach the current node; (3) `Info` stands for the information set of its node; (4) `[Node]` is the children of a non-terminal node; and (5) `[Rational]` is the payoff associated to that terminal node. Finally, a game is given by a tree associated to its extensive form.

```
prC::Pr
prC a b | and [(a `before` b) ,((chancePlay b) /= 0)] = prC' a
     b
        | otherwise     = 0

prC'  ::  Pr
prC'  (NonTerminal _ _ _ x)  b = (toRational (chancePlay b))
                                  / (toRational (sum (map chancePlay
                                       x)))

chancePlay::  Node -> Int
chancePlay  (NonTerminal p _ _ _) = sum (map chanceCheck p)
chancePlay  (Terminal p _ _) = sum (map chanceCheck p)

chanceCheck::  Player -> Int
chanceCheck  (P _) = 0
chanceCheck  (C i) = i
```

This block of code recovers the Pr distribution from the game. `PrC` checks if the second node is the child of the first node and if Chance is playing. `PrC'` is the numerical function that has the behavior of a distribution. Given a node `ChancePlay` returns the likelihood of that node to be played. And finally, `chaceCheck` just gives the numeric value of likelihood.

```
actions  ::  Info -> Game -> Actions
actions w g = nub  (actions' w g)

actions'  ::  Info -> Game -> Actions
actions' v (NonTerminal _ a v' x) |
                              v == v' = (concat (map actionAux x)) ++
                                          (concat (map (actions' v)
                                              x))
                                        | v /= v' =
                                        concat (map (actions' v) x)
actions' v (Terminal _ a _) = []

actionAux::Node -> Actions
actionAux (NonTerminal _ a _ _) = a
actionAux (Terminal _ a _) = a
```

Given a set of information state (`V Int`) `actions (V Int)` is the function that finds the actions which `V Int` can perform in game `g`.

```
infos :: Game -> InfoSet
infos g = nub (infos' g)

infos' :: Game -> InfoSet
infos' (NonTerminal _ _ v x) = v : concat (map infos' x)
infos' (Terminal _ a _) = []

before :: Node -> Node -> Bool -- <
before (Terminal _ _ _) _ = False
before (NonTerminal _ _ _ x) y = y ‘elem‘ x
```

`infos` is the function that returns the information sets of a game `g`. `before` checks for imediate successor

```
utility :: Node -> [Rational]
utility (Terminal _ _ u) = u
utility (NonTerminal _ _ _ _) = []

utilityI :: Int -> Node -> Rational
utilityI i (NonTerminal _ _ _ _) = error "Non Terminal Node"
utilityI i (Terminal _ _ u) = last (take i u)

utilityV :: Info -> Node -> Rational
utilityV _ (NonTerminal _ _ _ _) = error "Non Terminal Node"
utilityV (V _ i) t = utilityI i t

utilityA :: Game -> Info -> [Action] -> Rational
utilityA (Terminal p a x) v _ = utilityV v (Terminal p a x)
utilityA (NonTerminal p a' v x) v' (a:as) = utilityA (giveStep
    ((NonTerminal p a' v x)) a) v' as

giveStep :: Node -> Action -> Node
giveStep (NonTerminal _ _ _ x) a = head ( filter (\t -> a ‘
   elem‘ (actionAux t) ) x)
giveStep _ _ = error "Not a terminal"
```

This last block of code handles the utilities of the game, `utility` returns the vector of utilities while `utilityI` and `utilityV` return the payoff of player i and information set $v \in V_i$, respectively.

```
whosPlay :: Node -> Players -- P
whosPlay (NonTerminal p _ _ _) = z p
                        where z [] = []
                              z (p:ps) = case p of
                                         P _ -> p : (z ps)
                                         C 0 -> p : (z ps)
                                         C _ -> z ps

whosPlay (Terminal _ _ _) = []
```

The `whosPlay` function takes a node and returns the set of players performing an action in that node, the `case` is done to get rid of Nature's likelihood (`prC`)

4.2 Awareness Structure

Remind that $\mathcal{A}_G = \langle W, w_0, \{R_v\}_{vV}, L\rangle$, then its type is:

```
data World =W Int deriving (Show, Eq, Ord)
type Worlds = [World]
type Relation = Info -> [(World, World)]
type L = World -> Game
```

Defining a data type for Franke's Awareness Structure is straight forward. `World` is just a type constructor and a interger, `Worlds` is a short notation for set of worlds, and a relation for an information set is a set of pair of worlds (and the as usual, $w`R_{V_i}w$ iff $(w`, w) \in R_{V_i}$). `L` is the type of functions that given a world returns a game.

```
circRel :: Relation -> Info -> World -> Worlds
circRel rel v w = nub [w' | (w'',w') <- (rel v), w''==w ]
```

For the sake of simplicity we usually state $\{w'|(w, w') \in R_{V_i}\}$ by $R_{V_i}(w)$, so `circRel` is a function that lets us handle this issue returning a list of worlds that the information set `v` can reach with the relation `rel` from `w`.

4.3 Modal Game Model

As previously said $M_{AG} = \langle W, w_0, \{R_v, P_v\}_{v\in V}, L, \{\sigma\}_{w\in W}\rangle$.

```
type Cred = Info -> Worlds -> Rational --P
type Sigma = World -> Info -> Dist Action -- \sigma
```

```
probBelief ::Cred -> Relation -> (World, Info) -> World ->
    Rational
probBelief credence rel (w, v) w' =
                    (credence v ((circRel rel v w) `intersect` [
                        w']))
                    / (credence v (circRel rel v w))
```

The last function is an implementation of:

$$\pi_{w,v}(w') = \frac{P_v(R_v(w) \cap \{w'\})}{P_v(R_v(w))} \tag{4}$$

```
behBelief :: Sigma -> Relation -> Cred -> (World, Info) ->
    Info -> Action -> Rational
behBelief s rel c (w,v) v' a = (sum (map (\t ->(probBelief c
    rel (w,v) t) * (h (s t v') [a])) (circRel rel v w) ))
```

`behBelief` computes:

$$\sigma_{\langle w,v\rangle}(v',a) = \sum_{w'\in W} \pi_{w,v}(w') \times \sigma_{w'}(v',a) \tag{5}$$

```
subjectBel :: Worlds -> ((World, Info) -> World -> Rational)
    ->
               ((World, Info) -> Info -> Action -> Rational)
                   ->
               (World, Info) -> Info -> Action -> Rational
subjectBel ws probBel behBel (w,v) v' a =
  sum ( map ((\w' -> (probBel (w,v) w')*(behBel (w',v') v' a)
      )) ws)
```

`subjBel` is the function:

$$\theta_{\langle w,v\rangle}(v',a) = \sum_{w'\in W} \pi_{\langle w,v\rangle}(w') \times \sigma_{\langle w',v\rangle}(v',a) \tag{6}$$

```
eu::Game ->
      (World, Info) ->
      ((World, Info) -> Info -> Action -> Rational) ->
      Action -> Rational
eu g (w, v) s a = sum
      ( map (\t -> (z t) * (utilityA g v t)) (vActionProd g v
          a) )
               where z a' = product
  (map (\v' ->s (w, v) v' (ai v' a')) ( (infos g) ‘listMinus‘
      [v]))
                           ai vi a= case vi of
                                  V i j -> head (drop (j-1) a
                                      )

vActionProd :: Game -> Info -> Action -> [Actions]
vActionProd g v a = vProd (subs v a [actions w g | w<-(infos g
    )])
                       where subs (V _ i) a x = (take (i-1) x
                           ) ++ [[a]] ++ (drop i x)
```

The expected utility (`eu`) is given by:

$$EU_{\langle w,v\rangle}(a^*,\sigma_{w,v}) = \sum_{a\in A}\Big(\prod_{v'\in V\setminus\{v\}} \sigma_{\langle w,v\rangle}(v',a)\Big)\cdot U_v(a,a^*)) \tag{7}$$

and `vActionProd` computes $\Pi_{i\in I}A_i$ such that $i = jA_j = \{a\}$

4.4 Naive Update

When a player becomes aware of a new action she expands her previous perception of reality, but from our perspective she has as subjective game a pruned version of the game ($L(w_0)$), which coincides with the game where she is completely aware. In details it is:Let `g` be a game and let `a` and `b` be action to be updated by player `p1 = P 1` and `p2 = P 2`, respectively. Let A and B be the set of actions of game `g'` in $L(w')$ for player `p1` and `p2`, respectively. Then let `aS` and `bS` be the set of actions to be updated in `g'`. Then the updated version of `g'` is given by:

`> prune g [(p1,aS), (p2,bS)]`

Where `prune` is given by:

```
prune :: Game -> [(Player,Actions)] -> Game
prune (NonTerminal p a v x) pAs = NonTerminal p a v (wrt x pAs
      p)

wrt :: [Node] -> [(Player,Actions)] -> Players -> [Node]
wrt x pAs p = concat (map (wrt' pAs p) x)

wrt' :: [(Player,Actions)] -> Players -> Node -> [Node]
wrt' pAs p (Terminal ps a u) | all (\t ->(a `sublist` (wrtAux
     t pAs))) p  = [Terminal ps a u]
                                    | otherwise      = []
wrt' pAs p (NonTerminal p' a v x) | all (\t ->(a `sublist` (
     wrtAux t pAs))) p = [NonTerminal p a v (wrt x pAs p' )]
                                            | otherwise = []

wrtAux::Eq a => a -> [(a,[b])] -> [b]
wrtAux p [] = []
wrtAux p (x:xs) | p == (fst x) = snd x
                |otherwise = wrtAux p xs
```

5 Appendix

```
module SetsStatistics where
type Function a b = [(a,b)]
type Prob = Rational
type Dist a = Function a Prob
type Event a = a -> Bool

uniform :: [a] -> Dist a
uniform [] = []
uniform x = uniformAux x (toRational (length x))
uniformAux :: [a] -> Rational -> Dist a
uniformAux [] _ = []
uniformAux (x:xs) r = ((x,(1/r)):(uniformAux xs r))
infix 8 ??
--(== t1) ?? pr -- gives you the value of pr(1)
(??) :: Event a -> Dist a -> Prob
(??) p = sumP . filter (p . fst)
```

```
sumP :: Dist a -> Prob
sumP = sum . map snd
--Pr(x)
h:: (Eq a ) => Dist a -> [a] -> Prob
h p x = (`elem` x) ?? p
sublist :: (Eq a) => [a] -> [a] -> Bool
sublist x y = all (\z -> z `elem` y) x
intersect:: Eq a => [a] -> [a] -> [a]
intersect = intersectBy (==)
intersectBy:: (a -> a -> Bool) -> [a] -> [a] -> [a]
intersectBy eq xs ys =  [x | x <- xs, any (eq x) ys]
nub:: Eq a => [a] -> [a]
nub [] = []
nub (x:xs) = x:nub (filter (\y -> (y /= x)) xs )
listMinus:: (Eq a) => [a] -> [a] -> [a]
listMinus bigX bigY = [x | x <- bigX, not (x `elem` bigY)]
vProd :: [[a]] -> [[a]]
vProd x = z
              where z = foldl f k l
                    k = head (wrap x)
                    l = tail (wrap x)
                    f r t = [a++b| a<-r, b<-t]
wrap:: [[a]]  ->[[[a]]]
wrap [] = []
wrap (x:xs) =  (wrap' x):(wrap xs)
wrap'::[a] -> [[a]]
wrap' [] = []
wrap' (x:xs)= (wrap'' x):(wrap' xs)
wrap'' :: a -> [a]
wrap'' x = [x]
```

References

1. van Eijck, J., Unger, C.: Computational Semantics with Functional Programming. Cambridge University Press (2010)
2. Fagin, R., Halpern, J.Y.: Belief, Awareness and Limited Reasoning. Artificial Intelligence 34, 39–76 (1988)
3. Franke, M.: Modal Models for Game with Unawareness. University of Tüingen (2010) (unpublished manuscript)
4. Franke, M.: Signal to act: Game Theory in Pragmatics, PhD Thesis (2009)
5. Osbourne, M.J., Rubstein, A.: A course in Game Theory. MIT Press (1994)
6. Shoham, Y., Leyton-Brown, K.: Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations. Cambridge University Press (2010)

Emphatic NPI/FCI and Adversative Discourse Relations, a Probabilistic Approach[*]

Margot Colinet[1] and Grégoire Winterstein[2]

[1] Université Paris Diderot-Paris 7, Laboratoire de Linguistique Formelle
`margot.colinet@linguist.jussieu.fr`
[2] Université Paris 8, Laboratoire Structures Formelles du Langage
`gregoire.winterstein@linguist.jussieu.fr`
`http://www.linguist.univ-paris-diderot.fr/~gwinterstein/`

Abstract. This paper deals with the discursive effects of the use of *emphatic* Negative Polarity Items and Free Choice Items such as *any*. In this work, we show that the use of the emphatic *any*, be it an NPI or an FCI, has a direct effect on the introduction of subsequent discourse segments. Our theoretical observations are backed up by experimental results. To account for the data, an explicit link between two probabilistic approaches to natural language semantics is proposed. The first one deals with the semantics of NPIs and FCIs, respectively van Rooy (2003), Jayez (2010), and the second one tackles the interpretation of discourse markers, Merin (1999). It is shown that, modulo some formal tinkering, the two accounts interact nicely together to explain the data.

The meaning of NPIs, FCIs or ambivalent items such as *any* has long been studied, mainly in terms of denotation and alternative based semantics. The discursive effects of these elements have been less discussed. It is the point of this paper to investigate some of the discursive effects conveyed by the use of *any*.

Section 1 sets up the empirical domain we are interested in. The results of an experimental investigation are presented. They confirm that the data we study is stable and coherent. They also give an experimental basis for distinguishing *emphatic* uses of the element *any*.

In Sect. 2, various approaches to the semantics of NP/FCI are detailed and Sect. 3 does the same with the semantics of the discourse marker *but*. The analyses proposed share the characteristic of being based on probabilistic models: van Rooy (2003), Jayez (2010) for the semantics of NPIs and FCIs respectively, and Merin (1999), Winterstein (2010) for the semantics of discourse markers.

This enables us to formally link the two kinds of approaches in Sect. 4. To our knowledge, there is so far no attempt that tries to establish an explicit link between such accounts. One of the results of this paper is thus to evaluate the merits of integrating both propositions.

---

[*] Many thanks to M. Mameni, E. Henestroza for their native competence and to Y. Choi and D. Tribout for fruitful discussion.

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 27–39, 2012.

1 The Empirical Landscape of NP/FCIs and Adversative Relations

This section is empirical. Its main point is to show that the use of emphatic NPIs or FCIs carries with it some discursive effects. These effects can be observed by the (im)possibility to use certain discourse markers to introduce a subsequent discourse segment once an NP/FCI has been used.

1.1 Emphatic NP/FCIs

The prototypical example of NP/FCI that we use in this work is given by *any*. In Sect. 2 we give a detailed overview of some of the proposals that have been proposed to account for the semantics of this item. In this section, we only give an intuitive description of what *any* marks in linguistic contexts where its use would not be mandatory (i.e. in cases where we shall qualify the use of *any* as being *emphatic*).

Roughly, we call an item *emphatic* when its use in a certain context is not grammatically required but conveys discursive effects, see the following examples:

(1) a. I was lost all alone in the middle of a desert I didn't have any idea where to go.
    b. *I was lost all alone in the middle of a desert I didn't have some/an idea where to go.

(2) a. I was lost all alone in the middle of a desert I was lucky that I got any help (at all)!
    b. I was lost all alone in the middle of a desert I was lucky that I got some help!

The comparison between (1-a) and (1-b) on one hand and (2-a) and (2-b) on the other hand shows a contrast between two kinds of use of the NPI *any*. In (1) the use of *any* is grammatically required since the use of another neutral determiner, like *some* or *a*, instead is odd. In (2), the use of *any* is not grammatically required since *some* is also grammatical, therefore *any* must have been used in order to induce particular argumentative effects. Furthermore, the use of *any* in (2-a) sounds more natural to English native speakers if it is stressed or modified by the emphatic discourse marker *at all*. We will come back to the definition in more detail in section 2.

The difference between the interpretations of (2-a) and (2-b) relies on the nature of the help provided to the speaker. In (2-b), the speaker received substantial help, whereas in (2-a) the speaker received the minimal amount of help.

1.2 Observations

Our main observation pertains to the contrast observed in (3).

(3) a. #I'm glad you got us any tickets at all, but they're not front row.
b. I'm glad you got us tickets, but they're not front row.

Whereas the version in (3-b) appears natural enough, the use of an NP/FCI in the noun phrase in (3-a) seemingly forbids the speaker to criticize the placement on the tickets. The second segment is introduced by the connective *but* which in this case must receive a denial of expectation (also called argumentative) interpretation rather than a purely contrastive one (see Lakoff (1971), Umbach (2005), Winterstein (2010) for more about the distinction between these uses). This denial of expectation interpretation entails that the conjunct introduced by *but* must go against or contradict a conclusion that is made manifest by the first conjunct.

Thus, the intuitive explanation for the degraded character of (3-a) is that the use of *any* marks that the speaker is glad of his tickets, regardless of their nature. Introducing an exception to that with the use of *but* goes against the conventionally marked equality between the tickets, which makes the speaker sound dissonant.

1.3 Experimental Approach

In order to confirm our intuitions about the data presented above, we ran an experiment on native English speakers.

Experimental design. Twenty-six subjects agreed to participate in an online judgment task. They did not receive any compensation for their time. The subjects were asked to judge the naturality of sentences appearing on their screens. Naturality was judged by means of a scrolling bar without explicit graduation, except for the mentions *Weird* at the far left side and *Natural* at the far right side. The score on the bar translated to a figure between 0 and 100. In total, the subjects were shown fifteen screens, a third of which contained fillers. The sentences were presented after an introduction explaining the expected task, with examples to illustrate this task.

The target sentences were given in pairs: one that contained an emphatic item (usually *any*), and another that did not (labeled as "bare" sentences). Both versions included an exceptive segment introduced by *but*. It was assumed and confirmed by native speaker (but not tested) that without the *but*-introduced segment both versions were equally acceptable, albeit with a difference meaning. For example, both versions of (3) were presented on the same screen. The relative order of the sentences on one screen was randomized, and the relative order of the screens was also randomized, with a homogeneous distribution of fillers between the targeted items.

Results. The main results are presented on Fig 1 : the given figures correspond to the average of the scores attributed to the sentences of the "bare" type and the ones that contained an NP/FCI, in emphatic and non emphatic contexts.

As can be seen on the figure, the difference appears much larger in the emphatic contexts compared to the non-emphatic. The following table shows the

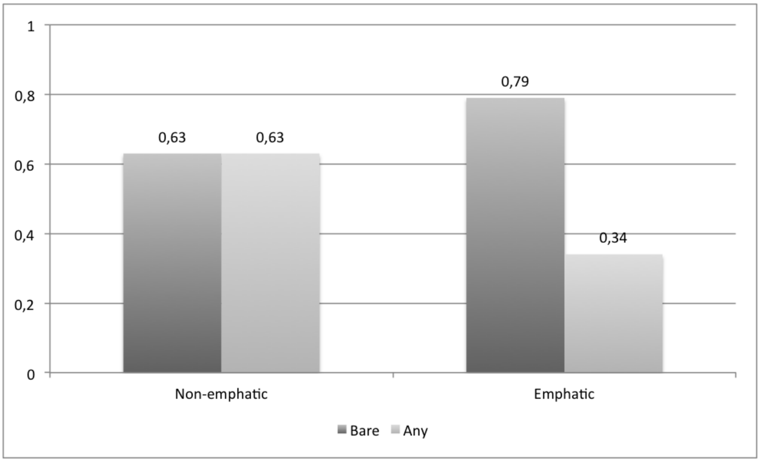


Fig. 1. Comparison between NP/FCI and bare sentences in emphatic and non-emphatic contexts

Table 1. Significance of the differences between NP/FCI and bare sentences

	Test statistic	p-value
Non-emphatic	7501	0.498
Emphatic	491.5	3.124e-15

results of the Mann-Whitney/Wilcoxon test. The results show that the difference is highly significant in the emphatic case.

The fillers provided the baseline for what is an acceptable and an agrammatical sentence. Acceptable sentences received an average score of 0.84 and agrammatical ones the average score of 0.08. In both cases, the standard deviation was low, confirming the non-controversial status of these sentences.

Discussion. While still preliminary, the experimental results confirm our intuitions regarding the interplay of the emphatic use of items such as *any* and the subsequent introduction of an exception with a *but*-segment. This strongly suggests that NP/FCI in emphatic contexts have a bearing on the discourse structure.

A more comprehensive survey would require to test the pairs of sentences without the *but*-segment in order to verify that adding the exception does not significantly affect the score of the "bare" sentences whereas it does for the NP/FCI sentences.

2 Probabilistic Accounts for NPIs and FCIs

Since Kadmon & Landman (1993) NPIs and FCIs are taken to have the same meaning and the same function in their respective licensing contexts. The authors assume that NPIs, FCIs or ambivalent items, like *any*, obey two constraints: a *widening* and a *strengthening* constraint. This analysis relies on a comparison between common noun phrases (CNPs) hosting plain indefinites (like *an N* or bare plurals) and those hosting an NPI or an FCI (like *any N(s)*). Indeed, CNPs of the form *any N(s)* are said to range over a wider quantificational domain than plain indefinite CNPs, including in their quantificational domain 'extreme cases' or 'exceptions' (*widening* effect). Those widening-based CNPs are licensed only if the statement of the sentence they occur in entails the statement of the same sentence with a plain indefinite CNP (*strengthening* requirement), see (4) and (5).

(4) Negative polarity *any*
YOU: - Will there be French fries tonight?
ME: - No, I don't have [potatoes]$_D$
YOU: - Have you just a couple of potatoes that I could take and fry in my room?
ME: - Sorry, I don't have [ANY potatoes]$_{D'}$

(31) in Kadmon & Landman (1993)

(5) Free choice *any*
A: - [An owl]$_D$ hunts mice.
B: - A healthy one, that is?
A: - No, [ANY owl]$_{D'}$. (38) in Kadmon & Landman (1993)

Where:

- D is a subset of D' (*widening*).
- S$_{D'}$, the host sentence of the NP ranging over the enlarged domain D', entails S$_D$, the host sentence of the NP ranging over D(*strengthening*).

This analysis has been really successful and many accounts on NPIs and FCIs are built on Kadmon & Landman (1993)'s proposal. Among them, both van Rooy (2003) and Jayez (2010) respectively reinterpret *strengthening* and *widening* constraints in a probabilistic framework.

2.1 van Rooy (2003): ***Strengthening*** as ***Entropy***

It has been already noted by Kadmon & Landman (1993) that for two assertions *strengthening* has to be defined in terms of *logical entailment*: an assertion hosting an NPI logically entails the same assertion with a plain indefinite; but the authors had already noted that this definition does not hold anymore for questions. A question hosting an NPI cannot logically entail the same question without the NPI. van Rooy (2003) observes that the only way one could understand the notion of entailment between two questions is that a question Q' entails another question Q if any complete answer to Q' is also a complete answer to Q. But van Rooy (2003) demonstrates that for two questions Q' and Q, with Q' stronger than Q because it ranges over a wider set of possible answers than Q, it is not true that any complete answer to Q' will be a complete answer to Q, see van Rooy (2003) for more detail.

van Rooy (2003) takes as starting point the commonly held view that NPIs are used in questions in order to reduce the bias towards a negative answer. For example, in a scenario where a speaker does not consider it to be highly probable that his addressee has ever been in China, she would better ask (6-b) than (6-a).

(6) a. Have you been in China (recently)?
b. Have you *ever* been in China (in your life)? (11) in van Rooy (2003)

Using the NPI *ever*, the speaker enlarges the domain of situations over which the question ranges. Intuitively, *widening* makes a question more general and in our scenario less biased. More precisely, the *widening* effect in a negatively biased question consists in increasing the probability of the positive answer and decreasing the probability of the negative one. In order to formalize this idea, van Rooy (2003) borrows from Information Theory the notion of *entropy*. In this framework, *entropy* is a measure of uncertainty on the result of a certain experiment that is calculated on the sum of the probability of each possible outcome. Intuitively, if one outcome is more probable than another, you have more information on the result of your experiment and *entropy* is low but if all possible outcomes are equiprobable you are as ignorant as possible and *entropy* is maximal. According to van Rooy, an NPI makes its hosting question stronger in the sense that it increases its entropy and makes its possible answers, on average, less expectable, that is to say more informative.

2.2 Jayez (2010): *Widening* as *Entropy*

Jayez (2010) also makes use of the notion of *entropy* but in order to reinterpret the notion of *widening*. According to many authors *widening* is far too strong a constraint that doesn't hold in many kinds of contexts. For example, this constraint doesn't hold anymore when the domain over which the NP ranges is contextually restricted (7), that is to say cannot be widened.

(7) a. You can pick [an apple]$_D$ in this basket
b. You can pick [any apple]$_{D'}$ in this basket Vendler (1967)
Where $D = D'$

According to Jayez (2010) the so-called *widening* effect does not arise when the quantificational domain is enlarged but rather when all the alternatives in that domain are considered to be equivalent with respect to the satisfaction of the proposition. Here again the *widening* constraint is interpreted as an increase of the equivalence between all domain alternatives over which an NP like *any N* ranges and is formalized in terms of *entropy*.

2.3 Unification of the Probabilistic Analyses

van Rooy's analysis aims at an interpretation of strengthening applied to the specific case of NPIs. On the other hand Jayez's main interest is on FCIs and their widening effect. Kadmon and Landman's foundational claim is that a unique analysis for the NP and FC uses of *any* is possible and is given in terms of a two-sided strengthening/widening constraint (where widening and strengthening are dependent). Since van Rooy and Jayez base their analyses on the Kadmon and Landman account, it seems natural to try to unify the two accounts.

Under Jayez's analysis, the entropic effect is on the probability for each possible value of an FCI to satisfy some property. van Rooy's entropic effect applies on the probability to answer *yes* or *no* to a polar question hosting an NPI. If we adopt both analyses, it entails that the entropic effects occur on two distinct levels. The case of (8) illustrates this.

(8) a. I'm glad you got us any tickets.
b. I'm glad you got us tickets.

In (8-a) the use of *any* marks that, the speaker had a low expectation about being glad to get tickets, i.e. that there was a bias in favour of him not being glad of getting tickets. By using *any*, all tickets are considered to be equally satisfactory for the speaker (which matches Jayez's analysis). This has the effect of raising the probability of the speaker being glad to get tickets, i.e. it reduces the bias between him being glad or not glad (this parallels van Rooy's proposition regarding the effect of NPIs in questions).

It is important to notice that the aforementioned accounts are especially meaningful on the specific case of *emphatic* NP/FCI, as formally defined in the next section.

2.4 Emphatic NPIs

An analysis in terms of *widening* and *strengthening* fits particularly well the description of the meaning of NPIs or FCIs that are used emphatically. Indeed, Kadmon & Landman (1993)'s key examples are typical examples of *emphatic* items. In their examples, the licensing contexts of NP- and FC-*any* are dialog contexts where (i) *any* must be stressed and (ii) it is used in order to provide a correction to a misinterpretation of a previous negative or generic sentence, see (10) and (9).

(9) a. I don't have potatoes
b. I don't have ANY potatoes!

(10) a. An owl hunts mice
b. ANY owl hunts mice!

By *emphatic* we refer to NPIs or FCIs that are used within an NP in a context where a neutral NP could have been used instead, that is a bare plural NP or a singular *a N* NP. NPIs and FCIs are not always emphatic, at least in English, since they might be required by the computation of the meaning of their host sentence. In the following sentences, an NP of the form *any N* is required and a more neutral NP gives rise either to the ungrammaticality of the sentence (11) or to a specific reading of the NP under question, see (c) in both (12) for the NPI use and (13) for the FCI use:

(11) ?? I didn't have an idea where to go.

(12) a. Mary doesn't teach any class this semester.
b. Mary doesn't teach a/some class this semester
c. $\exists x(C(x) \wedge \neg T(Mary,x))$

(13) a. Mary has to marry any doctor
b. Mary has to marry a/some doctor
c. $\exists x(D(x) \wedge \Box M(Mary,x))$

The conditions under which an NPI or an FCI might be used emphatically seem to depend on many factors relative to a given language and that we cannot present in detail in this work. Just as an illustration, we can compare the use of the English negative polar pronoun *anything* with the French negative polar pronoun *quoi que ce soit*. The English NPI *anything* sounds neutral in negative sentences, while *nothing* is emphatic. We observe the opposite scenario in French where the N-word *rien* is neutral while *quoi que ce soit* is emphatic:

(14) a. I didn't eat anything
b. I ate nothing

(15) a. Je n' ai rien mangé
I NEG_1 have nothing eaten
'I didn't eat anything'
b. Je n' ai pas mangé quoi que ce soit
I NEG_1 have NEG_2 eaten anything
'I didn't eat ANYTHING'

It is not clear how our distinction between emphatic and non-emphatic NPIs or FCIs would match Krifka (1995) or Zwarts (1996)'s hierarchies of NPIs. First of all, contrary to Krifka (1995) and Zwarts (1996) we take into account the emphatic uses of FCI. Our distinction does not necessarily match Krifka's distinction between stressed and unstressed uses of *any-* forms. In other words, we have not tested whether the uses of emphatic *any* should necessarily be stressed and/or modified by a discourse marker such as *at all*. We will also have to verify if our proposition matches the distinction that Zwarts (1996) proposes between weak and strong NPIs on the basis of their licensing contexts.

3 Probabilistic Accounts for Discourse Markers

The probabilistic approach to discourse markers we will consider here corresponds to the one proposed by Merin (1999). Merin's goal was to give a concrete interpretation to the notion of *argumentation* as pioneered by Anscombre & Ducrot (1983).

One of the most famous example of argumentative item is given by the adversative connective *but*. Its description goes as follows (with a definition of “is an argument” to be precised below):

(16) A sentence of the form A *but* B is felicitous iff. there exists an argumentative goal H such that:

- A is an argument in favor of H
- B is an argument against H

This characterization makes the clear assumption that the core meaning of *but* is its denial of expectation reading. While the question of the “basic” meaning of *but* is disputed in the literature Blakemore (2002), Sæbø (2003), Umbach (2005), Winterstein (2010), we will stick to the proposition above since all the uses we cover are denial of expectations which means that this description easily fits our needs. To see how (16) covers cases of semantic contrast, the reader is invited to consult Winterstein (2010).

To give a concrete interpretation of the relation of *being an argument*, Merin proposes a probabilistic framework inspired by the works of Carnap. In this setting, the relation of being an argument is of a Bayesian nature. If A is an argument in favor of H, it means that the knowledge of A increases the subjective probability of H: $P_{S \oplus A}(H) > P_S(H)$[1] (it is then said to be positively *relevant* to H). If this knowledge decreases the same probability, then A is an argument against H. The probability measure is one on epistemic states and corresponds to the intuitive speaker judgement about the likelihood of various situations. To measure effectively the impact of an assertion to the probability of the goal, Merin uses a *relevance* function r, such that $r_H(A)$ is positive iff. A argues in favor of H. The precise definition of r can vary, as long as it satisfies certain rules (see van Rooij (2004) for various examples of building such a relevance function).

The reconstruction of the goal H (termed “*abduction*”) is often presented as a problematic feature of such approaches. In Winterstein (2010) it is shown that this reconstruction is not purely contextual as proposed by Merin, but that it is constrained by explicit clues such as the informational structure of the considered utterance.

Another important feature of this approach is that any type of speech act carries with it some argumentative goal. Questions, in particular, are seen as oriented towards a specific goal, and a question is relevant to a goal H, iff. one of its answers is relevant to H.

Finally, this approach has been shown to be effective to account for the combinations of discourse markers such as *but* and *too*. As such, it appears sensible to try to go for an articulation with elements that go beyond discourse structure marking, but that stay in a probabilistic descriptive model.

---

[1] S denotes the epistemic state before an assertion, and $S \oplus A$ denotes this state to which the speaker has added the content A.

4 Bridging the Two Accounts

In this last section, we articulate the analyses presented in Sect. 2 and 3 to explain the data in (3). van Rooy (2003) already draw some parallels between an entropic approach to NPIs and Merin's notion of relevance. We here focus on the specific opposition marked by *but* and on the argumentative effect of emphatic *any* on its host sentence.

4.1 Questions and Probabilities

Before going on an example involving both NP/FCI and discourse markers, we will show that the probabilistic effects of *but* interact with the expectations of the speakers regarding the answers to his question. This is illustrated in (17) where, out of the blue, the use of *but* to indicate that the potatoes are rotten (17-b) is better than a plain juxtaposition (17-a).

(17) *A:* Do you have potatoes?
  a. ?*B:* Yes, rotten ones.
  b. *B:* Yes, but rotten ones.

This is explained in the following manner:

- By asking its question, the speaker has a goal H in mind, to which the question must be relevant. Since the question is polar, each of the two answers must have an effect on H (either increase or decrease its probability). For simplicity, let's assume that H is a goal positively affected by the *yes* answer.
- Out of the blue, the kind of things one wants to do with potatoes involves them not being rotten (the most plausible one is eating the said potatoes). So the probability of H is raised iff. one has non-rotten potatoes. In Jayez's terms, the potatoes are not all equivalent to satisfy H.
- Thus, the *yes* answer indeed raises the probability of H: it excludes the worlds without potatoes, and by itself augments the chances to do H. But having rotten potatoes is a counter argument to this same H, which means that the conditions of use of *but* are met in this context. This explains the preference to use it, since, as noted for example by Asher & Lascarides (2003), the so-called *denial of expectation* uses of *but* make the use of an adversative connective mandatory.

4.2 Discourse Markers and NP/FCI

The case of (18) shows that the use of *but* is no longer preferred if the speaker uses a NPI in his question instead of plain quantification.

(18) *A:* Do you have any potatoes?
  a. *B:* Yes, rotten ones.
  b. *B:* Yes, but rotten ones.

The improved character of (18-a) is due to the fact that the use of *any* is an indication that all potatoes are equal regarding the goal H that the speaker has in mind. Thus, them being rotten is no longer a counter argument for H, as it was in (17), and the use of *but* is not triggered. It is still possible to use *but* in this case, and by doing so, B indicates more clearly that he assumes that rotten potatoes might not be suitable for whatever purpose A has in mind.

To finish, let's turn to the case of the contrast in (3), repeated here for convenience.

(19) a. #I'm glad you got us any tickets at all, but they're not front row.
b. I'm glad you got us tickets, but they're not front row. =(3)

The use of *any* in the first segment of (19-a) indicates that all tickets have the same probability to satisfy the speaker. The bias reduction marked by *any* also indicates that, before the assertion the probability of the speaker being satisfied was lower than it is after the assertion.

To formalize this let's consider the propositions T_i ($i \in [1, n]$) that correspond to the event of getting the ticket (or set of tickets) t_i. Then the entropic effects marked by *any* entail the following:

(20) a. $\forall i, j : P_{S \oplus T_i}(glad'(t_i)) = P_{S \oplus T_j}(glad'(t_j))$
b. $\forall i : P_{S \oplus T_i}(glad'(t_i)) > P_S(glad'(t_i))$

(20-a) expresses that the probability of being satisfied by a ticket t_i is the same as for any other ticket and (20-b) marks that for all tickets the probability of it being satisfactory is higher than it was before the assertion.

Taken together, these two assertions entail the following:

(21) $\forall i : \; r_{glad'(t_i)}(T_i) > 0$

In other terms: getting any ticket t_i is an *argument* in favour of the conclusion "*The speaker is glad of getting* t_i". This comes as the direct effect of the use of *any* and its entropic effects on the set of tickets.

Following that, the use of *but* requires that there is a goal that is debated by its conjuncts. Given the context, the second conjunct of *but* is preferably interpreted as giving an argument against the fact that the speaker is glad of his tickets. If we denote the non-front row tickets as $t_{\neg fr}$, then the second conjunct would have the following argumentative effect:

(22) $r_{glad'(t_{\neg fr})}(T_{\neg fr}) < 0$

But the constraint (22) is not compatible with what is semantically marked by *any*, cf. (21), which means that the second conjunct cannot be taken as an argument against $glad'(t_{\neg fr})$. The bare plural in (19-b) does not impose equity between the tickets, therefore it allows the second conjunct to bear the argumentative contribution in (22) and the discourse is felicitous.

A direct consequence of this account is that we predict that (19-a) is acceptable if the context allows the construction of an argumentative goal which does not relate to the speaker being glad of his/her tickets. We believe that (23) is a case in point, which further confirms the validity of our approach.

(23) a. [*Context:*] Front-row tickets are far more expensive than the other tickets. The speaker has given the addressee some money to buy tickets, and enough money to buy front row ones.
b. I'm glad you got us any tickets at all, but they're not front row, so you should give me some money back.

In (23-b) the disputed goal does not relate on being glad of some tickets, but rather of being glad of the addressee's achievement. On this particular subject, *any* does not have the entropic effects it bears on the tickets. This means that nothing prevents this proposition of playing the role of argumentative goal. The question of why this other proposition cannot be abduced without a specific context to trigger it is not yet solved, but we acknowledge that a complete account of the phenomena treated here should include hints on an answer.

5 Conclusion

In this work, we proposed, following van Rooy, to reinterpret strengthening as a special case of the information-theoretic principle according to which the less likely an assertion is, the more informative it is. This general principle is directly derived from the widening effect of emphatic NP/FCIs that we propose to interpret in informational-theoretic terms as mark of equiprobability on the alternative values over which an emphatic NP/FCI ranges.

Support for this hypothesis is given by the way probabilistic accounts of discourse markers can interact with our account of emphatic NP/FCI to formulate explicit predictions on some discursive sequences.

More precisely, the maximization of entropy induced by *any* can be seen as being made relatively to some particular goal that actually matches the argumentative goal defended by the speaker. As a consequence, this affects the use of discourse connectives such as *but*, that make an explicit reference to this goal in their semantics.

References

Anscombre, J.-C., Ducrot, O.: L'argumentation dans la langue. Pierre Mardaga, Liège (1983)

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

Blakemore, D.: Relevance and Linguistic Meaning. The semantics and pragmatics of discourse markers. Cambridge University Press (2002)

Jayez, J.: Entropy and Free-choiceness. In: Proceedings of the Workshop on Alternative-Based Semantics. Laboratoire de Linguistique de Nantes–Université de Nantes (2010)

Kadmon, N., Landman, F.: Any. Linguistics and Philosophy 16(4), 353–422 (1993)

Krifka, M.: The Semantics and Pragmatics of Weak and Strong Polarity Items. Linguistic Analysis 25, 209–257 (1995)

Lakoff, R.: If's, And's and Buts about conjunction. In: Fillmore, C.J., Terence Langendoen, D. (eds.) Studies in Linguistic Semantics, pp. 114–149. de Gruyter, New York (1971)

Merin, A.: Information, Relevance and Social Decision-Making. In: Moss, L.S., Ginzburg, J., de Rijke, M. (eds.) Logic, Language, and Computation, vol. 2, pp. 179–221. CSLI Publications, Stanford (1999)

van Rooij, R.: Cooperative versus argumentative communication. Philosophia Scientia 2, 195–209 (2004)

van Rooy, R.: Negative Polarity Items in Questions: Strength as Relevance. Journal of Semantics 20, 239–273 (2003)

Sæbø, K.J.: Presupposition and Contrast: German aber as a Topic Particle. In: Weisgerber (ed.) Proceedings of Sinn und Bedeutung, Constance, vol. 7, pp. 257–271 (2003)

Umbach, C.: Contrast and Information Structure: A focus-based analysis of but. Linguistics 43(1), 207–232 (2005)

Vendler, Z.: Each and Every, Any and All. In: Vendler, Z. (ed.) Linguistics in Philosophy. Cornell University Press, Ithaca (1967)

Winterstein, G.: La dimension probabiliste des marqueurs de discours. Nouvelles perspectives sur l'argumentation dans la langue. Ph.D. thesis, Université Paris Diderot–Paris 7 (2010)

Zwarts, F.: A Hierarchy of Negative Expressions. In: Wansing, H. (ed.) Negation: A Notion in Focus, pp. 169–194. Mouton de Gruyter, Berlin (1996)

The Japanese Particle *yo* in Declaratives: Relevance, Priority, and Blaming

David Y. Oshima

GSID, Nagoya University, Japan 464-8601
`oshima@gsid.nagoya-u.ac.jp`
`http://gsid.nagoya-u.ac.jp/oshima`

Abstract. This paper 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* with a rising contour instructs to update 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* with a non-rising contour has a function to indicate that the hearer should have recognized the propositional content beforehand. The two uses of *yo* share the property of being concerned with the hearer's duties.

1 Introduction

This paper develops an analysis of two central functions of the Japanese discourse particle *yo* in declarative clauses. Section 2 presents basic facts about *yo*. Section 3 briefly reviews three influential analyses of *yo* within formal theories of discourse: Takubo and Kinsui (1997), McCready (2009), and Davis (2011), and discusses their limitations. Sections 4 and 5 present a novel analysis of *yo* occurring in declaratives, which is similar to Davis' account in certain respects but improves on it. The main claims are: (i) *yo* accompanied by a rising intonation (Davis' "guide to action" use) 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) *yo* accompanied by a non-rising intonation (Davis' "correction" use) has a function to indicate that the hearer should have recognized the propositional content beforehand.

2 Basic Facts about *yo*

The discourse functions of discourse particles (also called 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. It is used in a wide variety of speech styles and registers, e.g., both in male and female speech,

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 40–53, 2012.

and both in formal and informal speech. Also, it may occur in a wide range of clause types including declaratives, interrogatives, imperatives, and exhortatives.

It has been recognized that *yo* exhibits rather different functions depending on the intonation accompanying it (Koyama 1997; Davis 2011).[1] *Yo* may occur either with (i) the rising contour commonly referred to as "question rise" and assigned the label "LH%" in Venditti's (2005) notational system, or (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.[2] Throughout the paper, I will use ↗ to indicate the rising contour (question rise) and ↘ the non-rising contour.

This work focuses on what Davis (2011) calls the "guide to action" use and "correction" use of *yo*, but it must be noted that *yo* is functionally diverse – even putting aside cases where it occurs in non-declaratives – and has other uses (discourse functions) that cannot be handled with the analysis to be presented.[3]

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 (1), 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 rising contour was added by the present author, assuming that it is the intonation intended by Takubo and Kinsui.[4]

(1) Ame-ga futteiru-yo↗
rain-Nom fall.Ipfv.Prs-*yo*
'It is raining.' (adapted from Takubo and Kinsui 1997:756)

---

[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* may also be accompanied by the "rise-fall" contour (HL%). The functions of *yo* with a rise-fall contour, which are similar to but not identical with that of *yo* with a non-rising contour, will not be discussed in the current work.

[3] To mention one, *yo* with a non-rising contour may indicate that the speaker feels a heightened emotion (e.g., surprise) towards the propositional content (see Tanaka and Kubozono 1999:122).

(i) (The speaker looks outside the window and notices that it is snowing.)
Are, yuki-ga futteru-yo↘
oh snow-Nom fall.Ipfv.Prs-*yo*
'Oh, it's snowing.'

[4] The abbreviations used in glosses are: Acc = accusative, Aux = auxiliary, Cond = conditional, Cop = copula, Dat = dative, DP = discourse particle, Gen = genitive, Ipfv = imperfective, Neg = Negation, Nom = nominative, Pro = pronoun, Prs = present, Pst = past, Q = question marker, Quot = quotative marker, Top = topic, Vol = volitional.

"Direction to make an inference", however, is not a sufficiently specific characterization of the function of *yo* in question. Compare (2) and (3), 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.

(2) A: Kami koppu-mo katte-okoo-ka-na.
paper cup-also buy-do.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.)

(3) 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.)

(2B) and (3A) 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 a rising contour, in the latter.

To give another example, (4) is more natural with *yo* accompanied by a 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).

(4) (**Situation**: A is driving and B is on the passenger seat. They are 100km 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 a rising contour specifically has to do with (inference regarding) what the *hearer* should do or be (see Davis 2011:97 for a similar remark).

3.2 McCready (2009)

McCready (2009) suggests that *yo* is essentially a marker of *importance* for the hearer. 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, or simply relevance — 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 conveyed by *yo* are formulated as below:

(5) $[\![yo(\phi)]\!]=$

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:

(6) $\sigma||sassert(\phi)||\sigma' =$

$\sigma||\phi||\sigma'$ if $\sigma||\phi|| \neq \emptyset$
$\sigma|| \downarrow \neg\phi; \phi||\sigma'$ else.

McCready's analysis, as it is, does not seem to account for the speaker/hearer asymmetry illustrated above. In (2) and (3), 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 (7), while it tends to be compulsory in a context where the speaker gives a suggestion or warning in an indirect manner, as in (8) and (9) (see Takubo and Kinsui 1997:756; Davis 2011:99–100 for relevant remarks).

(7) (**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↗/$\emptyset$}
7-Cop.Prs.Polite-*yo*
'It's a "7".'

(8) (**Situation**: A and B are at a noodle place. 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 oishiidesu-{yo↗/??$\emptyset$}
here-Gen soba-Top good.Prs.Polite-*yo*/$\emptyset$
'The soba here is good.'

B': Koko-no soba-wa amari oishikunaidesu-{yo↗/??$\emptyset$}
here-Gen soba-Top much good.Neg.Prs.Polite-*yo*/$\emptyset$
'The soba here is not particularly good.'

(9) (**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 (7B) is less informative than (8B,B')/(9A) in their respective context.

One may suspect that McCready's analysis is suitable for *yo*↘, though not for *yo*↗. I will show in Section 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 a rising and non-rising intonation. He characterizes the function of *yo* with a rising contour, illustrated in (10) (see also (2), (4), (8) and (9)), as "guide to action", and that of *yo* with a non-rising contour, illustrated in (11), as "(call for) correction".

(10) 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 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)

(11) 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. He hypothesizes, in line with Gunlogson (2003), that declaratives usually have the speaker's public beliefs (those beliefs that both the speaker and the hearer acknowledge that the speaker has) rather than the common ground (the intersection of the speaker's and the hearer's public beliefs, in the case of a two-agent conversation), 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 remarks that due to this contrast only a declarative with *yo* (either with a 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 (14) and (15)), 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 the "intonational morphemes" combined with *yo*.

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 (8B), 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 (12), where the propositional content may or may not affect what the optimal action for the hearer is, and (13), where the contextual decision problem remains unsolved in the post-update context.

(12) (**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.'

(13) (**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 (12), 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 is, however, not the intention of A here; what he intends 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 (13), 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 (12) and (13), 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 Section 4 I will present an analysis along this idea.

The "Correction" Use. Regarding *yo* accompanied by a 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. In the case of (11), 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:

(14) (**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↘/$\emptyset$}
day.off-Cop.Prs.Polite-*yo*/$\emptyset$
'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↘/$\emptyset$}
today-Top cold-by day.off-Cop.Prs.Polite-*yo*/$\emptyset$
'He is taking a day off because he has a cold.'

(15) (**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, denchi-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*/∅
'Oh, that's my girlfriend's. She left it here yesterday.'

The use of *yo*↘ is fine in (14a) and (15a,b), but is felt to be odd (unfairly accusing, unreasonably hostile) in (14b) and (15c,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 (14b) and (15c,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 (14a) and (15a,b) are intuitively felt to be more embarrassing for the hearer than those of (14b) and (15c,d). Thus, one would expect that a higher level of politeness is called for in (14a) and (15a,b) than in (14b) and (15c,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 (14a) (the proposition that Morino is taking a day off today as he does on other Sundays) is more informative than that of (14b) (the proposition that Morino is taking a day off because he has cold).

4 *Yo* with a 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.[5] To 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: The modal statement "John should be in New York now" 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). 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 or her 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 like

---

[5] A similar characterization of *yo*↗ is presented by Inoue (1997:64), who suggests that [ϕ *yo*↗] indicates that ϕ holds true in the circumstances surrounding the speaker and hearer, and further poses to the hearer the question: "What are you going to do in these circumstances?"; see also Izuhara (2003:5).

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):

(16) *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 $CG' = CG \cup \{[\![\phi_{decl}]\!]\}$.

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 presupposes that the common ground and the modal base are ones appropriate for hearer-relativized priority modality, and (ii) it adds the propositional content to the modal base, as well as to the common ground.

(17) *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 (1992:72–73) remark that *yo* serves to draw the hearer's attention to the propositional content, and Miyazaki et al.'s (2002:266) 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 (13), 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), although they are concerned with a specific kind of priority modality (i.e., hearer-relativized priority modality).[6]

Table 1. Means to update conversational grounds

	modal base	ordering source
priority modality	**declaratives with *yo*↗**	imperatives
epistemic modality	regular declaratives	evidentials

5 *Yo* with a Non-rising Intonation: Blame on Ignorance

It was observed above, with the data in (14) and (15), that an utterance with *yo*↘ is infelicitous in a context where the hearer cannot be reasonably expected to know the propositional content beforehand,[7] and also that corrective statements need not to be accompanied by *yo* (with a 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.

(18) *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 should 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; see (6))

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.

---

[6] A case can be made that imperatives too are concerned with hearer-relativized priority modality, rather than priority modality in general.

[7] This property of *yo*↘ is captured in Hasunuma's (1996) proposal 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 will depart from hers, however, in claiming that 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.

The 2005 version of McCready's analysis fares better with the data in (14) and (15) than the 2009 version. The utterances (14a) and (15a,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 (14b) or (15c,d). It is counterintuitive, however, to suppose that the utterer of [ϕ *yo*] *presupposes* (i.e., takes it for granted that both the speaker and the hearer believe) that (the speaker believes that) the hearer should come to believe ϕ at the time of utterance. In the context of (11), for example, obviously the speaker does not expect the hearer to believe that (the speaker believes that) he 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. (19B), 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 (cf. (20)).

(19) A: Eiga-wa shichi-ji-kara-da-yo↘
movie-Top 7-o'clock-from-Cop-*yo*
'The movie starts at 7.'
B: Iya, sonna koto-wa nai.
no that matter-Top absent.Prs
'No, that's not so.'

(20) A: Eiga-wa shichi-ji-kara-da-shi, kimi-wa sore-o
movie-Top 7-o'clock-from-Cop.Prs-and you-Top that-Acc
wakatteiru-bekidatta.
know.Ipfv.Prs-should.Pst
'The movie starts at 7, and you should have known it.'
B: Iya, sonna koto-wa nai. Kimi-ga
no that matter-Top absent.Prs you-Nom
roku-ji-da-to itta-sei-de machigaeta-noda.
6-o'clock-Cop.Prs-Quot say.Pst-reason-by err.Pst-Aux.Prs
'No, that's not so. I got it wrong because you told me it was 6.'

I propose that the semantic contribution of *yo*↘ belongs to the level of conventional implicature/expressive meaning (CIE meaning; 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. To convey such a message can be sensible only when the hearer had a chance to know the propositional content. In the cases of (14a) and (15c,d), the hearer did not have such a chance, and thus it is odd to use *yo*↘.

It is worth noting that the proposed discourse functions of *yo*↗ and *yo*↘ are both concerned with the hearer's duties. This commonality can be taken as a conceptual link between the two distinct uses of *yo*.

6 Conclusion

This paper presented an analysis of two central functions *yo* occurring in declarative clauses. *Yo* with a rising contour instructs to update the modal base of priority modality relativized to the hearer with the propositional content, thereby indicating that it is relevant to what the hearer should and may do. *Yo* with a non-rising contour indicates that the hearer should have recognized the propositional content beforehand.

As mentioned earlier, *yo* in declaratives has functions other than the two discussed in the current work. Also, *yo* occurs in clause types other than declaratives too, carrying out yet other functions. I leave it to future research to examine the conceptual links and diachronic relations between the uses of *yo* discussed in the current work and others.

References

Brown, P., Levinson, S.C.: Politeness: Some universals in language usage, Reissue edn. Cambridge University Press, Cambridge (1987)

Davis, C.: Constraining interpretation: Sentence final particles in Japanese. PhD thesis, University of Massachusetts (2011)

Gunlogson, C.: True to form: Rising and falling declaratives as questions in English. Routledge, New York (2003)

Hasunuma, A.: Shuujoshi *yo* no danwa kinoo [The discourse functions of the sentence-final particle *yo*]. In: Ueda, I., Sunakawa, Y., Takami, K., Noda, H., Hasunuma, A. (eds.) Gengo Tankyuu no Ryooiki [Domains of Linguistic Inquiry], pp. 383–395. Daigaku Shorin, Tokyo (1996)

Heim, I.: On the projection problem for presuppositions. In: Barlow, M., Flickinger, D., Wescoat, M.T. (eds.) Proceedings of WCCFL 2, pp. 114–125. Stanford University, Stanford (1983)

Inoue, M.: 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]. Gengo 26(2), 62–71 (1997)

Izuhara, E.: Shuujoshi "yo" "yone" "ne" saikou [A study of ending particles "yo" "yone" "ne"]. The Journal of Aichi Gakuin University: Humanities & Sciences 51(2), 1–15 (2003)

Kinsui, S.: Shuujoshi *yo* to *ne* [Sentence-final particles *yo* and *ne*]. Gengo 22(4), 118–121 (1993)

Kori, S.: Nihongo no intoneeshon: Kata to kinoo [Intonation in Japanese: Patterns and functions]. In: Kunihiro, T. (ed.) Akusento, Intoneeshon, Rizumu to Poozu [Accent, Intonation, Rythm and Pause], pp. 169–202. Sanseido, Tokyo (1997)

Koyama, T.: 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. Kurosio Publishers, Tokyo (1997)

Kratzer, A.: Modality. In: von Stechow, A., Wunderlich, D. (eds.) Semantik: Ein Internationales Handbuch der Zeitgenössischen Forschung, pp. 651–656. Walter de Gruyter, Berlin (1991)

McCready, E.: The dynamics of particles. PhD thesis, University of Texas (2005)

McCready, E.: Particles: Dynamics vs. utility. In: Takubo, Y., Kinuhata, T., Grzelak, S., Nagai, K. (eds.) Japanese/Korean Linguistics, vol. 16, pp. 466–481. CSLI Publications, Stanford (2009)

McCready, E.: Varieties of conventional implicature. Semantics and Pragmatics 3(8), 1–57 (2010)

Miyazaki, K., Adachi, T., Noda, H., Takanashi, S.: Modaritii [Modality]. Kurosio Publishers, Tokyo (2002)

Nihongo Kijutsu Bunpoo Kenkyuukai (ed.): Dai 8 bu: Modaritii [Part 8: Modality]. Gendai nihongo bunpoo [Contemporary Japanese grammar], vol. 4. Kurosio Publishers, Tokyo (2003)

Portner, P.: Imperatives and modals. Natural Language Semantics 15(4), 351–383 (2007)

Potts, C.: The logic of conventional impricatures. Oxford University Press, Oxford (2005)

Potts, C.: The expressive dimension. Theoretical Linguistics 33(2), 165–197 (2007)

Takubo, Y., Kinsui, S.: Discourse management in terms of mental spaces. Journal of Pragmatics 28(6), 741–758 (1997)

Tanaka, S., Kubozono, H.: Nihongo no hatsuon kyooshitsu: Riron to renshuu [Introduction to Japanese pronunciation: Theory and practice]. Kurosio Publishers, Tokyo (1999)

Uyeno, T.: Shuujoshi to sono shuuhen [On Japanese sentence particles]. Nihongo Kyooiku 17, 62–77 (1972)

Venditti, J.J.: The J_ToBI model of Japanese intonation. In: Sun-Ah, J. (ed.) Prosodic Typology: The Phonology of Intonation and Phrasing, pp. 172–200. Oxford University Press, Oxford (2005)

Conjoined Nominal Expressions in Japanese

Interpretation through Monad

J.-R. Hayashishita[1] and Daisuke Bekki[2]

[1] University of Otago
[2] Ochanomizu University

1 Introduction

This paper studies nominal expressions in Japanese that are formed by conjoining two or more nominal expressions with conjunctions, and presents a theory of nominal expressions that captures their behaviors. In what follows, such nominal expressions are referred to as conjoined nominal expressions.

In terms of their behaviors, the conjoined nominal expressions in Japanese form three categories. The members of the first category represented by *A to B* always 'refer to' a plural object, while those of the second represented by *A ya B* may 'refer to' a plural or singular object depending on the environment where they are used. The members of the third category −*A ka B*, for example− necessarily 'refer to' a singular object.

For example, for (1a) and (1b) to be true, both Mark and Luke must come; however, (1c) indicates that only one of Mark and Luke came.

(1) a. [Mark to Luke] ga kita. '(Lit.) [Mark to Luke] came.'
b. [Mark ya Luke] ga kita. '(Lit.) [Mark ya Luke] came.'
c. [Mark ka Luke] ga kita. '(Lit.) [Mark ka Luke] came.'

Thus, in these examples, *A to B* and *A ya B* 'refer to' a plural object, and *A ka B* a singular object.

The sentences in (2) depict a different picture. (2a) states that Mary offers tea if both Mark and Luke come, but with (2b) and (2c), Mary offers tea as long as one person, Mark or Luke, comes.

(2) a. Mary wa [Mark to Luke] ga kita ra, otya o dasu.
'(Lit.) Mary offers tea if [Mark to Luke] come.'
b. Mary wa [Mark ya Luke] ga kita ra, otya o dasu.
'(Lit.) Mary offers tea if [Mark ya Luke] come.'
c. Mary wa [Mark ka Luke] ga kita ra, otya o dasu.
'(Lit.) Mary offers tea if [Mark ka Luke] come.'

It is thus indicated that while *A to B* cannot 'refer to' a singular object, *A ya B* may pattern with *A ka B* being understood to 'refer to' a singular object. (The 'singular' nature of *A ya B* is originally documented in [11]).

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 54–67, 2012.

Although we do not supply actual examples here, all the conjoined nominal expressions in Japanese fall into one of these three categories. We provide a partial list in (3).

(3) a. Category 1 – Those always 'referring to' a plural object:
A to B, A oyobi B, A narabi B, etc.

b. Category 2 – Those 'referring to' a plural object or a singular object:
A ya B, A toka B, A, B nado, etc.

c. Category 3 – Those always 'referring to' a singular object:
A ka B, A matawa B, A mosikuwa B, etc.

In what follows, we mostly discuss *A to B*, *A ya B*, and *A ka B* as their representatives.

The data we have just observed must be accounted for by a theory of nominal expressions rather than a theory of sentential conjunctions (e.g., the conjunction reduction theory in [6]), as the conjunctions under discussion can only conjoin nominal expressions. For example, *to*, *ya*, and *ka* cannot conjoin adjectival phrases (see (4a)) or verb phrases (see (4b)).

(4) a. * Luke wa kasikoi {to / ya/ ka} yasasii.

'Luke smart and/or kind.'

b. * Luke wa Tokyoo de nihongo o benkyoosita {to / ya / ka} Syanhai de tyuugokugo o benkyoosita.

'Luke studied Japanese in Tokyo, and/or studied Chinese in Shanghai.'

Observing the above data, one might pursue an analysis that treats the conjoined nominal expressions under discussion as generalized conjunctions and disjunctions in the sense of [5]. We, however, demonstrate below that such an analysis is insufficient; it cannot account for the behavior of the Category 2 items. We propose a novel theory of nominal expressions, which makes use of *interpretation through monad* in [2]. The proposed theory in effect treats all the conjoined nominal expressions as individuals. In particular, the Category 1 items and the Category 3 items are in effect treated as sums of singular-individuals and singular-individuals, respectively. Crucially, the theory allows the Category 2 items to be singular-individuals and sums of singular-individuals at the same time.

The remainder of the paper is organized as follows. We first argue in Section 2 that any analyses that treat the conjoined nominal expressions under discussion as generalized conjunctions and disjunctions cannot be maintained. We then propose a theory of nominal expressions making use of interpretation through monad in Section 3. In Section 4, we maintain that together with pragmatic considerations, the proposed theory sufficiently accounts for the data considered in the foregoing sections. We conclude the paper with a brief summary in Section 5.

2 Arguments against Treating the Conjoined Nominal Expressions as Generalized Conjunctions and Disjunctions

Towards the building of a theory of nominal expressions, in this section we further examine the behaviors of the conjoined nominal expressions in Japanese. The objective of this section is to argue that any theories analyzing the conjoined nominal expressions under discussion to be generalized conjunctions and disjunctions fail to capture their behaviors. To this end, we demonstrate that *A ya B* cannot be treated as a generalized conjunction or disjunction.

2.1 Problems of Treating *A ya B* as a Generalized Conjunction

In the pursuit of analyzing the conjoined nominal expressions under discussion as generalized conjunctions and disjunctions, it is reasonable to treat *A to B* as a generalized conjunction and *A ka B* as a generalized disjunction; see (5).

(5) a. $[\![A\ to\ B]\!] = \lambda P.\ (P(a) \wedge P(b))$ (i.e., $\lambda P\ \forall x \in \{a, b\}\,(P(x))$)
b. $[\![A\ ka\ B]\!] = \lambda P.\ (P(a) \vee P(b))$ (i.e., $\lambda P\ \exists x \in \{a, b\}\,(P(x))$)

This allows us to capture the contrasts between (1a) and (1c) and between (2a) and (2c). The question is how to analyze *A ya B*.

Let us first consider analyzing it as a generalized conjunction being on a par with *A to B*. As we will observe shortly, such an analysis fails to account for the intuition that *A ya B* may 'refer to' a singular object. Pursuing this analysis, one may argue that (2b) is analyzed as (6).

(6) $\forall x(x \in \{m, l\} \rightarrow \forall w'(wRw' \wedge (x \text{ comes in } w') \rightarrow \text{Mary offers tea in } w'))$

Once so analyzed, the fact that with (2b), Mary offers tea as long as Mark or Luke comes is expected, for (6) is logically equivalent to (7).

(7) $\forall w'(wRw' \wedge \exists x(x \in \{m, l\}\ \wedge\ (x \text{ comes in } w')) \rightarrow \text{Mary offers tea in } w')$

However, the assumption that *A ya B* can take scope over the entire conditional is not founded; for example, the scope of *oozei no gakusei* 'a large number of students' in (8a) and that of *sannin no gakusei* 'three students' in (8b) are restricted to the clause in which they originate.

(8) a. Mary wa oozei no gakusei ga kita ra, otya o dasu.
'Mary offers tea if a large number of students come.'
b. Mary wa sannin no gakusei ga kita ra, otya o dasu.
'Mary offers tea if three students come.'

(8a), for example, cannot be understood to mean that there are a number of students such that for each of them, if he or she comes, Mary offers tea; it must be taken to mean that Mary offers tea if a large number of students come.

In addition, *A ya B* appears to 'refer to' a singular object even in the situation where the logical equivalence between the universal and existential quantifiers does not hold. For example, like (9c), (9b) (in contrast to (9a)) is understood to mean that the coming of Mark or Luke alone is a possibility.

(9) a. Rainen [Mark to Luke] ga nihon ni kuru kamosirenai.
'(Lit.) Next year, [Mark to Luke] may come to Japan.'
b. Rainen [Mark ya Luke] ga nihon ni kuru kamosirenai.
'(Lit.) Next year, [Mark ya Luke] may come to Japan.'
c. Rainen [Mark ka Luke] ga nihon ni kuru kamosirenai.
'(Lit.) Next year, [Mark ka Luke] may come to Japan.'

However, (10) is not logically equivalent to (11), nor does it entail (11).

(10) $\forall x(x \in \{m, l\} \rightarrow \exists w'(wRw' \wedge (x \text{ comes in } w')))$

(11) $\exists w'(wRw' \wedge \exists x(x \in \{m, l\} \wedge (x \text{ comes in } w')))$

In short, analyzing *A ya B* as a generalized conjunction does not capture the cases where it appears to 'refer to' a singular object, and hence is problematic.

2.2 Problems of Treating *A ya B* as a Generalized Disjunction

Even when *A ya B* appears to 'refer to' a singular object, it cannot be analyzed as a generalized disjunction being on a par with *A ka B*. For example, (12a) is infelicitous while (12b) is not.

(12) a. ??Mary wa [Mark ya Luke] ga kita ra otya o dasu ga, Mark, Luke ryoohoo ga kita ra dasanai.
'(Lit.) Mary offers tea if [Mark ya Luke] come, but if both Mark and Luke come, she does not.'
b. Mary wa [Mark ka Luke] ga kita ra otya o dasu ga, Mark, Luke ryoohoo ga kita ra dasanai.
'(Lit.) Mary offers tea if [Mark ka Luke] come, but if both Mark and Luke come, she does not.'

In other words, in the situation where both Mark and Luke come, the first sentence of (12a) makes Mary offer tea, but that of (12b) does not. We may thus say that descriptively, the disjunction involved in *A ya B* is 'inclusive' while that involved in *A ka B* is 'exclusive'. Since admitting two kinds of disjunctions in a theory does not have theoretical appeal, analyzing *A ya B* to be a generalized disjunction is not reasonable.

2.3 Further Problems

There are some other phenomena of *A ya B* indicating that it cannot be analyzed as a generalized conjunction or disjunction. First, (9b), for example, states that the coming of both Mark and Luke and the coming of Mark or Luke alone are both among the possibilities. We thus end up saying that *A ya B* can be a generalized conjunction and a generalized disjunction at the same time. Needless to say, this is inconceivable. (We can show the same point, using any Category 2 items.)

Second, *A ya B* gives rise to what we call the 'someone-else' effect; for example, for (1b) to be felicitous, besides Mark and Luke, some additional person must come. Treating *A ya B* as a generalized conjunction or disjunction, it is not clear how this 'someone-else' effect can be handled. (The 'someone-else' effect can be observed with any Category 2 items.)

The discussion above in this section is sufficient for us to conclude that *A ya B* cannot be analyzed as a generalized conjunction or disjunction. Nominal expressions can be recursively combined with conjunctions of any categories; thus, the assumption that in terms of theoretical categories, *A ya B* is the same as *A to B* and *A ka B* is reasonable. Together with this assumption, the above conclusion, in turn, leads us to conclude that the conjoined nominal expressions under discussion cannot be analyzed as generalized conjunctions and disjunctions.

3 Proposals

Having concluded that the conjoined nominal expressions under discussion cannot be analyzed as generalized conjunctions and disjunctions, we pursue an analysis that treats them as individuals. Towards accounting for the behaviors of the conjoined nominal expressions – in particular the complex behaviors of the Category 2 items, which may 'refer to' a singular or plural object – we propose a theory of nominal expression, making use of *interpretation through monad* in [2]. The theory we will introduce in effect analyzes *A to B* and *A ka B* as a sum of singular-individuals and a singular-individual, respectively, and allows *A ya B* to be a sum of singular-individuals and a singular-individual at the same time.

Leaving a formal articulation of the proposed theory until the appendix of the paper, we summarize the main points of the theory here. First, the set of individuals and a binary operator '+' form a join-semilattice (cf. [8]). Second, conjoined nominal expressions and verbs are represented as sets of individuals and sets of predicates, respectively at Semantic Representations (= SRs). For example, we assume that *A to B*, *A ya B*, and *A ka B* are defined as (13a), (13b), and (13c), respectively; in the context where the singular-individuals of the domain are Mark, Luke, and John, *Mark to Luke*, *Mark ya Luke*, and *Mark ka Luke* are represented as (14a), (14b), and (14c), respectively.

(13) a. $[\![A\ to\ B]\!] = \{a + b \mid a \in [\![A]\!],\ b \in [\![B]\!]\}$
b. $[\![A\ ya\ B]\!] = \bigcup_{x \in [\![A]\!] \cup [\![B]\!]} \{z \mid x \leq z\}$
c. $[\![A\ ka\ B]\!] = [\![A]\!] \cup [\![B]\!]$

(14) (The singular-individuals of the domain are Mark, Luke, and John)

a. $[\![Mark\ to\ Luke]\!] = \{m + l\}$

b. $[\![Mark\ ya\ Luke]\!] = \{m, l, m + l, m + j, l + j, m + l + j\}$

c. $[\![Mark\ ka\ Luke]\!] = \{m, l\}$

Third, when a conjoined nominal expression is combined with a verb, each member of the set is combined with the verb, yielding a set of propositions, and the resulting propositions are conjoined with disjunctions; see (15).

(15) a. $[\![Mark\ to\ Luke\ ga\ kita]\!] = \mathit{came}(m + l)$

b. $[\![Mark\ ya\ Luke\ ga\ kita]\!] = \mathit{came}(m) \vee \mathit{came}(l) \vee \mathit{came}(m + l) \vee \mathit{came}(m + j) \vee \mathit{came}(l + j) \vee \mathit{came}(m + l + j)$

c. $[\![Mark\ ka\ Luke\ ga\ kita]\!] = \mathit{came}(m) \vee \mathit{came}(l)$

With these proposals, we can account for most of the phenomena we have observed so far. We start with the plural and singular contrast between (i) *A to B* on the one hand and (ii) *A ya B* and *A ka B* on the other, illustrated in (2). The meanings of (2a), (2b), and (2c) are expressed by (16).

(16) $\forall w'(wRw' \wedge (p \text{ in } w') \rightarrow \text{ Mary offers tea in } w')$

What differentiates them is p there. With (2a), for p to be true, $\mathit{come}(m + l)$ must be true (see (15a)); thus, for Mary to offer tea, both Mark and Luke need to come. With (2b) and (2c), on the other hand, p in (16) can be true when $\mathit{come}(m)$ or $\mathit{come}(l)$ is true (see (15b) and (15c)). It is thus expected that Mary offers tea if one person, Mark or Luke, comes.

Let us turn to the dual status of *A ya B*, illustrated in (9b) − (9b) states that the coming of both Mark and Luke and the coming of Mark or Luke alone are among the possibilities. The meaning of (9b) is expressed by (17).

(17) $\exists w'(wRw' \wedge (p \text{ in } w'))$

For p in (17) to be true, there are a number of possibilities (see (15b)), including when $\mathit{come}(m)$ is true and when $\mathit{come}(m + l)$ is true. It is thus expected that (9b) is understood to mean that the coming of both Mark and Luke and the coming Mark or Luke alone are among possibilities.

The contrast between *A ya B* and *A ka B* in (12) is also accounted for, provided that the first clause of the second sentence in (12a) and (12b) is represented as $\mathit{come}(m + l)$, and $\mathit{come}(m + l)$ does not entail $\mathit{come}(m)$ or $\mathit{come}(l)$ With the first sentence of (12a), when $\mathit{come}(m + l)$ is true, Mary must offer tea; thus, it contradicts with the second sentence. In (12b), no contradiction arises, for the first sentence does not force Mary to offer tea in the situation where $\mathit{come}(m+l)$ is true.

We point out the assumption that $\mathit{come}(m + l)$ does not entail $\mathit{come}(m)$ or $\mathit{come}(l)$ is supported by (18a) in contrast with (18b).

(18) a. Mary wa Mark ga piano o motiageta ra gohoobi o dasu ga, [Mark to Luke] ga hutari de motiageta ra dasanai.

'Mary offers a reward to Mark if he lifts a piano, but if Mark and Luke do together, she does not.'

b. ??Mary wa Mark ga piano o motiageta ra gohoobi o dasu ga, Mark ga saisyoni motiagete, Luke ga sono ato motiageta ra dasanai.

'Mary offers a reward to Mark if he lifts a piano, but if Mark first lifts it and Luke does subsequently, she does not.'

The contrast in (18) indicates that $\mathit{lift\_a\_piano}(m + l)$ does not entail $\mathit{lift\_a\_piano}(m)$ (see (18a)) while $\mathit{lift\_a\_piano}(m) \wedge \mathit{lift\_a\_piano}(l)$ does (see (18b)).

4 Pragmatic Considerations

When the singular-individuals of the domain are Mark, Luke, and John, (1a), (1b), and (1c) are represented as (15a), (15b), and (15c) at SR, respectively. The fact regarding (1a) – it indicates that both Mark and Luke came – is straightforward, because for (1a) to be true, $\mathit{came}(m + l)$ needs to be true; see (15a). But the question we have to address is why for (1b) to be true, it must be the case that both Mark and Luke – more accurately Mark, Luke, and someone else – came, despite the fact that $\mathit{came}(m)$ and $\mathit{came}(l)$ are among the conjuncts conjoined with disjunctions in its SR. This question is important, for (1c), on the other hand, is taken to mean that either Mark or Luke came. In addressing this question, we call for pragmatic considerations.

In the proposed theory, the SR of a given sentence may consist of two or more conjuncts which are combined with disjunctions (e.g., (1b) and (1c)). In such a situation, we submit, there are two scenarios, as described in (19).

(19) In the situation where the speaker utters a sentence p whose SR consists of n conjuncts, where n is a positive integer:

a. Scenario 1
The speaker is conveying the proposition corresponding to one of the n conjuncts. Uttering p instead of a sentence whose SR does not involve a disjunction is a pragmatic necessity.

b. Scenario 2
The speaker is unsure or unwilling to say which propositions hold among m propositions corresponding to m conjuncts of the n conjuncts, where m is a positive integer and $m \leqq n$.

We assume that in interpreting p in (19), the hearer first applies Scenario 1, and when s/he judges Scenario 1 to be not appropriate, s/he then considers Scenario 2.

As we explain directly, (1b) turns out to be an instance of Scenario 1, and (1c) an instance of Scenario 2. We start with the reasoning that applies to (1b). It is not conceivable that uttering (1b), the speaker conveys the propositions corresponding to $\mathit{came}(m)$, $\mathit{came}(l)$, and $\mathit{came}(m+l)$, for if s/he wanted, s/he should have said *Mark ga kita*, *Luke ga kita*, and *Mark to Luke ga kita*, respectively. The proposition corresponding to $\mathit{came}(m+j)$ should not be the one that the speaker would like to convey, as the mentioning of Luke is not justified. Similarly, the proposition corresponding to $\mathit{came}(l+j)$ cannot be so considered, as the mentioning of Mark is not justified. However, it can be understood that the speaker would like to convey the proposition corresponding to $\mathit{came}(m+l+j)$. The reason is that uttering (1b), instead of uttering *Mark to Luke to John ga kita*, the speaker can convey the proposition without unveiling the identity of the extra person, namely John. Consequently, (1b) is understood to mean that Mark, Luke, and someone else came.

Turning to the reasoning that applies to (1c). If the speaker wanted to assert the propositions corresponding to $\mathit{came}(m)$ and $\mathit{came}(l)$, s/he should have said that *Mark ga kita* and *Luke ga kita*, respectively. Thus, Scenario 1 is judged to be not appropriate, and Scenario 2 is considered. Consequently, (1c) is taken to indicate that the speaker is unsure or unwilling to say which proposition holds between the one corresponding to $\mathit{came}(m)$ and the one corresponding to $\mathit{came}(l)$.

5 Summary

After demonstrating that any theories analyzing the conjoined nominal expressions in Japanese to be generalized conjunctions and disjunctions fail, we proposed an alternative theory of nominal expressions, making use of interpretation through monad in [2]. The proposed theory in effect analyzes *A to B* and *A ka B* as a sum of singular-individuals and a singular-individual, respectively, and crucially allows *A ya B* to be a sum of singular-individuals and a singular-individual at the same time. As we mentioned earlier, the three types of conjunctions in Japanese we have considered can combine nominal expressions recursively in any orders. It is thus important to see to what extent the proposed theory can handle 'complex cases'. We will investigate this in our future research.

Appendix: A Semantic Theory with Disjunctive Monad

In our theory, the meaning of a sentence is calculated in steps, as illustrated in the following diagram:

Layer	Example
Sentence	"John or Bill met Mary or Susan"
↓ *Categorial derivation*	
Direct-style SR	$\mathit{meet}(j \cup b, m \cup s)$
↓ *Translation by monad*	
Disjunctive-style SR	$\{\mathit{meet}(j,m), \mathit{meet}(j,s), \mathit{meet}(b,m), \mathit{meet}(b,s)\}$
↓ *Infinite join*	
Proposition	$\mathit{meet}(j,m) \vee \mathit{meet}(j,s) \vee \mathit{meet}(b,m) \vee \mathit{meet}(b,s)$

In the subsequent sections, we will explain what each "layer" represents and how the operations between the layers are defined and carried out to yield the final output ("a proposition") from a given sentence.

A.1 Direct-Style SRs

The semantic representations of lexical entries are described with *direct-style semantic representations* (henthforth direct-style SRs), which are basically terms of a typed lambda calculus with finite products and quantifiers (cf. [7], [4]) but extended with the *control operators*: $\oplus$, $\cup$ and $\bigvee$, whose exact syntax is defined by the following BNF grammar.

Definition 1 (Syntax of direct-style SR)

$$\Lambda ::= x \mid c \mid \lambda x.\Lambda \mid \Lambda\Lambda \mid () \mid (\Lambda,\Lambda) \mid \forall x(\Lambda) \mid \exists x(\Lambda) \mid \Lambda \oplus \Lambda \mid \Lambda \cup \Lambda \mid \bigvee \Lambda$$

To describe the representations of two Japanese nominal conjunctions *ya* and *ka*, $\oplus$ and $\cup$ are used.

For the syntactic calculus, we assume a classical (or combinatory) categorial grammar (with the ground types $\{NP, N, S, \bar{S}\}$), extended with the typing rules for the control operators.

Definition 2 (Typing rules of direct-style SR). *For any syntactic categories X, Y and direct-style semantic representations M, N, the following rules hold.*

$$>\frac{X/Y : M \qquad Y : N}{X : MN} \qquad\qquad <\frac{Y : N \qquad X\backslash Y : M}{X : MN}$$

$$\oplus\frac{NP : M \qquad NP : N}{NP : M \oplus N} \qquad \cup\frac{NP : M \qquad NP : N}{NP : M \cup N} \qquad \oplus\frac{X : M}{X : \bigvee M}$$

We also use a join operator $+ : NP \times NP \to NP$ to form *plural objects.* As in Link's plural semantics, the objects of the type *NP* are assumed to form a *join-semilattice*, which is ensured by the following (standard) three axioms.

Axiom 3 (Axioms of join-semilattice). *For any x, y, z of the type NP, the following equiations hold.*

$$\begin{array}{rll} (Asssociativity) & x + (y + z) & = (x + y) + z \\ (Commutativity) & x + y & = y + x \\ (Idempotency) & x + x & = x \end{array}$$

Note also that we use the infix notation for "+", i.e., $x + y \overset{def}{\equiv} +(x, y)$, and we treat $+(x, y)$ as a function application structure. The partial-order $x \leq y$ that represents an inclusion relation between two plural objects x and y is defined as $y = x + y$ in the standard way, where "=" is a two-place predicate of the type $NP \times NP \to S$. (We also use the infix notation: $x = y \overset{def}{\equiv} = (x, y)$)

The lexical entries of the three conjunctions "to","ya" and "ka" in Japanese are defined in terms of the two control operators "$\oplus$" and "$\cup$" and the join operator "+":

Definition 4 (Lexical entries for conjoined NPs in Japanese)

$$\begin{aligned} to &\vdash NP/NP\backslash NP : \lambda y.\lambda x.(y + x) \\ ya &\vdash NP/NP\backslash NP : \lambda y.\lambda x.(y \oplus x) \\ ka &\vdash NP/NP\backslash NP : \lambda y.\lambda x.(y \cup x) \end{aligned}$$

These conjunctions participate in the semantics composition in a straightforward way:

Example 5 (Categorial derivations of nominals)

$$>\dfrac{<\dfrac{\dfrac{\text{John}}{NP : j} \quad \dfrac{\text{to}}{NP/NP\backslash NP : \lambda y.\lambda x.y + x}}{NP/NP : \lambda x.j + x} \quad \dfrac{\text{Bill}}{NP : b}}{NP : j + b}$$

$$>\dfrac{<\dfrac{\dfrac{\text{John}}{NP : j} \quad \dfrac{\text{ya}}{NP/NP\backslash NP : \lambda y.\lambda x.y \oplus x}}{NP/NP : \lambda x.j \oplus x} \quad \dfrac{\text{Bill}}{NP : b}}{NP : j \oplus b}$$

$$>\dfrac{<\dfrac{\dfrac{\text{John}}{NP : j} \quad \dfrac{\text{ka}}{NP/NP\backslash NP : \lambda y.\lambda x.y \cup x}}{NP/NP : \lambda x.j \cup x} \quad \dfrac{\text{Bill}}{NP : b}}{NP : j \cup b}$$

Example 6 (Categorial derivations of sentences)

$$<\dfrac{\dfrac{\text{John ka Bill ga}}{NP : j \cup b} \quad <\dfrac{\dfrac{\text{Mary to Susan ni}}{NP : m + s} \quad \dfrac{\text{atta}}{S\backslash NP\backslash NP : \lambda y.\lambda x.meet(x, y)}}{S\backslash NP : \lambda x.(x, m + s)}}{S : meet(j \cup b, m + s)}$$

$$<\dfrac{\dfrac{\text{John to Bill ga}}{NP : j + b} \quad <\dfrac{\dfrac{\text{Mary ya Susan ni}}{NP : m \oplus s} \quad \dfrac{\text{atta}}{S\backslash NP\backslash NP : \lambda y.\lambda x.meet(x, y)}}{S\backslash NP : \lambda x.(x, m \oplus s)}}{S : meet(j + b, m \oplus s)}$$

$$<\dfrac{\dfrac{\text{John ya Bill ga}}{NP : j \oplus b} \quad <\dfrac{\dfrac{\text{Mary ka Susan ni}}{NP : m \cup s} \quad \dfrac{\text{atta}}{S\backslash NP\backslash NP : \lambda y.\lambda x.meet(x, y)}}{S\backslash NP : \lambda x.(x, m \cup s)}}{S : meet(j \oplus b, m \cup s)}$$

A.2 Disjunctive-Style SRs

Direct-style SRs are translated to *disjunctive-style representations* whose syntax are defined by the following BNF grammar:

Definition 7 (Syntax of disjunctive-style SRs)

$$\mathcal{D} ::= \{\mathit{\Lambda}\} \ | \ \{X \mid \mathit{\Lambda}, \ldots, \mathit{\Lambda}\} \ | \ \mathcal{D} \cup \mathcal{D} \mid \mathcal{D} \cap \mathcal{D}$$

Although the status of three conjunctions are not distinguished at the layer of direct-style SRs, their differences are captured in the course of translation into disjunctive-style SRs. This process is called the *translation by disjunctive monad.*

Technically, this translation is an instance of the translations by monad proposed in [1]. The translations by monad are a general framework which uniformly handles a number of the pragmatic effects in natural language, using various kinds of monad as parameters of the translations (cf. [9][10]). The definition of disjunctive monad given below[1] is actually the same as the *non-deterministic monad* in [1].

Definition 9 (Translation by disjunctive monad)

$$\begin{array}{ll}
[\![x]\!] & = \{x\} \\
[\![c]\!] & = \{c\} \\
[\![\lambda x.M]\!] & = \{\lambda x.m \mid m \in [\![M]\!]\,\} \\
[\![MN]\!] & = \{mn \mid m \in [\![M]\!], n \in [\![N]\!]\,\} \\
[\![\forall x(M)]\!] & = \{\forall x(m) \mid m \in [\![M]\!]\,\} \\
[\![\exists x(M)]\!] & = \{\exists x(m) \mid m \in [\![M]\!]\,\} \\
[\![()]\!] & = \{()\} \\
[\![(M_1, \ldots, M_n)]\!] & = \{(m_1, \ldots, m_n) \mid m_1 \in [\![M_1]\!], \ldots, m_n \in [\![M_n]\!]\,\} \\
[\![M \oplus N]\!] & = \bigcup_{x \in [\![M]\!] \cup [\![N]\!]} \{z \mid x \leq z\} \\
[\![M \cup N]\!] & = [\![M]\!] \cup [\![N]\!] \\
[\![\bigvee M]\!] & = \{\bigvee [\![M]\!]\}
\end{array}$$

Following the Definition 9, the direct-style representations, which we obtain from the syntactic calculus in the previous section, are translated as below.

---

[1] Strictly speaking, the disjunctive-style SRs are the syntactic-alias of *meta-level* terms in [3]. The disjunctive-style SRs and the corresponding meta-level terms are shown below. In the interests of space, we refer the readers to [3] for the details of meta-lambda calculus.

Definition 8 (Set-theoretic notations in disjunctive-style SRs)

$$\begin{array}{rcl}
x \in M & \overset{def}{\equiv} & M \lightning x \\
\{M_1, \ldots, M_n\} & \overset{def}{\equiv} & \zeta X.(X = M_1) \vee \cdots \vee (X = M_n) \\
\{X \mid M_1, \ldots, M_n\,\} & \overset{def}{\equiv} & (\zeta X.M_1 \wedge \cdots \wedge M_n) \\
M \cup N & \overset{def}{\equiv} & \zeta X.(M \lightning X) \vee (N \lightning X) \\
M \cap N & \overset{def}{\equiv} & \zeta X.(M \lightning X) \wedge (N \lightning X)
\end{array}$$

Example 10 (Translation by disjunctive monad)

$$
\begin{aligned}
[\![j]\!] &= \{j\} \\
[\![meet]\!] &= \{meet\} \\
[\![(j,b)]\!] &= \{(m,n) \mid m \in [\![j]\!], n \in [\![b]\!]\} \\
&= \{(m,n) \mid m \in \{j\}, n \in \{b\}\} \\
&= \{(j,b)\} \\
[\![j \cup b]\!] &= [\![j]\!] \cup [\![b]\!] \\
&= \{j\} \cup \{b\} \\
&= \{j,b\} \\
[\![j + b]\!] &= \{mn \mid m \in [\![+]\!], n \in [\![(j,b)]\!]\} \\
&= \{mn \mid m \in [\![+]\!], n \in \{(j,b)\}\} \\
&= \{+(j,b)\} \\
&= \{j+b\} \\
[\![j \oplus b]\!] &= \bigcup_{x \in [\![j]\!] \cup [\![b]\!]} \{z \mid x \leq z\} \\
&= \bigcup_{x \in \{j\} \cup \{b\}} \{z \mid x \leq z\} \\
&= \{z \mid j \leq z\} \cup \{z \mid b \leq z\} \\
&= \{j, b, j+m, b+m, j+b, j+b+m\} \\
[\![(j \cup b, m \cup s)]\!] &= \{(m,n) \mid m \in [\![j \cup b]\!], n \in [\![m \cup s]\!]\} \\
&= \{(m,n) \mid m \in \{j,b\}, n \in \{m,s\}\} \\
&= \{(j,m), (j,s), (b,m), (b,s)\} \\
[\![meet(j \cup b, m \cup s)]\!] &= \{mn \mid m \in [\![meet]\!], n \in [\![(j \cup b, m \cup s)]\!]\} \\
&= \{meet(j,m), meet(j,s), meet(b,m), meet(b,s)\} \\
[\![(j + b, m \cup s)]\!] &= \{(m,n) \mid m \in [\![j+b]\!], n \in [\![m \cup s]\!]\} \\
&= \{(m,n) \mid m \in \{j+b\}, n \in \{m,s\}\} \\
&= \{(j+b,m), (j+b,s)\}
\end{aligned}
$$

A.3 Proposition

The proposition for a given sentence S is $\bigvee [\![P]\!]$, where P is the direct-style SR of S. The *infinite join* operator is defined as below: it takes the disjunctive-style SR of type S and returns a proposition (i.e. a direct-style SR without control operators).

Definition 11 (Infinite join)

$$\bigvee \{X_1, \ldots, X_n\} \overset{def}{\equiv} X_1 \vee \cdots \vee X_n$$

With the application of the infinite join operator, the disjunctive-style SR of the sentence is flatten to form a disjunctive proposition in a sense of the usual first-order predicate calculus.

Example 12 (Infinite join)

$$\begin{aligned}
&\bigvee \{meet(j,m), meet(j,s), meet(b,m), meet(b,s)\}\\
&= meet(j,m) \vee meet(j,s) \vee meet(b,m) \vee meet(b,s)\\
&\bigvee \{meet(j+b,m), meet(j+b,s)\}\\
&= meet(j+b,m) \vee meet(j+b,s)\\
&\bigvee \{meet(j,m+s), meet(b,m+s)\}\\
&= meet(j,m+s) \vee meet(b,m+s)
\end{aligned}$$

Definition 13 (Lexical entries for conditionals in Japanese)

$$tara/reba \vdash S/S\backslash S : \lambda P.\lambda Q.(\bigvee P) \to Q$$

Example 14 (Categorial derivations of conditional sentences)

$$< \dfrac{< \dfrac{> \dfrac{\dfrac{\text{Mark ka Luke ga}}{\begin{matrix} NP \\ : m \cup l \end{matrix}} \quad \dfrac{\text{ki}}{\begin{matrix} S\backslash NP \\ : \lambda x.come(x) \end{matrix}}}{\begin{matrix} S \\ : come(m \cup l) \end{matrix}} \quad \dfrac{\text{tara}}{\begin{matrix} S/S\backslash S \\ : \lambda P.\lambda Q.(\bigvee P) \to Q \end{matrix}}}{\begin{matrix} S/S \\ : \lambda Q.(\bigvee come(m \cup l)) \to Q \end{matrix}} \quad \dfrac{\text{Susan wa ocha o dasu}}{\begin{matrix} S \\ : serve\_tea(s) \end{matrix}}}{S \ : \ (\bigvee come(m \cup l)) \to serve\_tea(s)}$$

Example 15 (Translation by disjunctive monad)

$$\begin{aligned}
&[\![\bigvee come(m \cup l)]\!]\\
&= \{\bigvee [\![come(m \cup l)]\!]\}\\
&= \{\bigvee \{mn \mid m \in [\![come]\!], n \in [\![m \cup l]\!]\ \}\}\\
&= \{\bigvee \{come(m), come(l)\}\}\\
&= \{come(m) \vee come(l)\}
\end{aligned}$$

$$\begin{aligned}
&[\![\bigvee come(m \cup l) \to serve\_tea(s)]\!]\\
&= \{come(m) \vee come(l) \to serve\_tea(s)\}
\end{aligned}$$

$$\begin{aligned}
&[\![\bigvee come(m + l) \to serve\_tea(s)]\!]\\
&= \{come(m + l) \to serve\_tea(s)\}
\end{aligned}$$

$$\begin{aligned}
&[\![\bigvee come(m \oplus l) \to serve\_tea(s)]\!]\\
&= \{come(m) \vee come(l) \vee come(m + j)\\
&\quad \vee come(l + j) \vee come(m + l)\\
&\quad \vee come(m + l + j) \to serve\_tea(s)\}
\end{aligned}$$

References

1. Bekki, D.: Monads and meta-lambda calculus. In: The Fifth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS 2008) in Conjunction with the 22nd Annual Conference of the Japanese Society for Artificial Intelligence 2008, Asahikawa, Japan, pp. 56–78 (2008)
2. Bekki, D.: Monads and Meta-lambda Calculus. In: Hattori, H., Kawamura, T., Idé, T., Yokoo, M., Murakami, Y. (eds.) JSAI 2008. LNCS (LNAI), vol. 5447, pp. 193–208. Springer, Heidelberg (2009)

3. Bekki, D., Masuko, M.: Meta-lambda calculus: Syntax and semantics. In: McCready, E., Yabushita, K., Yoshimoto, K. (eds.) Formal Approaches to Semantics and Pragmatics. Springer (2012)
4. Gallin, D.: Intensional and Higher-Order Modal Logic. With Application to Montague Semantics. North-Holland Publisher/Elsevier Publisher, Amsterdam, New York (1975)
5. Gazdar, G.: A cross-categorial semantics for conjunction. Linguistics and Philosophy 3(1980), 407–409 (1980)
6. Gleitman, L.: Coordinating conjunctions in english. Language 41, 260–293 (1965)
7. Henkin, L.: A theory of propositional types. Fundamenta Mathematicae 52, 323–344 (1963)
8. Link, G.: The logical analysis of plurals and mass terms: A lattice-theoretical approach. In: Bauerle, R., Schwarze, C., von Stechow, A. (eds.) Meaning, Use and Interpretation of Language, pp. 302–323. De Gruyter, Berlin (1983)
9. Ogata, N.: Towards computational non-associative lambek lambda-calculi for formal pragmatics. In: The Fifth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS 2008) in Conjunction with the 22nd Annual Conference of the Japanese Society for Artificial Intelligence 2008, Asahikawa, Japan, pp. 79–102 (2008)
10. Shan, C.C.: Monads for natural language semantics. In: Striegnitz, K. (ed.) The ESSLLI 2001 Student Session (13th European Summer School in Logic, Language and Information), pp. 285–298 (2001)
11. Tanaka, D.: Rensetumeisiku no sukoopu to bun no imirikai moderu 'The scope of conjoined nominal expressions and sentence interpretation'. Doctoral dissertation, Kyushu University (2009)

Dynamic Semantics as Monadic Computation

Christina Unger

CITEC, Bielefeld University

Abstract. This paper proposes a formulation of the basic ideas of dynamic semantics in terms of the state monad. Such a monadic treatment allows to specify meanings as computations that clearly separate operations accessing and updating the context from purely truth conditional meaning composition.

1 Introduction

In the Montegovian tradition, formal semantics of natural languages are formulated in terms of the lambda calculus, starting with a core set of types, lexical meanings and simple composition rules. To account for phenomena such as intensionality, new types are introduced and make it necessary to revise all existing lexical meanings and composition rules in order to incorporate the new meaning aspect. In order to simplify presentations and allow for uniform, compositional and modular analyses of different phenomena, Shan [12] proposed to phrase formal semantic accounts in terms of monads.

The concept of a monad stems from category theory and became a key tool for structuring the denotational semantics of programming languages [9] as well as for modelling computational effects such as non-determinism, continuations, state changes, exceptions and input-output [13]. Some of these concepts have also been applied to the semantics of natural language, for example continuations for a treatment of quantification [2] and exception handling for capturing presupposition projection [5].

Shan [12] considers several monads well-suited for capturing semantic phenomena: the (pointed) powerset monad for interrogatives and focus, the reader monad for intensionality, and the continuation monad for quantification. Furthermore there is a reasonable consensus that dynamic semantics can be phrased in terms of the state monad, representing common wisdom of dynamic semantic theories as stateful computations. Such a treatment was, e.g., provided by Ogata [10] and Bekki [3]. This paper proposes a slightly different way to do this, mainly focusing on a computational view on state updates.

In general, a monadic approach has two benefits. The first one is a clear separation of those meaning aspects that affect the context from static meaning aspects in a way that retains full compositionality. The second one is modularity. Since all monads rely on the same primitives and composition rules, our state monad for dynamic semantics could be composed with monads capturing other phenomena such as intensionality and presuppositions in a modular fashion.

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 68–81, 2012.

2 The State Monad

A monad is a triple $(M, \texttt{unit}, \star)$, where M is a type constructor mapping each type α to the corresponding monadic type $M\alpha$ (objects of type $M\alpha$ can be thought of as computations that yield a value of type α), `unit` is a function of type $\alpha \to M\alpha$ that injects the value into the monad (i.e., it transforms a value into a computation), and $\star$ (pronounced 'bind') is a function of type $M\alpha \to (\alpha \to M\beta) \to M\beta$ that composes two computations, where the second one depends on a value yielded by the first one.

The state monad represents computations that read and modify a state, where a state can be any kind of environment: a counter, a tree, a set of entities, and so on. The type constructor M in the case of the state monad constructs a function type that takes a state as input and returns a pair of a value and a (possibly new or modified) state as output:

$$M\alpha = \mathit{State} \to (\alpha \times \mathit{State})$$

We take *State* to be a type synonym for a set of terms of type e, representing the context that stores anaphoric possibilities. We use variables $s, s', \ldots$ for states.

The functions `unit` and $\star$ of the state monad are defined as follows:

$$\begin{aligned} \texttt{unit}\ x &= \lambda s.\langle x, s\rangle \\ v \star k &= \lambda s.k\ \pi_1(v\ s)\ \pi_2(v\ s) \end{aligned}$$

Where π_1 and π_2 are functions that return the first and second element of a pair, respectively.

Building on these, a monadic version of function application, @ of type $M(\alpha \to \beta) \to M\alpha \to M\beta$, can be defined:

$$k \mathbin{@} v = k \star \lambda f.(v \star \lambda x.\texttt{unit}\ (f\ x))$$

This can be read as the following sequence of computation steps: Compute k and name the result f, compute v and name the result x, then apply f to x and inject the result into the monad again. For the state monad, $k \mathbin{@} v$ reduces to $\lambda s.\langle f\ x, s'\rangle$ (where s' either equals s or results from some operation on s), i.e. the result of extracting the function and its argument from the monad, applying the former to the latter and injecting the result into the monad again.

For practical reasons, we additionally define a function $\triangleright$ of type $M\alpha \to M\beta \to M\beta$ for threading operations that only affect the state without producing a meaningful value, defined as follows: $k \triangleright v = k \star \lambda x.v$, where x must not occur free in v. Introducing entities into the context will be an example for such an operation.

3 Lifting Denotations to the State Monad

We first inject the familiar denotations of nouns, verbs, etc. into the state monad, i.e., every denotation of type α will be lifted to a denotation of type $M\alpha = \mathit{State} \to$

$\alpha \times State$. Then we will specify operations reading and updating the state and add them to the denotations of proper names, pronouns, and quantifiers.

Values are lifted into a monad by means of functions $\texttt{lift}_n$, with n indicating the arity of the value to be lifted. That is, $\texttt{lift}_0$ applies to terms of some type α and lifts them to monadic terms of the general type $M\alpha$ (i.e. in our case $State \rightarrow \alpha \times State$), $\texttt{lift}_1$ applies to one-place functions and is of the general type $(\alpha \rightarrow \beta) \rightarrow (M\alpha \rightarrow M\beta)$, $\texttt{lift}_2$ applies to two-place functions and is of the general type $(\alpha \rightarrow \beta \rightarrow \gamma) \rightarrow (M\alpha \rightarrow M\beta \rightarrow M\gamma)$, and so on. The definition of $\texttt{lift}_n$ is systematic and straighforward:

$$\begin{aligned}
&\texttt{lift}_0\ x = \texttt{unit}\ x\\
&\texttt{lift}_1\ f = \lambda m.(m \star \lambda x.(\texttt{unit}\ (f\ x)))\\
&\texttt{lift}_2\ f = \lambda m_1 \lambda m_2.(m_1 \star \lambda x.(m_2 \star \lambda y.(\texttt{unit}\,(f\ x\ y))))
\end{aligned}$$

That is, $\texttt{lift}_0$ simply corresponds to $\texttt{unit}$, $\texttt{lift}_1$ lifts a one-place function f to another one-place function that wants its argument as a monadic value, and, once supplied with it, computes this argument, binds it to x, applies f to x, and injects the result into the monad again—and similar for $\texttt{lift}_2$. That is, all that is added to the familiar denotations is monadic glue, and with this, the possibility of the function arguments to contain state affecting operations, which would be processed in the order in which the arguments are supplied. We will see examples of this later.

For proper names like Alice we assume denotations of type e,[1] which are lifted to monadic denotations of type Me, i.e., $State \rightarrow e \times State$, by means of $\texttt{unit}$.

$$[\![\textsf{Alice}]\!] = \texttt{lift}_0\ a = \lambda s.\langle a, s\rangle$$

Denotations of type $e \rightarrow t$, for example denotations of common nouns like unicorn or intransitive verbs like whistle, are lifted to monadic denotations of type $Me \rightarrow Mt$, i.e., $(State \rightarrow e \times State) \rightarrow (State \rightarrow t \times State)$:

$$\begin{aligned}
[\![\textsf{unicorn}]\!] &= \texttt{lift}_1\ \textit{unicorn} = \lambda m.(m \star \lambda x.(\texttt{unit}\ (\textit{unicorn}\ x))\\
[\![\textsf{whistle}]\!] &= \texttt{lift}_1\ \textit{whistle} = \lambda m.(m \star \lambda x.(\texttt{unit}\ (\textit{whistle}\ x))
\end{aligned}$$

Analogously, denotations of transitive verbs like admire are lifted to monadic denotations by means of $\texttt{lift}_2$:

$$[\![\textsf{admires}]\!] = \texttt{lift}_2\ \textit{admire} = \lambda m_1 \lambda m_2.(m_1 \star \lambda x.(m_2 \star \lambda y.(\texttt{unit}\ (\textit{admire}\ x\ y)))$$

---

[1] Of course we could also assume generalized quantifier denotations like $\lambda P.(P\ a)$, but we decided to keep things as simple as possible for ease of exposition.

Lifting basic denotations like this, we can compute the meaning of Alice whistles by means of familiar function application:

$$\begin{aligned}
&[\![\textsf{whistles}]\!]\ [\![\textsf{Alice}]\!]\\
&= \lambda m.(m \star \lambda x.(\texttt{unit}\ (\mathit{whistle}\ x)))\ \ \lambda s.\langle a, s\rangle\\
&= \lambda s.\langle a, s\rangle \star \lambda x.\lambda s'.\langle \mathit{whistle}\ x, s'\rangle\\
&= \lambda s''.((\lambda x.\lambda s'.\langle \mathit{whistle}\ x, s'\rangle\ \ \pi_1(\lambda s.\langle a, s\rangle\ s''))\ \ \pi_2(\lambda s.\langle a, s\rangle\ s''))\\
&= \lambda s''.\langle \mathit{whistle}\ a, s''\rangle
\end{aligned}$$

Thus at the core of meaning computation nothing changes yet. We just added a context paramater and used `unit` and $\star$ to introduce this parameter and thread it through the otherwise familiar meaning composition. But what we actually want a proper name denotation to do is to introduce a new term into the context that can be picked up by pronouns later on, for example in a discourse like Alice whistles. Bob admires her. We therefore need a way to modify the state.

4 State Changing Denotations

In order to add terms to a context and later extract them again, we introduce two functions over contexts: (:) of type $e \to \mathit{State} \to \mathit{State}$ that adds some term x to a context s with $x : s$ being the enriched context, and a function `sel` of type $\mathit{State} \to e$ that selects a term from a context.[2]

Now, in order to read and modify the context, we define two state changing operations. The function `new` of type $e \to M\,()$ adds a term to the context and returns the unit value _ (actually it does not matter which value is returned, since we will thread this operation using $\triangleright$, which swallows the value):

$$\texttt{new}\ x = \lambda s.\langle\, \_, x : s\rangle$$

The function `get` of type $(e \to Me) \to Me$ replaces a value of type e by an entity selected from the context:

$$\texttt{get}\ m = \lambda s.\langle \texttt{sel}\ s, s\rangle$$

Now we can specify the denotations of proper names and pronouns as follows:

$$\begin{aligned}
[\![\textsf{Alice}]\!] &= \texttt{new}\ a \triangleright \texttt{lift}_0\ a = \lambda s.\langle a,\ a : s\rangle\\
[\![\textsf{her}]\!] &= \texttt{get} \triangleright \lambda x.(\texttt{lift}_0\ x) = \lambda s.\langle \texttt{sel}\ s, s\rangle
\end{aligned}$$

That is, we inject the familiar values of type e into the monad using $\texttt{lift}_0$ and additionally compose them with a state affecting operation.

---

[2] Which entity is selected should be determined by a pronoun resolution mechanism. Since this is out of the scope of this paper, we assume `sel` to function as an oracle here.

The meaning computation for Alice whistles and for Bob admires her give the following results:

$$\begin{aligned}&[\![\mathsf{whistles}]\!]\ [\![\mathsf{Alice}]\!]\\&= \lambda s.\langle a,\, a : s\rangle \star \lambda x.(\texttt{unit}\ (\mathit{whistle}\ x))\\&= \lambda s.\langle \mathit{whistle}\ a,\, a : s\rangle\end{aligned}$$

$$\begin{aligned}&([\![\mathsf{admires}]\!]\ [\![\mathsf{her}]\!])\ [\![\mathsf{Bob}]\!]\\&= \lambda s.\langle \texttt{sel}\ s, s\rangle \star \lambda x.(\lambda s.\langle b,\, b : s\rangle \star \lambda y.(\texttt{unit}\ (\mathit{admire}\ x\ y)))\\&= \lambda s.\langle \mathit{admire}\ (\texttt{sel}\ s)\ b,\, b : s\rangle\end{aligned}$$

We still consider these sentences as isolated units, but of course we rather want to sequence them, in order to capture the fact that the pronoun her in the second sentence can pick up the reference introduced by the proper name Alice in the first sentence.

5 From Sentences to Discourses

For sequencing sentences we specify a merge operation $\oplus$ of type $Mt \to Mt \to Mt$ that composes two sentences, where the second one should be interpreted w.r.t. the context that the first one returns, and the returned value should be the conjunction of the two sentence meanings. Since sequencing is already encoded in $\star$, $\oplus$ can be defined straightforwardly:

$$m_1 \oplus m_2 = m_1 \star \lambda p.(m_2 \star \lambda q.(\texttt{unit}\ (p \wedge q)))$$

This definition can be read as follows: Compute m_1 and name the result p, compute m_2 and name the result q, then build the conjunction of p and q and inject it into the monad again.

Consider, for example, the sentences Alice whistles and Bob admires her. The denotation of the discourse of the former followed by the latter is the following:

$$\begin{aligned}&\mathsf{Alice\ whistles}\ \oplus\ \mathsf{Bob\ admires\ her}\\&=\lambda s.\langle \mathit{whistle}\ a,\, a : s\rangle\ \oplus\ \lambda s.\langle \mathit{admire}\ (\texttt{sel}\ s)\ b,\, b : s\rangle\\&=\lambda s.\langle (\mathit{whistle}\ a) \wedge (\mathit{admire}\ (\texttt{sel}\ a : s)\ b), b : a : s\rangle\end{aligned}$$

That is, Alice whistles is interpreted w.r.t. the input context s and updates it by adding a. The subsequent sentence Bob admires her is then interpreted w.r.t. this updated context $a : s$ and adds another entity, b, which can be picked up by pronouns still to come.

That is, meaning composition proceeds as usual, but is enriched with a transition from an input context to an output context. In terms of DRT, a simple DRS $[x_1, \ldots, x_n \mid C_1, \ldots, C_m]$ with discourse referents $x_1, \ldots, x_n$ and conditions $C_1, \ldots, C_m$ would correspond to $\lambda s.\langle C_1 \wedge \ldots \wedge C_m,\, x_n : \ldots : x_1 : s\rangle$, i.e., a lambda term like the resulting one from above.

6 Quantification

The most interesting problem still remains: Quantifiers like most and every introduce entities into the context only temporarily: They are accessible only within the scope of the quantifier but not beyond. For example, in 1 the pronoun it cannot pick up the entity introduced by every unicorn.

1. Every unicorn is eating Bob's flowers. He adores it.

Entities introduced by proper names such as Bob, however, are usually accessible throughout whole discourses, for example he in the second sentence of 1 can pick up Bob as referent without a problem.

6.1 Monadic Quantifier Denotations

Quantifier denotations differ from the denotations of proper names in that they introduce an entity that is accessible only within the quantifier's scope. In order to capture this behavior, we assemble quantifier denotations using the following ingredients:

- the function `new`, that introduces a term into the state
- the usual quantifier denotation lifted into the monad, e.g., $\mathtt{lift}_2\ \lambda P \lambda Q.\forall x.P\ x \to Q\ x$
- a function `clear` that removes the introduced term from the state

Let us look at the lifting part first. Since quantifier denotations of type $(e \to t) \to (e \to t) \to t$ are two-place functions, we can lift them to monadic denotations using $\mathtt{lift}_2$. For example, applying $\mathtt{lift}_2$ to the denotation of every (the familiar $\lambda P \lambda Q.\forall x.P\ x \to Q\ x$) gives the following:

$$\lambda m_1 \lambda m_2.(m_1 \star \lambda P.(m_2 \star \lambda Q.(\mathtt{unit}\ \forall x.P\ x \to Q\ x)))$$

Note that this result is of type $M(e \to t) \to M(e \to t) \to t$, while the expected arguments will be of type $Me \to Mt$ (recall the lifted denotation of, e.g., unicorn or whistle from above). So we need a coercion function that turns a monadic function $M(\alpha \to \beta)$ into a function $M\alpha \to M\beta$ over monadic arguments. Specifying such a coercion function is a mere technicality. We call it $\uparrow$ and define it as follows:

$$\uparrow m = \lambda s.\langle \lambda x.\pi_1(m\ (\mathtt{unit}\ x)\ s), \pi_2((\mathtt{unit}\ x)\ s)\rangle$$

This looks much more complicated than it actually is; all that hapens is that we supply m with a monadically lifted argument x, over which we then abstract inside the monad, at the same time making sure that the state which will result from computing m is kept in the final denotation. For example, coercing the denotation of unicorn of type $M(e \to t)$, repeated in 2a, yields the denotation of type $Me \to Mt$ given in 2b.

2. (a) $\lambda m.(m \star \lambda x.(\texttt{unit}\ (\mathit{unicorn}\ x)))$
   (b) $\lambda s.\langle \lambda x.(\mathit{unicorn}\ x), s\rangle$

Now we can define a variant of function application, `app`, of type $(M(\alpha \to \beta) \to \gamma) \to (M\alpha \to M\beta) \to \gamma$, that coerces the argument before handing it to the function:

$$f\ \texttt{app}\ x = (f \quad \uparrow x)$$

Let us now turn to the other two parts of quantifier denotations, the state-affecting operations `new` and `clear`, that add an entity to the context and remove it again, respectively. The function `new` of type $e \to M\,()$ was already defined in Section 4 above. The function `clear` is of the same type, its definition differs only in that x is not added to the context but removed from it.

$$\texttt{new}\ x = \lambda s.\langle \_, x : s\rangle$$
$$\texttt{clear}\ x = \lambda s.\langle \_, s - x\rangle$$

Where we use the minus sign for a function of type $\mathit{State} \to e \to \mathit{State}$, that removes an element from a state.

Finally, we can specify quantifier denotations as follows (where we use the notation $f \triangleleft op$ as equivalent to $op \triangleright f$, in order to emphasize the order in which things happen):

$$\lambda m_1 \lambda m_2.\texttt{new}\ x \triangleright (m_1 \star \lambda P.(m_2 \star \lambda Q.((\texttt{unit}\ \forall x.P\ x \to Q\ x) \triangleleft \texttt{clear}\ x)))$$

This is the familiar quantifier denotation being injected into the monad and composed with two state affecting operations: adding the term that is bound to the context and removing it again. The whole denotation can be read as follows: Take to monadic arguments (the first usually being a computation of the noun denotation and the second being a computation of the verb phrase denotation), add the term x to the current state, compute the first argument and name the result P, compute the second argument and name the result Q, then lift the expression $\forall x.P\ x \to Q\ x$ into the monad and remove the term x again from the state. During the whole process, the state is handed from one step to the next: The input state is first updated with x, then handed to the computation of m_1, which possibly updates it, then it is handed to the computation of m_2 and again possibly updated, then x is removed, and the resulting state is handed to whatever sentence enters the stage next.

Let us look at a simple example with universal quantification, like the one in 3. Its denotation is computed by means of coerced function application, cf. 3a. This leads to 3b and finally to 3c.

3. Every unicorn whistles.
   (a) $(\llbracket \text{every} \rrbracket\ \text{app}\ \llbracket \text{unicorn} \rrbracket)\ \text{app}\ \llbracket \text{whistles} \rrbracket$
   (b) $\texttt{new}\ x \triangleright (\uparrow (\texttt{lift}_1\ \mathit{unicorn}) \star \lambda P.(\uparrow (\texttt{lift}_1\ \mathit{whistle}) \star \lambda Q.(\texttt{unit}\ \forall x.P\ x \to Q\ x) \triangleleft \texttt{clear}\ x))$
   (c) $\lambda s.\langle \forall x.\mathit{unicorn}\ x \to \mathit{whistles}\ x, s\rangle$

From the resulting expression you see that the input state is returned without any changes. This is because the only expression that affects it is every, which introduces a term and later deletes it again, in that sense leaves no traces. So let us consider the more interesting example in 4, where another element is added to the state, and where we have a pronoun picking up the universally bound variable. Function application proceeds again as expected, see 4a. The denotations of all words are listed in Table 1 below.

4. Every unicorn thinks that Alice likes it.
   (a) $(\llbracket \textsf{every} \rrbracket\ \textsf{app}\ \llbracket \textsf{unicorn} \rrbracket)\ \textsf{app}\ (\llbracket \textsf{thinks} \rrbracket\ ((\llbracket \textsf{likes} \rrbracket\ \llbracket \textsf{it} \rrbracket)\ \llbracket \textsf{Alice} \rrbracket))$
   (b) $\lambda m \lambda s.\langle \mathit{think}\ (\mathit{like}\ (\texttt{sel}\ s)\ a)\ \ \pi_1(m\ a:s), \pi_2(m\ s)\rangle$
   (c) $\lambda s.\langle \lambda x.(\mathit{think}\ (\mathit{like}\ (\texttt{sel}\ s)\ a)\ x), a:s\rangle$
   (d) $\lambda s.\langle \forall x.\mathit{unicorn}\ x \to \mathit{think}\ (\mathit{like}\ (\texttt{sel}\ x:s)\ a)\ x, a:s\rangle$

The denotation for thinks that Alice likes it is given in 4b. Coerced, this amounts to 4c. The input context s will be the context that results from the computation of every unicorn, i.e., will contain the term x. Applying the denotation of every unicorn (composed like in Example 3 above) to 4b yields 4d. Note that the reference for the pronoun is selected from the updated state $x : s$, and that the returned state is $a : s$. That is, two terms are introduced into the state, one by the quantifier and one by the proper name, but only the latter is accessible outside the quantifier's scope.

Table 1.

	Denotation
every	$\lambda m_1 \lambda m_2.\texttt{new}\ x \triangleright (m_1 \star \lambda P.(m_2 \star \lambda Q.((\texttt{unit}\ \forall x.P\ x \to Q\ x) \triangleleft \texttt{clear}\ x)))$
unicorn	$\lambda m.(m \star \lambda x.(\texttt{unit}\ (\mathit{unicorn}\ x)))$
thinks that	$\lambda m_1 \lambda m_2.(m_1 \star \lambda p.(m_2 \star \lambda x.(\texttt{unit}\ (\mathit{think}\ p\ x))))$
Alice	$\lambda s.\langle a, a : s\rangle$
likes	$\lambda m_1 \lambda m_2.(m_1 \star \lambda x.(m_2 \star \lambda y.(\texttt{unit}\ (\mathit{like}\ x\ y))))$
it	$\lambda s.\langle \texttt{sel}\ s, s\rangle$

To sum up, we now have means to compose lifted denotations with state affecting operations, such that we can capture the introduction and removal of terms from the state. This, however, does not yet enable us to handle donkey sentences like in 5.

5. (a) If a knight meets a knave, he greets him.
   (b) Every knight who meets a knave greets him.

The reason is that every term introduced by a quantifier is removed from the state once the quantifier's scope is closed, hence the term introduced by a knave will not be in the state anymore when the meaning of the pronoun him is computed. This problem is at the core of dynamic semantics and retrieved a wide range of different treatments, some changing the semantics of quantification, such as Dynamic Predicate Logic [6], and some changing the syntax of quantification, such Pagin & Westerståhl [11].

The direction we want to pursue here builds on two quite standard assumptions: First, indefinites like a are not quantifiers in the same sense that every is, but are *free indefinites*, i.e., introduce unbound variables. And second, quantificational binding is unselective. This kind of approach goes back already to Lewis [8] and Heim [7], but we want to give it a slightly new twist.

6.2 Free Indefinites

As mentioned above, in contrast to quantifiers such as every, most, or no, indefinites like a differ in their behavior regarding the availability of introduced referents. In a simple predication without scope-taking elements, their referents are available arbitrarily far to the right, even across sentences as in 6a (which has a denotation as shown in 6b).

6. (a) Alice saw a unicorn in the garden. It was eating flowers.
   (b) $\exists x. \mathit{unicorn}\ x \wedge \mathit{see}\ x\ a \wedge \mathit{eat\ flowers}\ x$

The availability of referents introduced by indefinites is, however, restricted if it occurs in the scope of a quantifier, see 7. Moreover, inside this scope they are free and seem to adopt the quantificational force of the enclosing quantifier, as can be seen in typical donkey sentences such as 8a and its denotation 8b.

7. Everyone saw a unicorn in the garden. #It was eating flowers.
8. (a) Everyone who saw a unicorn admired it.
   (b) $\forall y \forall x. \mathit{unicorn}\ x \wedge \mathit{see}\ x\ y \rightarrow \mathit{admire}\ x\ y$

Because of these properties, free indefinites are often assumed to not be quantifiers. Rather they are given a meaning x or $\lambda P.(P\ x)$, containing a free variable x that is interpreted existentially once truth conditions are assigned. Injected into the monad, the denotation of the determiner a would be $\texttt{new}\ x \triangleright (\texttt{unit}\ x)$, or as generalized quantifier: $\lambda m. \texttt{new}\ x \triangleright (m \star \lambda P.(\texttt{unit}\ (P\ x)))$.

The meaning of the first sentence in 6a thus yields 9a (ignoring in the garden), the second sentence yields 9b, resulting in the discourse 9c.

9. (a) $\lambda s. \langle \mathit{unicorn}\ x \wedge \mathit{see}\ x\ a,\ x : a : s \rangle$
   (b) $\lambda s. \langle \mathit{eat\ flowers}\ (\texttt{sel}\ s), s \rangle$
   (c) $\lambda s. \langle \mathit{unicorn}\ x \wedge \mathit{see}\ x\ a \wedge \mathit{eat\ flowers}\ (\texttt{sel}\ x : a : s),\ x : a : s \rangle$

That is, the free variable x is introduced into the context once and since it is never removed again, it will be accessible as long as the discourse lasts.

In order to also capture the cases in which a free indefinite adopts the quantificational force of another quantifier, we will now reformulate quantificational binding in an unselective fashion.

6.3 Quantification with Unselective Binding

We specify the core meaning of every as $\mathtt{every}\ (P\ x \to Q\ x)$, where $\mathtt{every}$ is an operation that universally binds all free variables in a formula unselectively, and additionally introduces a $\mathtt{clear}$ operation over all those variables, as we suggest that binding always goes together with making the bound variable unaccessible from outside the scope of the quantifier. That is, $\mathtt{every}\ (P\ x \to Q\ x)$ should correspond to $\lambda s.\langle \forall x.(P\ x \to Q\ x), s - x\rangle$.

In general, we want $\mathtt{every}$ to behave as in the following examples, especially it should not change the type of its argument besides injecting it into the monad:

$$\mathtt{every}\ (\mathit{unicorn}\ x) = (\mathtt{unit}\ \forall x.\mathit{unicorn}\ x) \lhd \mathtt{clear}\ x$$
$$\mathtt{every}\ \lambda y.(\mathit{unicorn}\ x \land \mathit{eat}\ y\ x) = (\mathtt{unit}\ \lambda y.\forall x.\mathit{unicorn}\ x \land \mathit{eat}\ y\ x) \lhd \mathtt{clear}\ x$$

To this end, we define $\mathtt{every}$ as a function of the general type $(\alpha_1 \to \dots \alpha_m \to \beta) \to M(\alpha_1 \to \dots \to \alpha_m \to \beta)$ as follows, where $\mathtt{clear}\ x_1 \dots n_n$ is short for $\mathtt{clear}\ x_1 \lhd \dots \lhd \mathtt{clear}\ x_n$:

$$\mathtt{every}\ P = (\mathtt{unit}\ \lambda y_1 \dots \lambda y_m.\forall x_1 \dots \forall x_n.(P\ y_1 \dots y_m)) \lhd \mathtt{clear}\ x_1 \dots x_n$$
$$\text{where } x_1, \dots, x_n \text{ are the free variables of } P$$

The denotation of the quantificational determiner every now reads as before, except that it uses $\mathtt{every}$ instead of $\forall x$ and does not need $\mathtt{unit}$ and $\mathtt{clear}\ x$, as this is taken care of by $\mathtt{every}$:

$$\lambda m_1 \lambda m_2.\mathtt{new}\ x \rhd (m_1 \star \lambda P.(m_2 \star \lambda Q.(\mathtt{every}\ (P\ x \to Q\ x))))$$

Which is the same as:

$$\lambda m_1 \lambda m_2.\mathtt{new}\ x \rhd (m_1 \star \lambda P.(m_2 \star \lambda Q.(\mathtt{unit}\ \forall x.(P\ x \to Q\ x) \lhd \mathtt{clear}\ x)))$$

So for simple quantifier denotations nothing changes. However, an unselective binding mechanism does make a difference for donkey sentences. Consider Every knight who meets a knave greets him. The denotations of the single words are given in Table 2. The meaning of who is as one would expect, just injected into the monad. For meets we now have to take a lifted denotation, so it can take a generalized quantifier as argument. The non-monadic lifted denotation would be $\lambda \mathcal{P} \lambda \mathcal{Q}.(\mathcal{P}\ \lambda x.(\mathcal{Q}\ \lambda y.(\mathit{meets}\ x\ y)))$ of type $((e \to t) \to t) \to ((e \to t) \to t) \to t$, the monadic version is, accordingly, of type $M((e \to t) \to t) \to M((e \to t) \to t) \to Mt$. The meaning composition for the whole sentence is specified in 10a, the final result is given in 10b.

10. Every knight who meets a knave greets him.
    (a) $([\![\text{every}]\!]\ \mathsf{app}\ (([\![\text{who}]\!]\ \mathsf{app}\ ([\![\text{meets}]\!]\ \mathsf{app}\ ([\![\text{a}]\!]\ \mathsf{app}\ [\![\text{knave}]\!])))\ \mathsf{app}\ [\![\text{knight}]\!]))$ $([\![\text{greets}]\!]\ [\![\text{him}]\!])$
    (b) $\lambda s.\langle \forall y \forall x.\mathit{knight}\ y \land \mathit{knave}\ x \land \mathit{meet}\ x\ y \to \mathit{greet}\ (\mathtt{sel}\ y : x : s)\ y,\ s\rangle$

Table 2.

	Denotation
every	$\lambda m_1 \lambda m_2.\texttt{new}\ x \triangleright (m_1 \star \lambda P.(m_2 \star \lambda Q.(\texttt{every}\ (P\ x \rightarrow Q\ x))))$
a	$\lambda m_1 \lambda m_2.\texttt{new}\ x \triangleright (m_1 \star \lambda P.(m_2 \star \lambda Q.(\texttt{unit}\ (P\ x \wedge Q\ x))))$
knight	$\lambda m.(m \star \lambda x.(\texttt{unit}\ (\mathit{knight}\ x)))$
knave	$\lambda m.(m \star \lambda x.(\texttt{unit}\ (\mathit{knave}\ x)))$
meets	$\lambda m_1 \lambda m_2.(m_1 \star \lambda \mathcal{P}.(m_2 \star \lambda \mathcal{Q}.(\texttt{unit}\ (\mathcal{P}\ \lambda x.(\mathcal{Q}\ \lambda y.(\mathit{meet}\ x\ y))))))$
greets	$\lambda m_1 \lambda m_2.(m_1 \star \lambda x.(m_2 \star \lambda y.(\texttt{unit}\ (\mathit{greet}\ x\ y))))$
he	$\lambda s.\langle \texttt{sel}\ s, s\rangle$
him	$\lambda s.\langle \texttt{sel}\ s, s\rangle$
who	$\lambda m_1 \lambda m_2 \lambda m_3.(m_1 \star \lambda P.(m_2 \star \lambda Q.(m_3 \star \lambda x.(\texttt{unit}\ (P\ x \wedge Q\ x)))))$

Let us look at details of the two most important steps in the derivation, namely the computation of the quantifier structures. First, we compute the meaning of a knave:

$$\begin{aligned}
&[\![\mathsf{a}]\!]\ \mathsf{app}\ [\![\mathsf{knave}]\!] = [\![\mathsf{a}]\!]\ \uparrow [\![\mathsf{knave}]\!] = [\![\mathsf{a}]\!]\ \lambda s.\langle \lambda x.\mathit{knave}\ x, s\rangle \\
&= \lambda m_2.\texttt{new}\ x \triangleright \lambda s.((m_2 \star \lambda Q.(\texttt{unit}\ (\mathit{knave}\ x \wedge Q\ x)))\ s) \\
&= \lambda m_2 \lambda s.\langle (m_2 \star \lambda Q.(\texttt{unit}\ (\mathit{knave}\ x \wedge Q\ x)))\ x : s,\ x : s\rangle
\end{aligned}$$

Here the quantifier introduces the free variable x into the state. In the subsequent steps, nothing happens with respect to the state, so let us jump to applying the denotation of every to the denotation of knight who meets a knave:

$$\begin{aligned}
&[\![\mathsf{every}]\!]\ \mathsf{app}\ [\![\mathsf{knight\ who\ meets\ a\ knave}]\!] \\
&= [\![\mathsf{every}]\!]\ \lambda s.\langle \lambda y.(\mathit{knight}\ y \wedge \mathit{knave}\ x \wedge \mathit{meet}\ x\ y),\ x : s\rangle \\
&= \lambda m_2.(m_2 \star \lambda Q.(\texttt{every}\ (\mathit{knight}\ y \wedge \mathit{knave}\ x \wedge \mathit{meet}\ x\ y \rightarrow Q\ y))) \\
&= \lambda m_2.(m_2 \star \lambda Q.(\texttt{unit}\ \forall y \forall x.(\mathit{knight}\ y \wedge \mathit{knave}\ x \wedge \mathit{meet}\ x\ y \rightarrow Q\ y)) \\
&\qquad\qquad \triangleleft\ \texttt{clear}\ x \triangleleft \texttt{clear}\ y)
\end{aligned}$$

Here the important step is the last one: Since $\texttt{every}$ unselectively binds all free variables, it also binds x, which was introduced by the indefinite in the very beginning; x thus ends up universally bound. Moreover, this binding procedure makes sure that x will be deleted from the state once the scope of the universal quantifier is closed, but until then it is available for pronouns such as he and him, hence the resulting denotation in 10b above.

Note that since only variables can be bound, constants introduced by proper names will never be deleted from the state, unless explicitely done so, thus are always available throughout the whole discourse.

7 Negation

Negation acts as a scope-taking element in the sense that it restricts the availability of referents introduced by indefinites just like quantifiers, as can be seen in 11.

11. Alice did not see a unicorn in the garden. #It was eating the flowers.

The maybe most straightforward way to capture the scope-taking character of negation is to treat it analogously to quantification, by means of *existential closure*. Existential closure was used by Heim to capture the fact that also non-quantificational indefinites receive an existential interpretation. The rationale for using existential closure for sentential negation is that sentential negation takes a proposition as argument, and propositions are closed formulas. So in order to negate an expression of type t, we need to first bind all remaining free variables. If we use `some` for existential quantification, like we used `every` for universal quantification—where `some` is defined exactly like `every`, only using ∃ instead of ∀—then the denotation of sentential negation can then be specified as follows:

$$[\![\textsf{it is not the case that}]\!] = \lambda m.(m \star \lambda p.(\neg > (\texttt{some}\ p))$$

Where $>$ is an operation of type $(\alpha \to \beta) \to M(\alpha \to \beta) \to M\beta$, that only affects the value but not the state (i.e., is a kind of counterpart to $\triangleright$):

$$f > m = \lambda s.\langle f\ \pi_1(m\ s),\, \pi_2(m\ s)\rangle$$

We need this because $\neg$ is a logical constant of type $t \to t$ and cannot directly be applied to the monadic result of `some` p.

As a simple example consider 12. The meaning computation proceeds as shown in 12a, resulting in 12b.

12. It is not the case that a unicorn barked at Alice.
    (a) $[\![\textsf{not}]\!]\ (([\![\textsf{a}]\!]\ \textsf{app}\ [\![\textsf{unicorn}]\!])\ \textsf{app}\ ([\![\textsf{bark at}]\!]\ [\![\textsf{alice}]\!]))$
    (b) $\lambda s.\langle \neg\exists x.\mathit{unicorn}\ x \wedge \mathit{barkat}\ a\ x, a : s\rangle$

Here is how we get there: The meaning of a unicorn barked at Alice is computed like the examples we saw in the previous section, with a unicorn being a free indefinite, yielding $\lambda s.\langle \mathit{unicorn}\ x \wedge \mathit{barkat}\ a\ x,\ a : x : s\rangle$. Applying negation to it gives the following:

$$\begin{aligned}
&\lambda s.((\neg > (\texttt{some}\ (\mathit{unicorn}\ x \wedge \mathit{barkat}\ a\ x)))\ a : x : s)\\
&= \lambda s.((\neg > (\texttt{unit}\ (\exists x.\mathit{unicorn}\ x \wedge \mathit{barkat}\ a\ x)) \triangleleft \texttt{clear}\ x)\ a : x : s)\\
&= \lambda s.((\neg > \lambda s'.\langle \exists x.\mathit{unicorn}\ x \wedge \mathit{barkat}\ a\ x, s' - x\rangle)\ a : x : s)\\
&= \lambda s.\langle \neg\exists x.\mathit{unicorn}\ x \wedge \mathit{barkat}\ a\ x, a : x : s - x\rangle
\end{aligned}$$

In addition to sentential negation, there is verb phrase negation, applying to denotations of type $e \to t$ (or $Me \to Mt$, in our case). An example is given in 13. Just like sentential negation it restricts the scope of free indefinites, for example him cannot be bound by a knave.

13. Every knight who does not meet a knave greets him.
    (a) $([\![\textsf{every}]\!]\ \textsf{app}\ ([\![\textsf{who}]\!]\ \textsf{app}\ ([\![\textsf{not}]\!]\ \textsf{app}\ ([\![\textsf{meet}]\!]\ \textsf{app}\ ([\![\textsf{a}]\!]\ \textsf{app}\ [\![\textsf{knave}]\!])))$
    $\textsf{app}\ [\![\textsf{knight}]\!]))$
    $\textsf{app}\ ([\![\textsf{greets}]\!]\ [\![\textsf{him}]\!])$

(b) $\lambda s.\langle \forall x.(\mathit{knight}\ x \wedge \neg\exists y.(\mathit{knave}\ y \wedge \mathit{meet}\ y\ x)) \rightarrow \mathit{greet}\ (\mathtt{sel}\ x : s)\ x, s\rangle$

In order to allow the denotation of non-sentential `not` to apply to denotations of form $\lambda x_1 \ldots \lambda x_n.p$ (resulting in $\lambda x_1 \ldots \lambda x_n.\neg p$), we generalize negation to an operation `not` that can skip over arbitrarily many lambda abstractions and is defined as follows (with p of type t):

$$\mathtt{not}\ \lambda x.P = \lambda x.\mathtt{not}\ P$$

$$\mathtt{not}\ p = \neg p$$

Now, for specifying the denotation of non-sentential negation, we use `not` instead of $\neg$: $[\![\mathtt{not}]\!] = \lambda m.(m \star \lambda p.(\mathtt{not} > (\mathtt{some}\ p))$. Then composing the meaning of 13 as shown in 13a results in 13b.

8 Conclusion

We used the state monad to formulate a core concept of dynamic semantics: accessing and updating contexts. All denotations are injected into the monad, thus implicitly carry a state, which is passed on during meaning composition and can be accessed and updated by state affecting operations. Proper names are assumed to add a referent to the state which is kept throughout the whole discourse, thus being available as antecedent without any restrictions. Quantifiers, on the other hand, are assumed to add a referent to the state only temporarily, deleting it again once the scope of the quantifier is closed. However, we saw that this simple mechanism for updating states does not suffice to treat donkey-type anaphora. We therefore proposed a treatment of donkey sentences along the line of Heim, assuming that determiners such as `a` are not quantifiers but free indefinites and implementing quantification with unselective binding. This offers a quite natural treatment of donkey sentences, although suggesting that donkey-type anaphora is not tightly connected to the dynamics of context updates.

Another approach that captures context updates and donkey-type anaphora in a non-dynamic setting is the Montagovian continuation semantics by Philippe de Groote [4] and approaches building on it, e.g., Asher & Pogodalla [1]. The continuation approach differs from ours in lifting only sentence types, while we inject all denotations into the monad. Moreover, the treatment of quantification and donkey sentences differs. For instance, de Groote hands on empty continuations in order to close off the scope of a quantifier, which ultimately requires a distinction between a local context, containing referents that should be deleted, and a global context, containing referents introduced by proper names, that should stay available also outside the scope of quantifiers and negation. Since we tied the removal of referents to binding, the local vs. global distinction amounts to the difference between variables and constants (which cannot be bound and thus are not deleted) in our approach.

The maybe nicest property of a monadic approach is that it allows us to specify denotations in a way such that they consist of the familiar denotation lifted into the monad and composed with state affecting operations, thus separating the truth conditional part of the meaning from the context changing part. The monadic approach thus offers a computational view on meaning composition that separates stateful, i.e. context updating or context accessing operations, from static, truth conditional meaning composition.

Furthermore, a monadic treatment fits nicely into Shan's picture of a modular treatment of semantic phenoma such as intensionality, variable binding, presuppositions, and so on. How our proposed state monad would combine and interact with other monads, however, still remains to be worked out.

References

1. Asher, N., Pogodalla, S.: A Montegovian treatment of modal subordination. In: Proceedings of Semantics and Linguistic Theory XX (2010)
2. Barker, C.: Continuations and the nature of quantification. Natural Language Semantics 10, 211–242 (2002)
3. Bekki, D.: Monads and Meta-lambda Calculus. In: Hattori, H., Kawamura, T., Idé, T., Yokoo, M., Murakami, Y. (eds.) JSAI 2008. LNCS (LNAI), vol. 5447, pp. 193–208. Springer, Heidelberg (2009)
4. de Groote, P.: Towards a Montegovian account of dynamics. In: Proceedings of Semantics and Linguistic Theory XVI (2006)
5. de Groote, P., Lebedeva, E.: Presupposition accommodation as exception handling. In: Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue, SIGDIAL 2010, pp. 71–74 (2010)
6. Groenendijk, J., Stokhof, M.: Dynamic Predicate Logic. Linguistics and Philosophy 14(1), 39–100 (1991)
7. Heim, I.: The Semantics of Definite and Indefinite Noun Phrases. PhD thesis, University of Massachusetts, Amherst (1982)
8. Lewis, D.: Adverbs of quantification. In: Keenan, E. (ed.) Formal Semantics of Natural Language, pp. 3–15. Cambridge University Press (1975)
9. Moggi, E.: Computational lambda-calculus and monads. In: Symposium on Logic in Computer Science, Asilomar, California. IEEE (June 1989)
10. Ogata, N.: Towards computational non-associative lambek lambda-calculi for formal pragmatics. In: Proceedings of the Fifth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS 2008) in Conjunction with the 22nd Annual Conference of the Japanese Society for Artificial Intelligence 2008, pp. 79–102 (2008)
11. Pagin, P., Westerståhl, D.: Predicate logic with flexibly binding operators and natural language semantics. Journal of Logic, Language and Information 2(2), 89–128 (1993)
12. Shan, C.-C.: Monads for natural language semantics. In: Striegnitz, K. (ed.) Proceedings of the 2001 European Summer School in Logic, Language and Information Student Session, pp. 285–298 (2002)
13. Wadler, P.: Comprehending monads. Mathematical Structures in Computer Science 2(4), 461–493 (1992)

Measurement-Theoretic Foundations of Gradable-Predicate Logic

Satoru Suzuki

Faculty of Arts and Sciences, Komazawa University
1-23-1, Komazawa, Setagaya-ku, Tokyo 154-8525, Japan
bxs05253@nifty.com

Abstract. Two approaches have predominated in research on the semantics of gradable adjectives: the Vague Predicate Analysis and the Scalar Analysis. Kennedy discusses four problems of the Vague Predicate Analysis, the two of which can be regarded as main: the Cross-Polarity Problem and the (In)commensurability Problem. The aim of this paper is to propose a new version of logic for gradable adjectives—Gradable-Predicate Logic (GPL). The model of the language $\mathcal{L}_{\mathsf{GPL}}$ of GPL is based on the Vague Predicate Analysis, and can give an answer both to the Cross-Polarity Problem and to the (In)commensurability Problem.

Keywords: cross-polar anomaly, gradable adjective, incommensurability, measurement theory, representation theorem, scalar analysis, uniqueness theorem, vagueness, vague predicate analysis.

1 Introduction

Kennedy [3, pp.3–82] argues on the vagueness of gradable adjectives as follows. Sentences involving gradable adjectives are vague. (1.1), for example, may be true in one context and false in another.

(1.1) The Mars Pathfinder mission is expensive.

(1.1) faces the following problem:

Problem 1 (Vagueness Problem). Suppose that 'a' is a proper name and that 'F' is a gradable adjective. Is then 'a is F' true in the context c of utterance?

Two approaches to the Vagueness Problem have predominated in research on the semantics of gradable adjectives: the *Vague Predicate Analysis* (see, for example, [4]) and the *Scalar Analysis* (see, for example, [3]). There are two primary differences between them. The first difference concerns the semantic type of a gradable adjective. The Vague Predicate Analysis assumes the following condition:

Condition 1 (Sameness Condition). *Gradable adjectives have the same semantic type as such non-gradable adjectives as 'dead' and 'octagonal': they denote functions form individuals to truth values.*

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 82–95, 2012.

The Scalar Analysis, on the other hand, assumes that gradable adjectives denote *relations* between individuals and *degrees*. The second difference concerns the nature of the ordering on the domain of the adjective. Both analyses claim that a partial order can be imposed on the domain of the adjective, but they differ in their assumptions about the ordering on the domain. The Vague Predicate Analysis assumes the following condition:

Condition 2 (Primitiveness Condition). *The ordering is primitive.*

The Scalar Analysis, on the other hand, assumes that the ordering is *derived* in the sense that the adjective imposes an ordering on its domain by relating individuals to degrees on a scale. Kennedy [3] discusses four problems of the Vague Predicate Analysis, the two of which can be regarded as main: the Cross-Polarity Problem and the (In)commensurability Problem. $POS_c(F)$ is a positive extension, which contains only individuals that are F in c. $NEG_c(F)$ is a negative extension, which contains only individuals that are not F in c. The partitioning of the domain into positive and negative extensions is context-dependent, that is, determined by the choice of *comparison class*. A comparison class is a subset of the universe of discourse that is relevant in the context of utterance. The interpretation of a gradable adjective in a context c is a member of a family of functions that partition a partially ordered set. We introduce a *degree function* d that applies to a gradable adjective and returns some member of this family. The underlying idea is that a degree function performs the role normally played by *context*. The Vague Predicate Analysis runs into a problem when confronted with the following *interadjective-comparison* sentence.[1]

(1.2) Mona is happier than Jude is sad.

Kennedy considers (1.2) to be *meaningless*. But the Vague Predicate Analysis can give the truth condition of (1.2) by (1.3).

$$(1.3)\quad \exists d((d(happy))(Mona) \wedge \neg((d(sad))(Jude)).$$

When $\mathcal{I}_{happy(sad)}$ is a comparison class relative to 'happy(sad)' and $\succ_{happy(sad)}$ is a partial order relative to 'happy(sad)', consider a context in which the following conditions are satisfied:

$$(1.4)\quad \begin{array}{l} \mathcal{I}_{happy} = \mathcal{I}_{sad} = \{Albert, Bernard, Catherine, Jude, Mona\}, \\ Mona \succ_{happy} Catherine \succ_{happy} Jude \succ_{happy} Bernard \succ_{happy} Albert, \\ Albert \succ_{sad} Bernard \succ_{sad} Jude \succ_{bad} Catherine \succ_{sad} Mona. \end{array}$$

In this context, there is a degree function d that satisfies (1.3), for example, the one that induces the partitioning shown in (1.5):

$$(1.5)\quad \begin{array}{l} POS_d(happy) = NEG_d(sad) = \{Catherine, Jude, Mona\}, \\ NEG_d(happy) = POS_d(sad) = \{Albert, Bernard\}. \end{array}$$

---

[1] In [17] we propose a logic called ICL designed especially for various interadjective comparisons.

As a result, (1.2) should be true in (1.5). The Vague Predicate Analysis cannot explain why comparatives like (1.2) constructed out of *cross-polar* pairs of adjectives should be meaningless in terms of the semantics of gradable adjectives and the comparative construction. However, for example, van Rooij [9] may consider (1.2) to be meaningful. Now the following problem arises:

Problem 2 (Cross-Polarity Problem)

1. How can anyone who considers comparatives like (1.2) constructed out of *cross-polar* pairs of adjectives to be *meaningless* explain *why* they are so?
2. On the other hand, in what *model* can anyone who considers comparatives like (1.2) to be *meaningful* evaluate them?

The Vague Predicate Analysis runs into another problem when confronted with the following sentence:

(1.6) My copy of *The Brothers Karamazov* is heavier than my copy of *The Idiot* is old.

Kennedy considers (1.6) to be *meaningless*. But the Vague Predicate Analysis can give the truth condition of (1.6) by (1.7).

$$\exists d((d(heavy))(The\ Brothers\ Karamazov) \wedge \neg((d(old))(The\ Idiot)). \quad (1.7)$$

Consider a context in which the following conditions are satisfied:

$$\begin{aligned}&\mathcal{I}_{heavy} = \mathcal{I}_{old}\\&= \{Crime\ and\ Punishment, The\ Brothers\ Karamazov,\\&The\ Devils, The\ Idiot\},\\&The\ Brothers\ Karamazov \succ_{heavy} The\ Idiot\\&\succ_{heavy} The\ Devils \succ_{heavy} Crime\ and\ Punishment,\\&Crime\ and\ Punishment \succ_{old} The\ Brothers\ Karamazov\\&\succ_{old} The\ Devils \succ_{old} The\ Idiot.\end{aligned} \quad (1.8)$$

In this context, there is a degree function d that satisfies (1.8), for example, the one that induces the partitioning shown in (1.9):

$$\begin{aligned}&POS_d(heavy) = \{The\ Brothers\ Karamazov\},\\&NEG_d(heavy) = \{The\ Idiot, The\ Devils, Crime\ and\ Punishment\},\\&POS_d(old) = \{Crime\ and\ Punishment\},\\&NEG_d(old) = \{The\ Brothers\ Karamazov, The\ Devils, The\ Idiot\}.\end{aligned} \quad (1.9)$$

As a result, (1.6) should be true in (1.9). The Vague Predicate Analysis cannot explain why such adjectives as in (1.6) should be *incommensurable* in terms of the semantics of gradable adjectives and the comparative construction. However, for example, van Rooij [9] and Bale [1] may consider such adjectives as in (1.6) to be *commensurable*. Now the following problem arises:

Problem 3 ((In)commensurability Problem)

1. How can anyone who considers such adjectives as in (1.6) to be *incommensurable* explain *why* they are so?

2. On the other hand, in what *model* can anyone who considers such adjectives as in (1.6) to be *commensurable* evaluate comparatives like (1.6)?

By means of comparative constructions, we can compare individuals according to different properties. Such comparisons may be divided into two types: *direct* and *indirect* comparisons. The former are comparisons of *direct measurements.* As an example of them, we can give the following sentence:

(1.10) Albert is taller than he is wide.

The latter are comparisons of *relative positions on different scales* as in

(1.11) Albert is handsomer than Catherine is intelligent.

Moreover, the following sentences are examples of direct comparisons:

(1.12) Catherine is twice as tall as Doris is wide.
(1.13) Albert is 10 centimetres taller than Bernard is wide.

The aim of this paper is to propose a new version of logic for gradable adjectives—Gradable-Predicate Logic (GPL). The model of the language $\mathcal{L}_{\mathsf{GPL}}$ of GPL is based on the Vague Predicate Analysis, can give an answer both to the Cross-Polarity Problem and to the (In)commensurability Problem, and can give the truth conditions of such sentences as (1.10), (1.11), (1.12), and (1.13).

The structure of this paper is as follows. We define the language $\mathcal{L}_{\mathsf{GPL}}$ of GPL, give descriptions of meaningfulness, scale types, magnitude estimation, and measurement theory, define a model $\mathfrak{M}$ of $\mathcal{L}_{\mathsf{GPL}}$, formulate the representation theorem for interadjective-comparison ordering and the uniqueness theorem for it, formulate the representation theorem for magnitude estimation and the uniqueness theorem for it, provide GPL with a satisfaction definition and a truth definition, touch upon the non first-order axiomatisability of models of $\mathcal{L}_{\mathsf{GPL}}$, furnish solutions to the cross-polarity problem and the (in)commensurability problem, and give the truth conditions of (1.10), (1.11), (1.12), and (1.13).

2 Gradable-Predicate Logic GPL

2.1 Language

We define the language $\mathcal{L}_{\mathsf{GPL}}$ of GPL in which gradable adjectives in natural languages can be expressed by gradable predicate symbols as follows:

Definition 1 (Language). *Let $\mathcal{V}$ denote a set of individual variables, $\mathcal{C}$ a set of individual constants, $\mathcal{P}$ a set of n one-place gradable predicate symbols. The language $\mathcal{L}_{\mathsf{GPL}}$ of* GPL *is given by the following rule:*

$$
\begin{aligned}
&t ::= x \mid a,\\
&\varphi ::= P_i(t) \mid t_i = t_j \mid \mathbf{NER}_{P_i,P_j}(t_i,t_j) \mid \mathbf{LER}_{P_i,P_j}(t_i,t_j)\\
&\mid \mathbf{TER}^k_{P_i,P_j}(t_i,t_j) \mid \mathbf{UER}^k_{P_i,P_j}(t_i,t_j)\\
&\mid \top \mid \neg\varphi \mid \varphi \wedge \varphi \mid \forall x\varphi,\ \textit{where } x \in \mathcal{V},\ a \in \mathcal{C},\ \textit{and } P_i, P_j \in \mathcal{P}.
\end{aligned}
$$

- $\mathbf{NER}_{P_i,P_j}(t_i,t_j)$ *means that* t_i *is quantitatively* P_i*-er than* t_j *is* P_j*.*
- $\mathbf{LER}_{P_i,P_j}(t_i,t_j)$ *means that* t_i *is qualitatively* P_i*-er than* t_j *is* P_j*.*
- $\mathbf{TER}^k_{P_i,P_j}(t_i,t_j)$ *means that* t_i *is* k *times as* P_i *as* t_j *is* P_j*.*
- $\mathbf{UER}^k_{P_i,P_j}(t_i,t_j)$ *means that* t_i *is* P_i*-er than* t_j *is* P_j *by* k *units of measurement (e.g. (centi)metre, (kilo)gram, ...).*

The set of all well-formed formulae of $\mathcal{L}_{\mathsf{GPL}}$ *is denoted by* $\Phi_{\mathcal{L}_{\mathsf{GPL}}}$*.*

2.2 Semantics

Meaningfulness and Scale Types. Roberts [8, p.52, pp.57–59] argues on the *meaningfulness* of sentences involving *scales*. He begins with the the following sentences and considers which seem to be meaningful.

> (2.1) The number of cans of corn in the local super market at closing time yesterday was at least 10.
> (2.2) One can of corn weighs at least 10.
> (2.3) One can of corn weighs twice as much as a second.
> (2.4) The temperature of one can of corn at closing time yesterday was twice as much as that of a second time.

(2.1) seems to be meaningful, but (2.2) does not, for the number of cans is specified without reference to a particular scale of measurement, whereas the weight of a can is not. Similarly, (2.3) seems to be meaningful, but (2.4) does not, for the ratio of weights is the same regardless of measurement used, whereas that of temperature is not necessarily the same. Meaningfulness can be studied by analysing the following *admissible transformations* of scale defined by the concept of a *homomorphism*:

Definition 2 (Homomorphism). *Suppose a relational system* $\mathfrak{U} := (A, R_1, R_2, \ldots, R_p, \circ_1, \circ_2, \ldots, \circ_q)$ *and another* $\mathfrak{V} := (B, R'_1, R'_2, \ldots, R'_p, \circ'_1, \circ'_2, \ldots, \circ'_q)$*, where* A *and* B *are sets,* $R_1, R_2, \ldots, R_p$ *are relations on* A*,* $R'_1, R'_2, \ldots, R'_p$ *are relations on* B*,* $\circ_1, \circ_2, \ldots, \circ_q$ *are operations on* A*, and* $\circ'_1, \circ'_2, \ldots, \circ'_q$ *are operations on* B*.* f *is called a homomorphism from* $\mathfrak{U}$ *into* $\mathfrak{V}$ *if, for any* $a_1, a_2, \ldots, a_{r_i} \in A$*,*

$$R_i(a_1, a_2, \ldots, a_{r_i}) \text{ iff } R'_i(f(a_1), f(a_2), \ldots, f(a_{r_i})),\ i = 1, 2, \ldots, p,$$

and for any $a, b \in A$*,*

$$f(a \circ_i b) = f(a) \circ'_i f(b),\ i = 1, 2, \ldots, q.$$

Definition 3 (Admissible Transformation of Scale). *Suppose that a scale* f *is one homomorphism from a relational system* $\mathfrak{U}$ *into another* $\mathfrak{V}$*, and suppose that* A *is the set underlying* $\mathfrak{U}$ *and* B *is the set underlying* $\mathfrak{V}$*. Suppose that* Φ *is a function that maps the range of* f*, that is, the set* $f(A) := \{f(a) : a \in A\}$ *into* B*. Then the composition* $\Phi \circ f$ *is a function from* A *into* B*. If* $\Phi \circ f$ *is a homomorphism from* $\mathfrak{U}$ *into* $\mathfrak{V}$*, we call* Φ *an admissible transformation of* f*.*

The following provides an example:

Example 1. Suppose $\mathfrak{U} := (\mathbb{N}, >)$, $\mathfrak{V} := (\mathbb{R}, >)$, and $f : \mathbb{N} \to \mathbb{R}$ is given by $f(x) := 2x$. Then f is a homomorphism from $\mathfrak{U}$ into $\mathfrak{V}$. If $\Phi(x) := x + 5$, then $\Phi \circ f$ is a homomorphism from $\mathfrak{U}$ into $\mathfrak{V}$, for we have $(\Phi \circ f)(x) = 2x + 5$, and

$$x > y \text{ iff } 2x + 5 > 2y + 5.$$

Thus, $\Phi : f(A) \to B$ is an admissible transformation of f. However, if $\Phi(x) := -x$ for any $x \in f(A)$, then Φ is not an admissible transformation, for $\Phi \circ f$ is not a homomorphism from $\mathfrak{U}$ into $\mathfrak{V}$.

We define meaningfulness in terms of admissible transformations as follows:

Definition 4 (Meaningfulness). *A sentence involving scales is meaningful iff the truth or falsity is unchanged under admissible transformations of all the scales in question.*

Roberts [8, pp.64–67] defines *scale types* in terms of the class of admissible transformations as follows:

1. The simplest example of a scale is where only admissible transformation is $\Phi(x) = x$. Such a scale is called an *absolute scale*. Counting is an example of an absolute scale.
2. When the admissible transformations are all the functions $\Phi : f(A) \to B$ of the form $\Phi(x) = \alpha x,\ \alpha > 0$, Φ is called a *similarity transformation*, and a scale with the similarity transformations as its class of admissible transformations is called a *ratio scale*. Mass and temperature on the Kelvin scale are examples of ratio scales. According to Stevens [10], various *sensations* such as loudness and brightness can also be measured in ratio scales.
3. When the admissible transformations are all the functions $\Phi : f(A) \to B$ of the form $\Phi(x) = \alpha x + \beta,\ \alpha > 0$, Φ is called a *positive linear transformation*, and a corresponding scale is called an *interval scale*. Temperature on the Fahrenheit scale and temperature on the Celsius scale are examples of interval scales.
4. When a scale is unique up to *order*, the admissible transformations are monotone increasing functions $\Phi(x)$ satisfying the condition that $x \geq y$ iff $\Phi(x) \geq \Phi(y)$. Such scales are called *ordinal scales*. The Mohs scale of hardness is an example of an ordinal scale.
5. In some scales, all one-to-one functions Φ define admissible transformations. Such scales are called *nominal scales*. Examples of nominal scales are numbers on the uniforms of baseball players.
6. A scale is called a *log-interval scale* if the admissible transformations are functions of the form $\Phi(x) = \alpha x^{\beta},\ \alpha, \beta > 0$. Log-interval scales are important in *psychophysics*, where they are considered as scale types for the *psychophysical laws* relating a physical quantity (for example, intensity of a sound) to psychological quantity (for example, loudness of a sound).
7. When the admissible transformations are functions of the form $\Phi(x) = x + \beta$, a corresponding scale is a *difference scale*. The so-called Thurstone Case V scale, which is a measure of response strength, is an example of a difference scale.

Magnitude Estimation, Cross-Modality Matching, and Interadjective Comparison. Judgments of subjective loudness can be made in laboratory in various ways. Stevens [10] classifies four methods. The method of *magnitude estimation* is one of the most common. The following provides an example:

Example 2 (Magnitude Estimation). A subject hears a reference sound and is told to assign it a fixed number. Then he is presented other sounds and asked to assign them numbers proportional to the reference sound.

Stevens argues that magnitude estimation gives rise to a *ratio scale*. Moreover, he uses the idea of *cross-modality matching* to test the *power law*. It must be noted that the scale corresponding to the power law is a *log-interval scale*. Krantz [5] puts Stevens's argument that magnitude estimation gives rise to a ratio scale and his idea of cross-modality matching to test the power law on a rigorous *measurement-theoretic* foundation. In this paper, we try to propose a logic for *interadjective comparison* the model of which is based on Krantz's measurement theory for magnitude estimation and cross-modality matching. There are two main problems with measurement theory[2]:

1. the *representation problem*—justifying the assignment of numbers to objects, and
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 specified conditions on a qualitative relational system are (necessary and) sufficient for the assignment of numbers to objects that represents (preserves) all the relations in the system. A solution to the latter can be furnished by a *uniqueness theorem*, which specifies the transformation up to which this assignment is unique.

Model. We define a model $\mathfrak{M}$ of $\mathcal{L}_{\mathsf{GPL}}$ as follows:

Definition 5 (Model). $\mathfrak{M}$ *is a sequence* $(\mathcal{I}_{F_1}, \dots, \mathcal{I}_{F_n}, a_1^{\mathfrak{M}}, b_1^{\mathfrak{M}}, \dots, \spadesuit_{F_1}, \dots, \spadesuit_{F_n}, \heartsuit_{F_1}, \dots, \heartsuit_{F_n}, F_1^{\mathfrak{M}}, \dots, F_n^{\mathfrak{M}}, \succeq)$*, where*

- $\mathcal{I}_{F_i}$ *is a nonempty set of individuals for evaluation of* F_i*, called a comparison class relative to* F_i*.*
- $a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}, \dots \in \mathcal{I}_{F_i}$*.*
- $\spadesuit_{F_i}$ *is an average individual in* $\mathcal{I}_{F_i}$*.*
- $\heartsuit_{F_i}$ *is a zero-point individual in* $\mathcal{I}_{F_i}$*.*
- $F_i^{\mathfrak{M}} \subseteq \mathcal{I}_{F_i}$*.*
- $\succeq$ *is a binary relation on* $\bigcup_{i=1}^{n} \mathcal{I}_{F_i} \times \mathcal{I}_{F_i}$*, called the interadjective-comparison ordering relation, that satisfies the following conditions:*
  1. $\succeq$ *is a weak order (transitive and connected).*
  2. *For any* $a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}} \in \mathcal{I}_{F_i}$ *and any* $a_j^{'\mathfrak{M}}, b_j^{'\mathfrak{M}} \in \mathcal{I}_{F_j}$*, if* $(a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}) \succeq (a_j^{'\mathfrak{M}}, b_j^{'\mathfrak{M}})$*, then* $(b_j^{'\mathfrak{M}}, a_j^{'\mathfrak{M}}) \succeq (b_i^{\mathfrak{M}}, a_i^{\mathfrak{M}})$*.*
  3. *For any* $a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}, c_i^{\mathfrak{M}} \in \mathcal{I}_{F_i}$ *and any* $a_j^{'\mathfrak{M}}, b_j^{'\mathfrak{M}}, c_j^{'\mathfrak{M}} \in \mathcal{I}_{F_j}$*, if* $(a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}) \succeq (a_j^{'\mathfrak{M}}, b_j^{'\mathfrak{M}})$ *and* $(b_i^{\mathfrak{M}}, c_i^{\mathfrak{M}}) \succeq (b_j^{'\mathfrak{M}}, c_j^{'\mathfrak{M}})$*, then* $(a_i^{\mathfrak{M}}, c_i^{\mathfrak{M}}) \succeq (a_j^{'\mathfrak{M}}, c_j^{'\mathfrak{M}})$*.*

---

[2] [8] gives a comprehensive survey of measurement theory. The mathematical foundation of measurement had not been studied before Hölder [2] developed his axiomatisation for the measurement of mass. [6], [11] and [7] are seen as milestones in the history of measurement theory.

4. *For any* $a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}} \in \mathcal{I}_{F_i}$*, there exist* $a_1'^{\mathfrak{M}}, b_1'^{\mathfrak{M}} \in \mathcal{I}_{F_1}$ *such that* $(a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}) \sim (a_1'^{\mathfrak{M}}, b_1'^{\mathfrak{M}})$*, where* $(a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}) \sim (a_j'^{\mathfrak{M}}, b_j'^{\mathfrak{M}}) := (a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}) \succeq (a_j'^{\mathfrak{M}}, b_j'^{\mathfrak{M}})$ *and* $(a_j'^{\mathfrak{M}}, b_j'^{\mathfrak{M}}) \succeq (a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}})$.
5. *For any* $a_1^{\mathfrak{M}}, b_1^{\mathfrak{M}}, c_1^{\mathfrak{M}}, d_1^{\mathfrak{M}} \in \mathcal{I}_{F_1}$*, if* $(d_1^{\mathfrak{M}}, c_1^{\mathfrak{M}}) \succeq (a_1^{\mathfrak{M}}, b_1^{\mathfrak{M}}) \succeq (d_1^{\mathfrak{M}}, d_1^{\mathfrak{M}})$*, then there exist* $a_1'^{\mathfrak{M}}, b_1'^{\mathfrak{M}} \in \mathcal{I}_{F_1}$ *such that* $(d_1^{\mathfrak{M}}, b_1'^{\mathfrak{M}}) \sim (a_1^{\mathfrak{M}}, b_1^{\mathfrak{M}}) \sim (a_1'^{\mathfrak{M}}, c_1^{\mathfrak{M}})$.
6. *Suppose that* $a_1^{(1)\mathfrak{M}}, a_1^{(2)\mathfrak{M}}, \ldots, a_1^{(i)\mathfrak{M}}, \ldots$ *is a sequence of equally spaced elements of* $\mathcal{I}_{F_1}$*, that is,* $(a_1^{(i+1)\mathfrak{M}}, a_1^{(i)\mathfrak{M}}) \sim (a_1^{(2)\mathfrak{M}}, a_1^{(1)\mathfrak{M}}) \succ (a_1^{(1)\mathfrak{M}}, a_1^{(1)\mathfrak{M}})$ *for any* $a_1^{(i+1)\mathfrak{M}}, a_1^{(i)\mathfrak{M}}$ *in the sequence. If the sequence is strictly bounded (that is, if there exist* $b_1^{\mathfrak{M}}, c_1^{\mathfrak{M}} \in \mathcal{I}_{F_1}$ *such that* $(b_1^{\mathfrak{M}}, c_1^{\mathfrak{M}}) \succ (a_1^{(i)\mathfrak{M}}, a_1^{(1)\mathfrak{M}})$ *for any* $a_1^{(i)\mathfrak{M}}$ *in the sequence), then it is finite.*

Remark 1 (Primitiveness Condition). In the model $\mathfrak{M}$ of $\mathcal{L}_{\mathsf{GPL}}$, $\succeq$ is primitive. So GPL satisfies the Primitiveness Condition.

Condition 2 postulates that reversing pairs should be reversing the ordering. The following provides an example:

Example 3 (Reversal of Pairs). In (1.11), if a_i is much handsomer than b_i, and a_j' is slightly more intelligent than b_j', so that $(a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}) \succeq (a_j'^{\mathfrak{M}}, b_j'^{\mathfrak{M}})$, then b_i is much uglier than a_i, but b_j' is only slightly duller than a_j', so that $(b_j'^{\mathfrak{M}}, a_j'^{\mathfrak{M}}) \succeq (b_i^{\mathfrak{M}}, a_i^{\mathfrak{M}})$.

Condition 3 says that pairs $(a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}})$ behave *qualitatively* like ratios with respect to $\succeq$. The following provides an example:

Example 4 (Ratios). Pairs $(a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}})$ behave with respect to $\succeq$ in much the same as $\frac{a_i}{b_i} \geq \frac{a_j'}{b_j'}$ and $\frac{b_i}{c_i} \geq \frac{b_j'}{c_j'}$ implies $\frac{a_i}{c_i} \geq \frac{a_j'}{c_j'}$ for positive real numbers.

Condition 4 postulates that any pair $(a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}})$ should be equivalent to some $\mathcal{I}_{F_1} \times \mathcal{I}_{F_1}$ pair. Condition 5 postulates that intermediate-level individuals can be chosen *densely* within $\mathcal{I}_{F_1}$. Condition 6 postulates that $\succeq$ should have an *Archimedean Property.* The following provides an example:

Example 5 (Archimedean Property). However small the pair $(a_1^{\mathfrak{M}}, b_1^{\mathfrak{M}})$ in handsomeness may be, if one can find a sequence of handsomeness $a_1^{(1)\mathfrak{M}}, a_1^{(2)\mathfrak{M}}, \ldots$ such that any pair $(a_1^{(i+1)\mathfrak{M}}, a_1^{(i)\mathfrak{M}})$ in handsomeness is equivalent to $(a_1^{\mathfrak{M}}, b_1^{\mathfrak{M}})$ in handsomeness, then the overall interval $(a_1^{(i)\mathfrak{M}}, a_1^{(1)\mathfrak{M}})$ in handsomeness becomes indefinitely large.

We can prove the following representation theorem for interadjective-comparison ordering by modifying the method of [6]:

Theorem 1 (Representation for Interadjective-Comparison Ordering). *If* $\succeq$ *is an interadjective-comparison ordering relation of Definition 5, then there exist functions* $f_i : \mathcal{I}_{F_i} \to \mathbb{R}^+$ $(1 \leq i \leq n)$ *such that for any* $a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}} \in \mathcal{I}_{F_i}$ $(1 \leq i \leq n)$ *and any* $a_j'^{\mathfrak{M}}, b_j'^{\mathfrak{M}} \in \mathcal{I}_{F_j}$ $(1 \leq j \leq n)$,

$$(2.5) \quad (a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}) \succeq (a_j'^{\mathfrak{M}}, b_j'^{\mathfrak{M}}) \text{ iff } \frac{f_i(a_i^{\mathfrak{M}})}{f_i(b_i^{\mathfrak{M}})} \geq \frac{f_j(a_j'^{\mathfrak{M}})}{f_j(b_j'^{\mathfrak{M}})}.$$

We can also prove the following uniqueness theorem for interadjective-comparison ordering by modifying the method of [6]:

Theorem 2 (Uniqueness for Interadjective-Comparison Ordering). *If* f_i' $(1 \leq i \leq n)$ *are any other such functions as* f_i *of Theorem 1, then there exist* $\alpha_i, \beta \in \mathbb{R}^+$ $(1 \leq i \leq n)$ *such that*

$$f_i' = \alpha_i f_i^{\beta}.$$

Remark 2 (Log-Interval Scale). f_i defines a log-interval scale.

2.3 Magnitude Estimation and Ratio Scale

We specify some conditions under which *magnitude estimation* leads to a *ratio scale*[3]. These conditions follow from the following three consistency conditions:

1. the magnitude-pair consistency condition,
2. the pair consistency condition, and
3. the magnitude-interadjective-comparison consistency condition.

We state these three consistency conditions. Suppose that an experimenter is performing a magnitude estimation of a subject on $\mathcal{I}_{F_i}$. The experimenter fixes $a_i^{\mathfrak{M}} \in \mathcal{I}_{F_i}$ and assigns to $a_i^{\mathfrak{M}}$ the psychological magnitude p. We assume that all magnitudes are positive. Next, the experimenter asks the subject to assign to each $b_i^{\mathfrak{M}} \in \mathcal{I}_{F_i}$ a magnitude q depending on $a_i^{\mathfrak{M}}$ and p. This is written in symbols as follows:

$$ME_i(b_i^{\mathfrak{M}} | a_i^{\mathfrak{M}}, p) = q,$$

where ME_i is the magnitude estimate for $b_i^{\mathfrak{M}}$ when $a_i^{\mathfrak{M}}$ is assigned a magnitude p. In particular,

$$ME_i(a_i^{\mathfrak{M}} | a_i^{\mathfrak{M}}, p) = p.$$

In the variant of magnitude estimation called *pair estimation*, a pair of $a_i^{\mathfrak{M}}$ and $b_i^{\mathfrak{M}}$ from $\mathcal{I}_{F_i}$ are presented, and then the experimenter asks the subject to give a numerical estimate of, as it were, the *sensation ratio* of $a_i^{\mathfrak{M}}$ to $b_i^{\mathfrak{M}}$. We denote this estimate by $PE_i(a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}})$. It is reasonable to assume that PE_i *corresponds to* $\succeq$ as follows:

$$(2.6) \quad (a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}) \succeq (c_j^{\mathfrak{M}}, d_j^{\mathfrak{M}}) \text{ iff } PE_i(a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}) \geq PE_j(c_j^{\mathfrak{M}}, d_j^{\mathfrak{M}}).$$

Magnitude estimates and pair estimates are often assumed to satisfy the following *magnitude-pair consistency condition*: for any $c_i^{\mathfrak{M}} \in \mathcal{I}_{F_i}$ and any $p \in \mathbb{R}^+$,

$$(2.7) \quad PE_i(a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}) = \frac{ME_i(a_i^{\mathfrak{M}} | c_i^{\mathfrak{M}}, p)}{ME_i(b_i^{\mathfrak{M}} | c_i^{\mathfrak{M}}, p)}.$$

---

[3] We owe this subsection to [8, pp.186–192].

Moreover, it is often assumed that pair estimates behave like ratios, that is, they satisfy the following *pair consistency condition*: for any $a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}, c_i^{\mathfrak{M}} \in \mathcal{I}_{F_i}$,

$$(2.8) \quad PE_i(a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}) \cdot PE_i(b_i^{\mathfrak{M}}, c_i^{\mathfrak{M}}) = PE_i(a_i^{\mathfrak{M}}, c_i^{\mathfrak{M}}).$$

If $i = 1$, (2.8) yields

$$(2.9) \quad PE_1(a_1^{\mathfrak{M}}, b_1^{\mathfrak{M}}) \cdot PE_1(b_1^{\mathfrak{M}}, c_1^{\mathfrak{M}}) = PE_1(a_1^{\mathfrak{M}}, c_1^{\mathfrak{M}}).$$

In *interadjective-comparison matching*, we usually fix $a_i^{\mathfrak{M}} \in \mathcal{I}_{F_i}$ and $a_j^{\mathfrak{M}} \in \mathcal{I}_{F_j}$ and say that they match. The experimenter then asks the subject to find $b_i^{\mathfrak{M}} \in \mathcal{I}_{F_i}$ that matches a given individual $b_j^{\mathfrak{M}} \in \mathcal{I}_{F_j}$. This is written in symbols as follows:

$$IM_{ji}(b_j^{\mathfrak{M}} | a_j^{\mathfrak{M}}, a_i^{\mathfrak{M}}) = b_i^{\mathfrak{M}}.$$

In particular,

$$IM_{ji}(a_j^{\mathfrak{M}} | a_j^{\mathfrak{M}}, a_i^{\mathfrak{M}}) = a_i^{\mathfrak{M}}.$$

It is reasonable to assume that *if $a_j^{\mathfrak{M}}$ is matched by $a_i^{\mathfrak{M}}$ and $b_j^{\mathfrak{M}}$ by $b_i^{\mathfrak{M}}$, then the corresponding sensation ratios are judged equal*:

$$(2.10) \quad \text{If } IM_{ji}(b_j^{\mathfrak{M}} | a_j^{\mathfrak{M}}, a_i^{\mathfrak{M}}) = b_i^{\mathfrak{M}}, \text{ then } (b_j^{\mathfrak{M}}, a_j^{\mathfrak{M}}) \sim (b_i^{\mathfrak{M}}, a_i^{\mathfrak{M}}).$$

It is often assumed that magnitude estimation and interadjective-comparison matching are related by the following *magnitude-interadjective-comparison consistency condition*:

$$(2.11) \quad \frac{ME_i(IM_{ji}(b_j^{\mathfrak{M}} | a_j^{\mathfrak{M}}, a_i^{\mathfrak{M}}) | c_i^{\mathfrak{M}}, p)}{ME_i(a_i^{\mathfrak{M}} | c_i^{\mathfrak{M}}, p)} = \frac{ME_j(b_j^{\mathfrak{M}} | c_j^{\mathfrak{M}}, q)}{ME_j(a_j^{\mathfrak{M}} | c_j^{\mathfrak{M}}, q)}.$$

That is, if $b_j^{\mathfrak{M}}$ is matched with $b_i^{\mathfrak{M}}$ in the interadjective-comparison matching, where $a_j^{\mathfrak{M}}$ is given as matched with $a_i^{\mathfrak{M}}$, then the ratio of the magnitude estimate of $b_i^{\mathfrak{M}}$ to the magnitude estimate of $a_i^{\mathfrak{M}}$ on the ith adjective equals the ratio of the magnitude of $b_j^{\mathfrak{M}}$ to the magnitude estimate of $a_j^{\mathfrak{M}}$ on the jth adjective for any reference estimate p for $c_i^{\mathfrak{M}}$ and q for $c_j^{\mathfrak{M}}$. If $IM_{ji}(b_j^{\mathfrak{M}} | a_j^{\mathfrak{M}}, a_i^{\mathfrak{M}}) = b_i^{\mathfrak{M}}$ and $a_i^{\mathfrak{M}} = c_i^{\mathfrak{M}}$, (2.7) and (2.11) yield

$$(2.12) \quad \frac{ME_i(b_i^{\mathfrak{M}} | a_i^{\mathfrak{M}}, p)}{p} = PE_j(b_j^{\mathfrak{M}}, a_j^{\mathfrak{M}})$$

because $ME_i(a_i^{\mathfrak{M}} | a_i^{\mathfrak{M}}, p) = p$. It is reasonable to assume that if $(b_i^{\mathfrak{M}}, a_i^{\mathfrak{M}}) \sim (b_1^{\mathfrak{M}}, a_1^{\mathfrak{M}})$, then (2.12) holds for $j = 1$:

$$(2.13) \quad \text{If } (b_i^{\mathfrak{M}}, a_i^{\mathfrak{M}}) \sim (b_1^{\mathfrak{M}}, a_1^{\mathfrak{M}}), \text{ then } \frac{ME_i(b_i^{\mathfrak{M}} | a_i^{\mathfrak{M}}, p)}{p} = PE_1(b_1^{\mathfrak{M}}, a_1^{\mathfrak{M}}).$$

We can prove the following representation theorem for magnitude estimation by modifying the method of [5]:

Theorem 3 (Representation for Magnitude Estimation). *Suppose that* $\succeq$ *is an interadjective-comparison ordering relation of Definition 5. Moreover, suppose that* $\succeq, ME_i, PE_i$ *and* IM_{ji} *satisfy (2.6), (2.9), (2.10) and (2.13). Then there exists a power function* $f : \mathbb{R}^+ \to \mathbb{R}^+$ *such that if* f_i*s satisfy (2.5), then*

$$(2.14) \quad ME_i(b_i^{\mathfrak{M}}|a_i^{\mathfrak{M}}, p) = q \textit{ iff } \frac{f_i(b_i^{\mathfrak{M}})}{f_i(a_i^{\mathfrak{M}})} = \frac{f(q)}{f(p)},$$

$$(2.15) \quad PE_i(a_i^{\mathfrak{M}}, b_i^{\mathfrak{M}}) = r \textit{ iff } \frac{f_i(a_i^{\mathfrak{M}})}{f_i(b_i^{\mathfrak{M}})} = f(r), \textit{ and}$$

$$(2.16) \quad \textit{If } IM_{ji}(b_j^{\mathfrak{M}}|a_j^{\mathfrak{M}}, a_i^{\mathfrak{M}}) = b_i^{\mathfrak{M}}, \textit{ then } \frac{f_i(b_i^{\mathfrak{M}})}{f_j(b_j^{\mathfrak{M}})} = \frac{f_i(a_i^{\mathfrak{M}})}{f_j(a_j^{\mathfrak{M}})}.$$

We can also prove the following uniqueness theorem for magnitude estimation by modifying the method of [5]:

Theorem 4 (Uniqueness for Magnitude Estimation). *If* f_i' *and* f' *also satisfy (2.5) and (2.14) through (2.16), then there exist* $\alpha_i, \beta \in \mathbb{R}^+$ $(1 \le i \le n)$ *such that*

$$f_i' = \alpha_i f_i^{\beta} \textit{ and } f' = f^{\beta}.$$

Remark 3 (Log-Interval Scale). Both f_i and f define log-interval scales.

We now obtain the following corollary of Theorem 3 and Theorem 4:

Corollary 1 (Ratio Scale). ME_i *is a ratio scale.*

Satisfaction, Truth and Validity. We define an assignment function and its extension as follows:

Definition 6 (Assignment Function and Its Extension). *Let* $\mathcal{V}$ *denote a set of individual variables,* $\mathcal{C}$ *a set of individual constants and* $\mathcal{I}$ *a set of individuals.*

- *We call* $s : \mathcal{V} \to \mathcal{I}$ *an assignment function.*
- *We define the extension of* s *as a function* $\tilde{s} : \mathcal{V} \cup \mathcal{C} \to \mathcal{I}$ *such that*
  1. *For any* $x \in \mathcal{V}$, $\tilde{s}(x) = s(x)$, *and*
  2. *For any* $a \in \mathcal{C}$, $\tilde{s}(a) = a^{\mathfrak{M}}$.

We provide GPL with the following satisfaction definition relative to $\mathfrak{M}$, define the truth in $\mathfrak{M}$ by means of satisfaction, and then define validity as follows:

Definition 7 (Satisfaction, Truth and Validity). *What it means for* $\mathfrak{M}$ *to satisfy* $\varphi \in \Phi_{\mathcal{L}_{\mathsf{GPL}}}$ *with* s, *in symbols* $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \varphi[s]$ *is inductively defined as follows:*

- $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} P(t)[s]$ *iff* $\tilde{s}(t) \in P^{\mathfrak{M}}$.
- $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} t_1 = t_2[s]$ *iff* $\tilde{s}(t_1) = \tilde{s}(t_2)$.

- $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \mathbf{NER}_{P_i,P_j}(t_i,t_j)[s]$ *iff if* $\succeq, ME_i, ME_j, PE_i, PE_j$ *and* IM_{ji} *satisfy (2.6), (2.9), (2.10) and (2.13), then* $\tilde{s}(t_i) \in \mathcal{I}_{P_i}$ *and* $\tilde{s}(t_j) \in \mathcal{I}_{P_j}$ *and* $ME_i(\tilde{s}(t_i)|\heartsuit_{P_i}, 0) > ME_j(\tilde{s}(t_j)|\heartsuit_{P_j}, 0)$.
- $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \mathbf{LER}_{P_i,P_j}(t_i,t_j)[s]$ *iff* $\tilde{s}(t_i) \in \mathcal{I}_{P_i}$ *and* $\tilde{s}(t_j) \in \mathcal{I}_{P_j}$ *and* $(\tilde{s}(t_i), \spadesuit_{P_i}) \succ (\tilde{s}(t_j), \spadesuit_{P_j})$*, where* $(\tilde{s}(t_i), \tilde{s}(t_j)) \succ (\tilde{s}(t_k), \tilde{s}(t_l)) := (\tilde{s}(t_k), \tilde{s}(t_l)) \not\succeq (\tilde{s}(t_i), \tilde{s}(t_j))$.
- $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \mathbf{TER}^k_{P_i,P_j}(t_i,t_j)[s]$ *iff if* $\succeq, ME_i, ME_j, PE_i, PE_j$ *and* IM_{ji} *satisfy (2.6), (2.9), (2.10) and (2.13), then* $\tilde{s}(t_i) \in \mathcal{I}_{P_i}$ *and* $\tilde{s}(t_j) \in \mathcal{I}_{P_j}$ *and* $ME_i(\tilde{s}(t_i)|\heartsuit_{P_i}, 0) = k \cdot ME_j(\tilde{s}(t_j)|\heartsuit_{P_j}, 0)$.
- $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \mathbf{UER}^k_{P_i,P_j}(t_i,t_j)[s]$ *iff if* $\succeq, ME_i, ME_j, PE_i, PE_j$ *and* IM_{ji} *satisfy (2.6), (2.9), (2.10) and (2.13), then* $\tilde{s}(t_i) \in \mathcal{I}_{P_i}$ *and* $\tilde{s}(t_j) \in \mathcal{I}_{P_j}$ *and* $ME_i(\tilde{s}(t_i)|\heartsuit_{P_i}, 0) = ME_j(\tilde{s}(t_j)|\heartsuit_{P_j}, 0) + k$ *units.*
- $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \top$.
- $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \neg\varphi[s]$ *iff* $\mathfrak{M} \not\models_{\mathcal{L}_{\mathsf{GPL}}} \varphi[s]$.
- $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \varphi \wedge \psi[s]$ *iff* $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \varphi[s]$ *and* $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \psi[s]$.
- $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \forall x \varphi[s]$ *iff for any* $\mathfrak{d} \in \mathcal{I}$, $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \varphi[s(x|\mathfrak{d})]$,
  where $s(x|\mathfrak{d})$ *is the function that is exactly like* s *except for one thing: for the individual variable* x*, it assigns the individual* $\mathfrak{d}$*. This can be expressed as follows:*

$$s(x|\mathfrak{d})(y) := \begin{cases} s(y) & \text{if} \quad y \neq x \\ \mathfrak{d} & \text{if} \quad y = x. \end{cases}$$

If $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \varphi[s]$ *for all* s*, we write* $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \varphi$ *and say that* φ *is true in* $\mathfrak{M}$*. If* φ *is true in all models of* $\mathcal{L}_{\mathsf{GPL}}$*, we write* $\models_{\mathcal{L}_{\mathsf{GPL}}} \varphi$ *and say that* φ *is valid.*

Remark 4 (Sameness Condition). The satisfaction clause of a gradable predicate is the same as the standard one of a non-gradable predicate. So GPL satisfies the Sameness Condition.

2.4 Non First-Order Axiomatisability of Models of $\mathcal{L}_{\mathsf{GPL}}$

The semantic structure of GPL is so *rich* that GPL has the following meta-logical property.

Theorem 5 (Non First-Order Axiomatisability). *The class of models of* $\mathcal{L}_{\mathsf{GPL}}$ *is not first-order axiomatisable.*

Remark 5 (Infinitary Logic). We can express the Archimedean Property by means of infinite quantifier sequences. In order to express them, we need *infinitary logic.*

2.5 Solutions to Cross-Polarity Problem and (In)commensurability Problem

Let us now return to the Cross-Polarity Problem with (1.2)[Mona is happier than Jude is sad.]. Anyone who considers comparatives like (1.2) to be *meaningless* might reason as follows. There is no guarantee that for any $a_2^{\mathfrak{M}}, b_2^{\mathfrak{M}} \in \{Albert, Bernard, Catherine, Jude, Mona, \spadesuit_{sad}, \cdots\}$, there exist $a_1^{\mathfrak{M}}, b_1^{\mathfrak{M}} \in \{Albert, Bernard, Catherine, Jude, Mona, \spadesuit_{happy}, \cdots\}$ such that $(a_1^{\mathfrak{M}}, b_1^{\mathfrak{M}}) \sim (a_2^{\mathfrak{M}}, b_2^{\mathfrak{M}})$. Because at least the condition (4) that the interadjective-comparison ordering relation $\succeq$ for 'happy' and 'sad' should satisfy is not satisfied, we cannot give the truth condition for (1.2). In other words, (1.2) is meaningless. On the other hand, anyone who considers (1.2) to be *meaningful* can evaluate it in the model $\mathfrak{M}$ of Definition 5. Moreover, we can give an answer to the (In)commensurability Problem with (1.6)[My copy of *The Brothers Karamazov* is heavier than my copy of *The Idiot* is old.] in precisely the same way as the Cross-Polarity Problem.

2.6 Truth Conditions of Examples

When a model $\mathfrak{M}$ of $\mathcal{L}_{\mathsf{GPL}}$ is given, the truth conditions of the examples (1.10), (1.11), (1.12), and (1.13) are as follows:

- (1.10) Albert is taller than he is wide.
  $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \mathbf{NER}_{tall,wide}(Albert, Albert)$ iff if $\succeq, ME_1, ME_2, PE_1, PE_2$ and IM_{21} satisfy (2.6), (2.9), (2.10) and (2.13), then $Albert \in \mathcal{I}_{tall}$ and $Albert \in \mathcal{I}_{wide}$ and $ME_1(Albert|\heartsuit_{tall}, 0) > ME_2(Albert|\heartsuit_{wide}, 0)$.
- (1.11) Albert is handsomer than Catherine is intelligent.
  $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \mathbf{LER}_{tall,intelligent}(Albert, Catherine)$ iff $Albert \in \mathcal{I}_{tall}$ and $Catherine \in \mathcal{I}_{intelligent}$ and $(Albert, \spadesuit_{tall}) \succ (Catherine, \spadesuit_{intelligent})$.
- (1.12) Catherine is twice as tall as Doris is wide.
  $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \mathbf{TER}^2_{tall,wide}(Catherine, Doris)$ iff if $\succeq, ME_1, ME_2, PE_1, PE_2$ and IM_{21} satisfy (2.6), (2.9), (2.10) and (2.13), then $Catherine \in \mathcal{I}_{tall}$ and $Doris \in \mathcal{I}_{wide}$ and $ME_1(Catherine|\heartsuit_{tall}, 0) = 2 \cdot ME_2(Doris|\heartsuit_{wide}, 0)$.
- (1.13) Albert is 10 centimetres taller than Bernard is wide.
  $\mathfrak{M} \models_{\mathcal{L}_{\mathsf{GPL}}} \mathbf{UER}^{10}_{tall,wide}(Albert, Bernard)$ iff if $\succeq, ME_1, ME_2, PE_1, PE_2$ and IM_{21} satisfy (2.6), (2.9), (2.10) and (2.13), then $Albert \in \mathcal{I}_{tall}$ and $Bernard \in \mathcal{I}_{wide}$ and $ME_1(Albert|\heartsuit_{tall}, 0) = ME_2(Bernard|\heartsuit_{wide}, 0) + 10\ centimetres$.

3 Concluding Remarks

In this paper, we have proposed a new version of logic for gradable adjectives—GPL. The model of $\mathcal{L}_{\mathsf{GPL}}$ is based on the Vague Predicate Analysis, can give an answer both to the Cross-Polarity Problem and to the (In)commensurability Problem, and can give the truth conditions of such sentences as (1.10), (1.11), (1.12), and (1.13).

This paper is only a part of a larger measurement-theoretic study. We are now trying to construct such logics as

1. dynamic epistemic preference logic [13],
2. dyadic deontic logic [12],
3. vague predicate logic [15,16],
4. threshold utility maximiser's preference logic [14], and
5. a logic for better questions and answers

by means of measurement theory.

Acknowledgements. This work was supported by Grant-in-Aid for Scientific Research (C) (23520015). We would like to thank an anonymous reviewer of LENLS 8 for his helpful comments.

References

1. Bale, A.C.: A universal scale of comparison. Linguistics and Philosophy 31, 1–55 (2008)
2. Hölder, O.: Die Axiome der Quantität und die Lehre von Mass. Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig. Mathematisch-Physikaliche Classe 53, 1–64 (1901)
3. Kennedy, C.: Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison. Routledge, New York (1999)
4. Klein, E.: A semantics for positive and comparative adjectives. Linguistics and Philosophy 4, 1–45 (1980)
5. Krantz, D.H.: A theory of magnitude estimation and cross-modality matching. Journal of Mathematical Psychology 9, 168–199 (1972)
6. Krantz, D.H., et al.: Foundations of Measurement, vol. 1. Academic Press, New York (1971)
7. Luce, R.D., et al.: Foundations of Measurement, vol. 3. Academic Press, San Diego (1990)
8. Roberts, F.S.: Measurement Theory. Addison-Wesley, Reading (1979)
9. van Rooij, R.: Measurement and interadjective comparisons. Journal of Semantics 28, 335–358 (2011)
10. Stevens, S.S.: On the psychophysical law. Psychological Review 64, 153–181 (1957)
11. Suppes, P., et al.: Foundations of Measurement, vol. 2. Academic Press, San Diego (1989)
12. Suzuki, S.: Measurement-Theoretic Foundation of Preference-Based Dyadic Deontic Logic. In: He, X., Horty, J., Pacuit, E. (eds.) LORI 2009. LNCS (LNAI), vol. 5834, pp. 278–291. Springer, Heidelberg (2009)
13. Suzuki, S.: Prolegomena to Dynamic Epistemic Preference Logic. In: Hattori, H., Kawamura, T., Idé, T., Yokoo, M., Murakami, Y. (eds.) JSAI 2008. LNCS (LNAI), vol. 5447, pp. 177–192. Springer, Heidelberg (2009)
14. Suzuki, S.: A measurement-theoretic foundation of threshold utility maximiser's preference logic. Journal of Applied Ethics and Philosophy 3, 17–25 (2011)
15. Suzuki, S.: Measurement-Theoretic Foundations of Probabilistic Model of JND-Based Vague Predicate Logic. In: van Ditmarsch, H., Lang, J., Ju, S. (eds.) LORI 2011. LNCS (LNAI), vol. 6953, pp. 272–285. Springer, Heidelberg (2011)
16. Suzuki, S.: Prolegomena to Salient-Similarity-Based Vague Predicate Logic. In: Onoda, T., Bekki, D., McCready, E. (eds.) JSAI-isAI 2010. LNCS (LNAI), vol. 6797, pp. 75–89. Springer, Heidelberg (2011)
17. Suzuki, S.: Measurement-theoretic foundations of interadjective-comparison logic. In: Aguilar-Guevara, A., et al. (eds.) Proceedings of Sinn und Bedeutung 16. MIT Working Papers in Linguistics, Cambridge, Mass. (to appear)

Towards a Self-selective and Self-healing Evaluation

Alastair Butler[1,2] and Kei Yoshimoto[2]

[1] Japan Science and Technology Agency, PRESTO
ajb129@hotmail.com
[2] Center for the Advancement of Higher Education, Tohoku University
kei@compling.jp

Abstract. Assume a recursive routine for evaluating expressions against an assignment function that stores accumulated binding information for variable names. This paper proposes adding an `If` operation that allows for what is evaluated to be automatically selected during the runtime of evaluation based on the state of the assignment function. This can (a) allow a single encoding of content that would otherwise require distinct expressions, and (b) equip an expression with a way to recover from situations that would cause unwelcome results from evaluation. The new operation is demonstrated to be an essential component for allowing a robust interpretation of unknown lexical items and for feeding an automated regulation of binding dependencies determined on a grammatical basis.

Keywords: semantic evaluation, assignment function, binding dependencies, predicate logic, robust interpretation, grammatical roles.

1 Introduction

Suppose we have a formal language and an evaluation routine $(g, e)^\circ$ for expressions e of the formal language parameterised against assignment function g that stores accumulated binding information for variable names. The classical interpretation of predicate logic is one example of such an evaluation routine; Tarski and Vaught (1956). Let us add to the language a primitive constructor `If` with the following rule of evaluation:

- $(g, \texttt{If }(f, e_1, e_2))^\circ :=$ if $f(g)$ then $(g, e_1)^\circ$ else $(g, e_2)^\circ$

where f is a function taking an assignment and returning a boolean, while e_1 and e_2 are expressions of the formal language. If $f(g)$ returns `true` then e_1 is evaluated and e_2 ignored. If `false` is returned by $f(g)$ evaluation proceeds with e_2 and it is e_1 that is ignored.

An operation like `If` can only find application when the assignment function has structure. Typically assignment functions with the least amount of structure possible are favoured (to maintain minimal assumptions; particularly as

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 96–109, 2012.

assignment functions are essentially a representational layer). For example the assignment function for classical interpretations of predicate logic will assign to all variables a (possibly different) single value. Beyond assigned values being specific values there is nothing to test.

In a series of papers Vermeulen (1993, 2000) and Hollenberg and Vermeulen (1996) propose an altogether richer assignment function in which (possibly empty) sequences of values are assigned to variables. Such an assignment is utilised in the systems of, for example, Visser and Vermeulen (1996), van Eijck (2001), Dekker (2002) and Butler (2007). With a sequence assignment, in addition to assigned values being specific sequences, it is possible to test for sequence length. For example (1) and (2) demonstrate how selecting one of two expressions can come from a test for whether the sequence assigned to the variable name `"arg0"` has length greater than or equal to 1.

(1) $$(\left[\, \texttt{"arg0"} \to \texttt{[}d\texttt{]} \,\right], \texttt{ If } ((\lambda g.|g(\texttt{"arg0"})| \geq 1), e_1, e_2))^\circ = (\left[\, \texttt{"arg0"} \to \texttt{[}d\texttt{]} \,\right], e_1)^\circ$$

(2) $$(\left[\, \texttt{"arg0"} \to \texttt{[]} \,\right], \texttt{ If } ((\lambda g.|g(\texttt{"arg0"})| \geq 1), e_1, e_2))^\circ = (\left[\, \texttt{"arg0"} \to \texttt{[]} \,\right], e_2)^\circ$$

The goal of this paper is to cash out how useful `If` is for a language evaluated against a sequence assignment with the aim of simulating natural language. Specifically we show how a robust interpretation of unknown lexical items becomes possible, as well as enabling an automated feeding of regulation for binding dependencies determined on a grammatical basis.

The remainder of the paper is organised as follows. Section 2 demonstrates a way to encode verbs; section 3 considers nouns; section 4 adds non core arguments for nouns and verbs; section 5 treats relative clauses; section 6 captures passive verbs; and section 7 extends the analysis to embedding taking verbs. Section 8 offers a summary.

2 Encoding Verbs

We start by introducing constructors for a language that is to be evaluated against a sequence assignment and a first-order model (with domain D and interpretation function I): `T` for the construction of terms, `Pred` to form predicates, and `SOME` to bring about existential quantification. $\hat{e}$ is notation for a sequence of expressions; $\hat{e}_i$ is the i-th element of $\hat{e}$; `hd` is an operation returning the head (frontmost) element of a sequence; `map` applies a function to each element in a sequence; and $g[x/d]$ is notation for sequence assignment g only with the sequence value assigned to x extended to include d as the head element.

- $(g,\ \mathtt{T}\ x)^\circ := (\mathtt{hd}\ g(x))@x$
- $(g,\ \mathtt{Pred}\ (s,\hat{e}))^\circ := (s,\ \mathtt{map}\ (\lambda e.(g,\ e)^\circ)\ \hat{e})^\bullet$
- $(g,\ \mathtt{SOME}\ x\ e)^\circ := \exists d \in D : (g[x/d],\ e)^\circ$

Term evaluation ends with a role value construct (following Nakashima, Noda and Handa 1996) consisting of (i) the head value (domain entity) of the sequence assigned to the variable name of the term, (ii) constructor '@', and (iii) the variable name of the term to indicate grammatical role.

Predicate evaluation calls $(s, \hat{e})^\bullet$ such that when an argument with an `"event"` role is present in $\hat{e}$, evaluation completes with a neo-Davidsonian analysis (cf. Davidson 1967, Parsons 1990):

- $(s, x_1@\mathtt{a}_1\ ...\ e@\texttt{"event"}\ ...\ x_n@\mathtt{a}_n)^\bullet :=$
  $e \in I(s)$ and $I(\mathtt{a}_1(e)) = x_1$ ... and $I(\mathtt{a}_n(e)) = x_n$

When there is no argument with an `"event"` role, but there is an `"h"` role (as will be the case with nouns in section 3), evaluation completes as follows:

- $(s, x_1@\mathtt{a}_1\ ...\ y@\texttt{"h"}\ ...\ x_n@\mathtt{a}_n)^\bullet :=$
  $y \in I(s)$ and $(y, x_1) \in I(\mathtt{a}_1)$ ... and $(y, x_n) \in I(\mathtt{a}_n)$

2.1 Examples

Let us assume that an `"event"` binding is created whenever there is a clause, to be accomplished with starting each formal encoding of a sentence with the following `S` operation:

(3) `S = λe.SOME "event" e`

We can now consider applying the above language together with `If` to simulate how differing presences of noun phrases influence processing the pseudoverb *wuggs* in (4) (cf. Berko 1958). If (4a) is a well formed sentence then *wuggs* ought to be a transitive verb; for (4b), an intransitive verb; and for (4c), a verb taking no bound arguments (e.g, *rains*).

(4) a. Someone wuggs someone.
b. Someone wuggs.
c. It wuggs.

To capture (4) it is sufficient to assume a single formal encoding for *wuggs* as follows:

(5)
```
wuggs1 =
  If(λg.|g("arg0")| ≥ 1,
    If(λg.|g("arg1")| ≥ 1,
      Pred ("wuggs", [T "event", T "arg0", T "arg1"]),
      Pred ("wuggs", [T "event", T "arg0"])),
    Pred ("wuggs", [T "event"]))
```

The subexpressions of (5) comprise three different forms for predicate `"wuggs"`, all of which have an `"event"` bound argument while varying as to whether they have `"arg0"` and `"arg1"` bound arguments.

With the further assumptions that object *someone* provides the binding operation `SOME "arg1"` with scope over the verb and subject *someone* provides `SOME "arg0"` with scope over any object and the verb, we can expect to code the sentences of (4) as in (6).

(6) a. `S (SOME "arg0" (SOME "arg1" wuggs1))`
b. `S (SOME "arg0" wuggs1)`
c. `S wuggs1`

What occurs during evaluations of (6a–c) can be seen from (7)–(9), respectively, which depict states of the assignment reached after the evaluation of all noun phrase arguments.

(7)
$$\left(\begin{bmatrix} \texttt{"event"} \rightarrow \texttt{[}d_1\texttt{]} \\ \texttt{"arg0"} \rightarrow \texttt{[}d_2\texttt{]} \\ \texttt{"arg1"} \rightarrow \texttt{[}d_3\texttt{]} \end{bmatrix}, (13)\right)^\circ =$$
$$d_1 \in I(\texttt{"wuggs"}) \text{ and } I(\texttt{"arg0"}(d_1)) = d_2 \text{ and } I(\texttt{"arg1"}(d_1)) = d_3$$

(8)
$$\left(\begin{bmatrix} \texttt{"event"} \rightarrow \texttt{[}d_1\texttt{]} \\ \texttt{"arg0"} \rightarrow \texttt{[}d_2\texttt{]} \\ \texttt{"arg1"} \rightarrow \texttt{[]} \end{bmatrix}, (5)\right)^\circ = d_1 \in I(\texttt{"wuggs"}) \text{ and } I(\texttt{"arg0"}(d_1)) = d_2$$

(9)
$$\left(\begin{bmatrix} \texttt{"event"} \rightarrow \texttt{[}d_1\texttt{]} \\ \texttt{"arg0"} \rightarrow \texttt{[]} \\ \texttt{"arg1"} \rightarrow \texttt{[]} \end{bmatrix}, (5)\right)^\circ = d_1 \in I(\texttt{"wuggs"})$$

In (7) the state of the assignment selects an encoding with `"arg0"` and `"arg1"` bound arguments. In (8) the selected encoding has an `"arg0"` bound argument but no `"arg1"` bound argument. In (9) an encoding with only an `"event"` bound argument is selected.

However using (5) when processing the imperative (10) as encoded in (11) encounters problems since evaluation falls back to an encoding with no `"arg1"` bound argument, as (12) shows, despite a non-empty sequence value being assigned to `"arg1"`.

(10) Wugg someone!

(11) `S (SOME "arg1" wuggs1)`

(12)
$$\left(\begin{bmatrix} \texttt{"event"} \rightarrow \texttt{[}d_1\texttt{]} \\ \texttt{"arg0"} \rightarrow \texttt{[]} \\ \texttt{"arg1"} \rightarrow \texttt{[}d_2\texttt{]} \end{bmatrix}, (5)\right)^\circ = d_1 \in I(\texttt{"wuggs"})$$

The problem seen with (12) is alleviated by the encoding of *wuggs* in (13), such that if `"arg1"` is assigned a non-empty sequence when `"arg0"` is assigned the empty sequence, then an `"arg0"` binding is created.

(13)
```
wuggs2 =
    If($\lambda g.|g("arg0")| \geq 1$,
      If($\lambda g.|g("arg1")| \geq 1$,
        Pred ("wuggs", [T "event", T "arg0", T "arg1"]),
        Pred ("wuggs", [T "event", T "arg0"])),
      If($\lambda g.|g("arg1")| \geq 1$,
        SOME "arg0" (
          Pred ("wuggs", [T "event", T "arg0", T "arg1"])),
        Pred ("wuggs", [T "event"])))
```

Example (10) can now be coded as (14).

(14)
```
S (SOME "arg1" wuggs2)
```

We can see what occurs during an evaluation of (14) by focusing on what happens after evaluating `SOME "arg1"`, as (15) illustrates. The result simulates the effect found with (10) of a subject being at least implicitly present when a verb takes an object.

(15)
$$\left(\begin{bmatrix} \texttt{"event"} \to [d_1] \\ \texttt{"arg0"} \to [\,] \\ \texttt{"arg1"} \to [d_2] \end{bmatrix}, \ (13)\right)^\circ =$$

$$\exists d_3 \in D : d_1 \in I(\texttt{"wuggs"}) \text{ and } I(\texttt{"arg0"}(d_1)) = d_3 \text{ and } I(\texttt{"arg1"}(d_1)) = d_2$$

3 Encoding Nouns

Before we consider formal encodings for nouns we must first create the possibility of noun phrases with restrictions as environments able to support the presence of nouns. To do this we first introduce sequence relations `pop` and `shift(`*op*`)`. For `shift(`*op*`)`, *op* must be specified, with `cons` and `snoc` as suitable candidates to give `shift(cons)` and `shift(snoc)`.

$\texttt{cons}\ (y, [x_0, \ldots, x_{n-1}]) = [y, x_0, \ldots, x_{n-1}]$.

$\texttt{snoc}\ (y, [x_0, \ldots, x_{n-1}]) = [x_0, \ldots, x_{n-1}, y]$.

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

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

We now define language operations `Lam` and `Garb`:

- $(g, \texttt{Lam}\ (x, y, e))^\circ := \exists h : (g, h) \in \texttt{shift(cons)}_{x,y} : (h, e)^\circ$

- $(g, \texttt{Garb (}n\texttt{,}\hat{x}\texttt{,}y\texttt{,}e\texttt{))}^\circ :=$
  $\exists h_0 ... h_{|\hat{x}|} : h_0 = g$ and
  for $0 \leq i < |\hat{x}|, (h_i, h_{i+1}) \in \texttt{shift(snoc)}^{|h_i(\hat{x}_i)|-n}_{\hat{x}_i,y} : (h_{|\hat{x}|}, e)^\circ$

`Lam` $(x,\ y,\ e)$ changes the assignment with $\texttt{shift(cons)}_{x,y}$ and returns the evaluation of e against the new assignment state. For example:

(16) $$(\left[\ \texttt{"e"} \rightarrow \texttt{[}d_1\texttt{]}\ \right]\texttt{, Lam ("e", "arg0", }e\texttt{))}^\circ =$$
$$(\left[\ \texttt{"arg0"} \rightarrow \texttt{[}d_1\texttt{]}\ \right]\texttt{, }e)^\circ$$

`Garb` $(n,\ \hat{x},\ y,\ e)$ modifies the assignment with `shift(snoc)` shifting with `snoc` potentially multiple values from the sequences assigned to the names of $\hat{x}$ into the sequence that is assigned to y, so exactly the frontmost n bindings remain assigned to each name of $\hat{x}$. The evaluation of e against the new assignment state is returned. For example:

(17) $$(\left[\ \texttt{"h"} \rightarrow \texttt{[}d_3\texttt{,}d_2\texttt{,}d_1\texttt{]}\ \right]\texttt{, Garb (1, ["h"], "c", }e\texttt{))}^\circ =$$
$$(\left[\begin{array}{l} \texttt{"h"} \rightarrow \texttt{[}d_3\texttt{]} \\ \texttt{"c"} \rightarrow \texttt{[}d_2\texttt{,}d_1\texttt{]} \end{array}\right]\texttt{, }e)^\circ$$

This gives an operation of ‘unbinding’ like in Berkling (1976) and still more like the ‘end-of-scope’ operator in Hendriks and van Oostrom (2003) since `Garb` is not limited to the terminal level. However a notable difference is that binding values are not completely destroyed but rather entered as values into the sequence assigned to a privileged variable name (e.g. `"c"` in (17)).

`Lam` and `Garb` form the basis for operation `rest`, (18), enabling noun phrases with noun restrictions. The idea implemented by (18) is that noun phrases should insulate the restriction from the containing clause by shifting all open local bindings to a given binding name (here, `"garbage"`) with the exception of the binding that it is the purpose of the noun phrase to introduce which must shift to an `"h"` binding to make available the required binding for nouns. Insulation is accomplished with three parameters: `x` for the binding name that the noun phrase opens in the containing clause, `e` to provide the content of the noun phrase restriction, and `l` to specify a sequence of variable names.

(18)
```
rest = λx.λe.λl.
 Lam (x, "h",
   Garb (0, diff (l, ["h", "garbage"]), "garbage",
     Garb (1, ["h"], "garbage", e)))
```

An application of `rest` (i) turns the `x` binding into an `"h"` binding, (ii) shifts all other open bindings of the names in `l` (minus `"h"` and `"garbage"`) to `"garbage"` bindings, and (iii) shifts all values of the sequence assigned to `"h"` to `"garbage"` with the exception of the frontmost value. For example:

(19) $(\begin{bmatrix} \texttt{"arg0"} \to \texttt{[}d_1\texttt{]} \\ \texttt{"arg1"} \to \texttt{[}d_2\texttt{]} \end{bmatrix}, \texttt{rest "arg1"}\ e\ \texttt{["h", "arg0", "arg1", "garbage"]})^\circ =$

$$(\begin{bmatrix} \texttt{"h"} \to \texttt{[}d_2\texttt{]} \\ \texttt{"garbage"} \to \texttt{[}d_1\texttt{]} \end{bmatrix},\ e)^\circ$$

We add two more operations to our language, `AND` to enable the coordination of expressions, and `W` as an operation to feed e (an expression with an open parameter for taking a sequence) a sequence value (`variables_of` g) comprising all the variable names that are assigned values (empty sequences or otherwise) by the assignment g.

- $(g,\ e_1\ \texttt{AND}\ e_2)^\circ := (g,\ e_1)^\circ$ and $(g,\ e_2)^\circ$
- $(g,\ \texttt{W}\ e)^\circ := (g,\ e\ (\texttt{variables\_of}\ g))^\circ$

We also need determiner `A` which takes as parameters `e` that serves as the content of the noun phrase restriction, `x` the name under which the noun phrase binds in the containing clause, and `f` the part of the containing clause over which the noun phrase takes scope.

(20) `A = λe.λx.λf.SOME x ((W (rest x e)) AND f)`

Finally we are able to consider encoding nouns, the simplest form of which is to be a predicate with a bound `"h"` argument, as with (21).

(21) `wugg1 = Pred ("wugg", [T "h"])`

The ingredients introduced in this section are brought together with the formal analysis of (22) in (23).

(22) A wugg wuggs a wugg.

(23) `S (A wugg1 "arg0" (A wugg1 "arg1" (wuggs2)))`

An evaluation of (23) produces the result of (24).

(24) $\exists d_1 \in D($
$\exists d_2 \in D($
$d_2 \in I(\texttt{"wugg"})$ and
$\exists d_3 \in D($
$d_3 \in I(\texttt{"wugg"})$ and
$d_1 \in I(\texttt{"wuggs"})$ and $I(\texttt{"arg0"}(d_1)) = d_2$ and $I(\texttt{"arg1"}(d_1)) = d_3)))$

The analysis of (23) discriminates overtly between nouns (`wugg1`) and verbs (`wuggs2`), but we could also use the general encoding of (25) which tests for the presence of an `"event"` binding such that with success a verb encoding is selected while failure selects a noun encoding. This allows for the analysis of (26) to produce the result of (24).

(25) $\texttt{general\_wugg = If(}\lambda g.|g(\texttt{"event"})| \geq 1\texttt{, wuggs2, wugg1)}$

(26) `S (A general_wugg "arg0" (A general_wugg "arg1" general_wugg))`

4 Adding Non Core Arguments

So far we have only considered examples where arguments have the privileged grammatical status of being either subjects or objects, with subjects creating `"arg0"` bindings and objects `"arg1"` bindings. To cover such limited data we offered encodings that hard wired acceptable combinations of `"arg0"` and `"arg1"` bindings. This is obviously inadequate as soon as we consider the presence of other types of noun phrase arguments, as are created with preposition phrases in English, such as *as*, *following*, *including*, *on*, but also more productively *according to*, *compared with*, *out into*, *primarily because of*, and so on. In this section we incorporate such binding names.

We start by introducing recursive `add_args`, (27). This takes three parameters: `l1` and `l2` as sequences of binding names and `predicate` that will itself have an open parameter taking a sequence of binding names. If `l1` is `nil` then the content of `l2` is applied to `predicate`, else an expression is created with `If` that has (a) a test for a single `(hd l1)` binding, (b) an expression to evaluate if the test succeeds involving `l2` extended with `(hd l1)` and a recursive call to `add_args` on the tail of `l1`, and (c) an expression to evaluate if the test fails involving a recursive call to `add_args` on the tail of `l1` but no extension to `l2`.

(27)
```
rec add_args = λl1.λl2.λpredicate.
 if l1 = nil then predicate l2 else
   If (λg.|g(hd l1)| ≥ 1,
     add_args (tl l1) ([hd l1]^l2) predicate,
     add_args (tl l1) l2 predicate)
```

To create encodings for nouns `nn`, (28), calls `add_args` as a wrapper around `Pred` that takes at the very least a `T "h"` argument, and possibly others built from names supplied to an `l` parameter. The call of `add_args` adds arguments to `Pred` with names taken from the call of `W` minus the `"h"` name such that they will only have consequences for an evaluation when sufficient binding support is present from the assignment.

(28)
```
nn = λs.
 W (λl.add_args (diff (l, ["h"])) nil (
   λl.Pred (s, map (λx.T x) (l^["h"]))))
```

We can now create a noun encoding for *wugg* as in (29).

(29) `wugg = nn "wugg"`

We can also create encodings for verbs with `verb_base`, (30). This is similar to `nn`, except the ever present argument is `"event"` rather than `"h"`. Also there is an extra parameter `args` for taking a sequence of variable names that must be arguments of the verb.

(30)
```
verb_base = λs.λargs.
 W (λl.
   add_args (diff (l, args^["event"])) nil (
     λl.Pred (s, map (λx.T x) (l^args^["event"]))))
```

Since the range of arguments required by the pseudoverb *wuggs* will be unknown the value passed to the `args` parameter of `verb_base` will be the empty sequence, `nil`, as in (31).

(31) `wuggs3 = verb_base "wuggs" nil`

We are now able to consider examples like (32) which can be formally coded as in (33).

(32) To a wugg a wugg near a wugg wuggs with a wugg.

(33)
```
A wugg "to" (S (A (A wugg "near" wugg) "arg0" (
  A wugg "with" wuggs3)))
```

A successful evaluation of (33) is illustrated in (34).

(34)
$$
\begin{array}{l}
\exists d_1 \in D(\\
\quad \exists d_2 \in D(\\
\quad d_2 \in I(\texttt{"wugg"}) \text{ and} \\
\quad \exists d_3 \in D(\\
\quad\quad \exists d_4 \in D(\\
\quad\quad d_4 \in I(\texttt{"wugg"}) \text{ and} \\
\quad\quad d_3 \in I(\texttt{"wugg"}) \text{ and } (d_3, d_4) \in I(\texttt{"near"})) \text{ and} \\
\quad\quad \exists d_5 \in D(\\
\quad\quad d_5 \in I(\texttt{wugg}) \text{ and} \\
\quad\quad d_1 \in I(\texttt{"wuggs"}) \text{ and } I(\texttt{"arg0"}(d_1)) = d_3 \text{ and} \\
\quad\quad I(\texttt{"with"}(d_1)) = d_5 \text{ and } I(\texttt{"to"}(d_1)) = d_2))))
\end{array}
$$

5 Adding Relative Clauses

In this section we add the possibility of relative clauses. There are two parts to creating relative clauses: (i) accommodating an embedded clause within a noun phrase restriction, and (ii) linking the binding of the noun phrase internally to the embedded clause.

Part (i) is accomplished with `that` of (35), which conjoins `e1`, to be occupied by the head noun, and `e2` for the relative clause to fall inside the scope of a new `"event"` binding.

(35) `that = λe1.λe2.e1 AND (S e2)`

We will consider accomplishing part (ii) by modifying how verbs are encoded, so that no other syntactic trigger need be present internally to the formal representation of the relative clause that is additional to the verb. In short the idea is that the state of the assignment function provides a verb with sufficient information that it is inside a relative clause because both the `"event"` and `"h"` names will be assigned non-empty sequence values. Moreover we might well expect a verb that is a known lexical item to make available case frame or subcategorisation information about the bindings for arguments that are required to support the interpretation of the verb. For example, a transitive verb (e.g., *meets*) will expect both (subject) `"arg0"` and (object) `"arg1"` bindings, such that if one should be missing then it is the role of the available `"h"` binding to serve as the missing binding.

The above idea is implemented with `cover` of (36), which takes a sequence of names `l`, a function from a name and an expression to an expression, and an expression as parameters. If `l` is the empty sequence (`nil`) then return `a`; else create an `If` branch that (i) recursively reapplies with the tail of `l`, `f` and `a` as arguments, and (ii) applies `f` which takes the head of `l` and a recursive call that applies the tail of `l`, `(λx.λe.SOME x e)` and `a`. The effect is that `f` is applied only once on the first name of `l` that has no binding. For remaining names without bindings `(λx.λe.SOME x e)` is applied to ensure the presence of a binding, much as seen with the deviation of (13) from (12).

(36)
```
rec cover = λl.λf.λa.
  if l = nil then a else
    If (λg.|g(hd l)| ≥ 1,
      cover (tl l) f a,
      f (hd l) (cover (tl l) (λx.λe.SOME x e) a))
```

We now need to modify the encoding for verbs so `cover` is called, such that the `l` parameter of `cover` receives as its value the sequence of names corresponding to the arguments the predicate is known to be expecting, `f` is either an instruction to shift an `"h"` binding or an instruction to create a new binding, and `a` is the remainder of the verb constructed with `verb_base`, (30).

(37)
```
verb = λs.λargs.If (λg.|g("h")| ≥ 1,
  cover args (λx.λe.Lam ("h", x, e)) (verb_base s args),
  cover args (λx.λe.SOME x e) (verb_base s args))
```

We can now encode a known transitive verb like *meets* as in (38). Pseudoverb *wuggs* can be encoded as in (39).

(38) `meets = verb "meets" ["arg0", "arg1"]`

(39) `wuggs = verb "wuggs" nil`

To demonstrate the above encodings we can consider examples like in (40) which are formalised in (41).

(40) a. A wugg that meets a wugg wuggs.
b. A wugg that a wugg meets wuggs.
c. A wugg that a wugg wuggs wuggs.

(41) a. `S (A (that wugg (A wugg "arg1" meets)) "arg0" wuggs)`
b. `S (A (that wugg (A wugg "arg0" meets)) "arg0" wuggs)`
c. `S (A (that wugg (A wugg "arg0" wuggs)) "arg0" wuggs)`

The result of evaluating (41b) is illustrated in (42).

(42) $\exists d_1 \in D($
$\quad \exists d_2 \in D($
$\quad\; d_2 \in I(\texttt{"wugg"})$ and
$\quad\; \exists d_3 \in D($
$\quad\;\; \exists d_4 \in D($
$\quad\;\;\; d_4 \in I(\texttt{"wugg"})$ and
$\quad\;\;\; d_3 \in I(\texttt{"meets"})$ and $I(\texttt{"arg0"}(d_3)) = d_4$ and $I(\texttt{"arg1"}(d_3)) = d_2))$
$\quad\;$ and $d_1 \in I(\texttt{"wuggs"})$ and $I(\texttt{"arg0"}(d_1)) = d_2))$

6 Encoding Passives

To capture passives we introduce the operation `passive`, (43), to take scope over the encoding of a verb. With `Garb` (see section 3) `passive` shifts any `"arg0"` binding to an `"arg1"` binding, and any `"by"` binding to an `"arg0"` binding.

(43)
```
passive =
  λe.Garb (0, ["arg0"], "arg1", Garb (0, ["by"], "arg0", e))
```

That *was wugged* is passive tells us the encoding of the verb must, at the very least, contain requirements for `"arg0"` and `"arg1"` argument bindings, as with (44).

(44)
```
was_wugged = passive (verb "wuggs" ["arg0", "arg1"])
```

We can now consider the examples of (45) which are formally presented in (46).

(45) a. Someone was wugged.
b. Someone was wugged by someone.

(46) a. `S (Some "arg0" was_wugged)`
b. `S (Some "arg0" (Some "by" was_wugged))`

With evaluation both (46a) and (46b) result in (47) but through very different means. With (46a) d_3 is created internally to the evaluation of `verb` with a call to `cover` as a value assigned to `"arg0"`. With (46a) d_3 is created initially as a value assigned to `"by"` and subsequently moved by the call of `passive` to be a value assigned to `"arg0"`.

(47) $\exists d_1 \in D($
$\quad \exists d_2 \in D($
$\quad\; \exists d_3 \in D($
$\quad\;\; d_1 \in I(\texttt{"wuggs"})$ and $I(\texttt{"arg0"}(d_1)) = d_3$ and $I(\texttt{"arg1"}(d_1)) = d_2)))$

7 Adding Embedding Taking Verbs

In this section we consider adding the possibility of embedding taking verbs with the operation `emb_verb`, (49), which calls `emb_verb_base`, (48).

```
(48)  emb_verb_base = λs.λargs.λemb.
       W (λl.
         add_args (diff (l, args^["event", "h"])) nil (
           λl.Pred (s,
             (map (λx.T x) (l^args^["event"]))^[
               (Garb (0, l^args^["event"], "garbage",
                 S emb))@"embedding"])))
```

```
(49) emb_verb = λs.λargs.λemb.
       If (λg.|g("h")| ≥ 1,
        cover args (λx.λe.Lam("h",x,e)) (emb_verb_base s args emb),
        cover args (λx.λe.SOME x e) (emb_verb_base s args emb))
```

Complications over the encoding we already have for verbs with `verb` and `verb_base` come with the presence of an extra parameter for taking an embedded expression. Notably, internally to `emb_verb_base` before the embedded expression is reached, bindings that served to bind arguments of the embedding taking verb are shifted (with `Garb`) so they do not bind further into the embedded clause.

A more subtle difference to note is that the `"h"` name is removed together with the `"event"` name and the other names of `args` from the sequence of names that are input to `add_args`. `"event"` and the names of `args` are of course removed since they are already due to serve as bound arguments. The `"h"` name is also removed so that it will not act as an argument and so will not be collected by `Garb`, with the consequence that any open `"h"` binding will survive through to the embedding. However note that a `"h"` binding might get caught as a binding of the embedding predicate with the call to `cover` triggered with `emb_verb` that feeds off information for argument names stated to be required by the case frame information of a given verb encoding.

Applications of `emb_verb` are given in (50) and (51) with *know* and the pseudoverb *wuggs*, respectively. Presence of a clausal complement for a given instance of *wuggs* informs that the pseudoverb should be taken to be an embedding taking verb.

```
(50) knows = emb_verb "knows" ["arg0"]
```

```
(51) wuggs_emb = emb_verb "wuggs" nil
```

We can now provide the examples of (52) with the formal codings of (53).

(52) a. A wugg that knows a wugg wuggs wuggs.
b. A wugg that a wugg knows wuggs wuggs.
c. A wugg that it wuggs wuggs wuggs.

(53) a. `S (A (that wugg (knows (A wugg "arg0" wuggs))) "arg0" wuggs)`
b. `S (A (that wugg (A wugg "arg0" (knows wuggs))) "arg0" wuggs)`
c. `S (A (that wugg (wuggs_emb wuggs)) "arg0" wuggs)`

A partial evaluation of (53b) is illustrated in (54), sufficient to demonstrate how the binding dependencies are established.

(54)
$$
\begin{array}{l}
\exists d_1 \in D(\\
\quad \exists d_2 \in D(\\
\quad d_2 \in I(\texttt{"wugg"}) \text{ and} \\
\quad \exists d_3 \in D(\\
\quad\quad \exists d_4 \in D(\\
\quad\quad\quad d_4 \in I(\texttt{"wugg"}) \text{ and} \\
\quad\quad\quad (\texttt{"knows", [}d_3\texttt{@"event", }d_4\texttt{@"arg0",} \\
\quad\quad\quad\quad \exists d_5 \in D(\\
\quad\quad\quad\quad\quad d_5 \in I(\texttt{"wuggs"}) \text{ and } I(\texttt{"h"}(d_2)) = d_2)\texttt{@"embedding"])}^{\bullet})) \text{ and} \\
\quad d_1 \in I(\texttt{"wuggs"}) \text{ and } I(\texttt{"arg0"}(d_1)) = d_2))
\end{array}
$$

8 Summary

This paper introduced operator `If` to allow for content of an expression to be selected at the runtime of evaluation based on the state of the assignment function. We have seen examples of how this can (a) allow a single encoding of content that would otherwise require distinct expressions, and (b) equip expressions with ways to recover from situations that would otherwise lead to unwelcome results from evaluation. With these two properties evaluation was left to feed automatic regulation to enable a coverage of unknown lexical items as well as novel binding names, while also providing the means to get away with very little explicit coding of information about how binding dependencies should be established. Essentially explicit code was limited to the name of a binding upon creation. Thereafter evaluation itself shouldered the burden for how dependencies were established.

Acknowledgement. This research has been supported by the JST PRESTO program (Synthesis of Knowledge for Information Oriented Society).

References

Berkling, K.J.: A symmetric complement to the lambda-calculus. Interner Bericht ISF-76-7, GMD, St. Augustin, Germany (1976)

Berko, J.: The child's learning of English morphology. Word 14, 150–177 (1958)

Butler, A.: Scope control and grammatical dependencies. Journal of Logic, Language and Information 16, 241–264 (2007)

Davidson, D.: The logical form of action sentences. In: Rescher, N. (ed.) The Logic of Decision and Action. University of Pittsburgh Press, Pittsburgh (1967); Reprinted in: Davidson, D.: Essays on Actions and Events, pp. 105–122. Claredon Press, Oxford (1980)

Dekker, P.: Meaning and use of indefinite expressions. Journal of Logic, Language and Information 11, 141–194 (2002)

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

Hendriks, D., van Oostrom, V.: Adbmal-calculus. Department of Philosophy. Utrecht University (2003)

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

Nakashima, H., Noda, I., Handa, K.: Organic programming language GAEA for multi-agents. In: Proceedings of the Second International Conference on Multiagent Systems (ICMAS 1996). AAAI (1996)

Parsons, T.: Events in the Semantics of English. MIT Press, Cambridge (1990)

Tarski, A., Vaught, R.L.: Arithmetical extensions of relational systems. Compositio Mathematica 13, 81–102 (1956)

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

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

Visser, A., Vermeulen, C.F.M.: Dynamic bracketing and discourse representation. Notre Dame Journal of Formal Logic 37, 321–365 (1996)

The Fifth International Workshop on Juris-Informatics (JURISIN 2011)

Shozo Ota[1], Ken Satoh[2], and Makoto Nakamura[3]

[1] The University of Tokyo, Japan
[2] National Institute of Informatics and Sokendai, Japan
[3] Nagoya University, Japan

The Fifth International Workshop on Juris-Informatics (JURISIN 2011) was held with a support of the Japanese Society for Artificial Intelligence (JSAI) in association with Third JSAI International Symposia on AI (JSAI-isAI 2011). Although JURISIN was organized to discuss legal issues from the perspective of informatics, the scope of JURISIN is more wide-ranging than that of the conventional AI and law. In our 'call for papers' the following topics are mentioned: legal reasoning, argumentation/argumentation agent, legal term ontology, formal legal knowledge-base, intelligent management of legal knowledge-base, translation of legal documents, computer-aided law education, use of informatics and AI in law, legal issues on ubiquitous computing/multi-agent system/the Internet, social implications of use of informatics and AI in law, and so on.

Thus, the members of Program Committee (PC) are leading researchers in various fields: Thomas Ågotnes (University of Bergen, Norway), Katie Atkinson (The University of Liverpool, UK), Phan Minh Dung (AIT, Thailand), Tom Gordon (Franfoher FOKUS, Germany), Guido Governatori (The University of Queensland, Australia), Tokuyasu Kakuta (Nagoya University, Japan), Masahiro Kozuka (Kanazawa University, Japan), Makoto Nakamura (Nagoya University, Japan), Katsumi Nitta (Tokyo Institute of Technology, Japan), Paulo Novais (University of Minho, Portugal), Shozo Ota (The University of Tokyo, Japan), Jeremy Pitt (Imperial College, UK), Henry Prakken (University of Utrecht & Groningen, The Netherlands), Seiichiro Sakurai (Meiji Gakuin University, Japan), Ken Satoh (National Institute of Informatics and Sokendai, Japan), Akira Shimazu (Japan Advanced Institute of Science and Technology, Japan), Fumihiko Takahashi (Meiji Gakuin University, Japan), Satoshi Tojo (Japan Advanced Institute of Science and Technology, Japan), Katsuhiko Toyama (Nagoya University, Japan), Radboud Winkels (The University of Amsterdam, the Netherlands), and John Zeleznikow (Victoria University, Australia). The collaborative work of computer scientists, lawyers and philosophers of other fields is expected to contribute to advances of juris-informatics and it is also expected to open novel research areas.

Despite the terrible earthquake and the nuclear plant accident, eleven papers were submitted. Each paper was reviewed by three members of PC. While one paper was cancelled, eight papers were accepted in total. The collection of papers covers various topics such as legal reasoning, argumentation theory, and natural language processing of legal documents.

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 110–111, 2012.

After the workshop, seven papers including one invited paper were submitted for the post proceedings. They were reviewed by PC members again and five papers were finally selected. Followings are their synopses.

Robert Kowalski and Anthony Burton's paper is an invited paper. They describe a rule-based system, WUENIC, implemented as a logic program, developed by WHO and UNICEF for estimating global, country by country, infant immunization coverage. Their system possesses many of the characteristics of rule-based legislation, facilitating decisions that are consistent, transparent and replicable. Hirokazu Igari et al. propose a novel method analyzing structure for documents with common format like legal judgements. They implement a document structure parser which has high generality and extensibility and achieves high accuracy. Yasuhiro Ogawa et al. propose new evaluation metrics for statutory sentences translated from Japanese to English. Their method is based on recall-oriented metrics and they show that it is more suitable for consistent evaluation than the previous evaluation metric. Takanori Sato et al. propose a deliberation process support system for the Japanese citizen judge system started in 2009. The system assists the presiding judge to facilitate the deliberation by visualization of argument structure, suggesting orders of arguments and giving a chance to each member in the citizen judge meeting. Katsuhiko Sano et al. study the prerequisite-effect structures in legal documents from an intuitionistic logical viewpoint. They introduce a new operator into intuitionistic logic which corresponds with immediate consequence. They also give a complete sequent calculus.

Finally, we wish to express our gratitude to all those who submitted papers, PC members, discussant and attentive audience.

WUENIC – A Case Study in Rule-Based Knowledge Representation and Reasoning

Robert Kowalski[1] and Anthony Burton[2,*]

[1] Imperial College London
rak@doc.ic.ac.uk
[2] World Health Organization, Geneva
burtona@who.int

Abstract. WUENIC is a rule-based system implemented as a logic program, developed by WHO and UNICEF for estimating global, country by country, infant immunization coverage. It possesses many of the characteristics of rule-based legislation, facilitating decisions that are consistent, transparent and replicable. In this paper, we focus on knowledge representation and problem-solving issues, including the use of logical rules versus production rules, backward versus forward reasoning, and rules and exceptions versus argumentation.

Keywords: WUENIC, rules and exceptions, logic programming, argumentation.

1 Introduction

Rule-based law has many characteristics in common with logical reasoning, facilitating decision-making that is consistent, transparent and replicable. In this paper, we describe and discuss a rule-based system for estimating infant immunization coverage for 194 countries and territories.

The system, WUENIC (WHO and UNICEF Estimates of National Immunization Coverage), developed by WHO and UNICEF, is implemented in pure Prolog, and was partly inspired by the representation of the British Nationality Act as a logic program [1]. However, whereas the representation of the British Nationality Act required only the formalisation of an existing body of legal rules, the development of WUENIC involved the additional feature of further developing and refining the rules themselves. Not surprisingly, we found that the process of developing and testing the formalisation greatly assisted the process of rule development and refinement.

During the project, the working group preparing the estimates also explored the use of production rules and argumentation as alternative knowledge representation approaches. In this paper, we compare these approaches with our use of logic

---

[*] The author is a staff member of the World Health Organization. The author alone is responsible for the views expressed in this publication, and they do not necessarily represent the decisions, policy or views of the World Health Organization Department of Immunization, Vaccines and Biologicals.

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 112–125, 2012.

programming in the context of the WUENIC application. We also compare backward and forward reasoning as alternative ways of executing the WUENIC program.

1.1 WUENIC

Since 2000 the World Health Organization (WHO) and the United Nations Children's Fund (UNICEF) have made annual estimates of national infant immunization coverage for selected vaccines [2]. Among other uses, these estimates are used to track progress towards the Millennium Development Goal 4 of reducing child mortality. They are also used as an information resource by the GAVI Alliance (formally known as the Global Alliance for Vaccines and Immunisation) for funding immunisation in developing countries.

The WHO/UNICEF estimates are based on reports from national authorities and are supplemented with results from nationally representative surveys. In addition, local staff, primarily national immunization system managers and WHO/UNICEF regional and national staff, are also consulted for additional information that might influence or bias these two primary sources of empirical data.

Given this data and additional information, the WHO/UNICEF working group uses a set of rules to derive its official estimates, reconciling any inconsistencies in the data. Until we began formalizing the rules in the summer of 2009, the informal rules were imprecise and sometimes applied inconsistently.

In the summer of 2009, Anthony Burton, who had little previous experience of programming in Prolog and had read about the representation of the British Nationality Act as a logic program [1], approached Robert Kowalski to discuss the possibility of applying similar knowledge representation methods to the informal WHO/UNICEF rules. By May 2010, the working group had formalized a major portion of the rules, and Anthony Burton had implemented a Prolog program (called WUENIC). Since then, the program has been used to help produce the annual estimates of immunization coverage for the years 2009 and 2010 [11].

Prolog is a logic programming language with non-logical features that are primarily used for efficiency purposes. WUENIC is implemented in pure Prolog, without any of these non-logical features.

1.2 Logic Programs

Logic programs are a simplified form of logic, in which knowledge is expressed as conditional sentences of the form ***if*** *condition(s)* ***then*** *conclusion,* or equivalently in the form *conclusion* ***if*** *condition(s).* Such conditionals (also called *implications, clauses* or *rules*) combine an atomic *conclusion* with a conjunction of *conditions.*

The very first sentence of the British Nationality Act is expressed in an English style that is close to logic programming form:

> 1.-(1) A person born in the United Kingdom after commencement shall be a British citizen if at the time of the birth his father or mother is -
> (a) a British citizen; or
> (b) settled in the United Kingdom.

One difference between the English sentence and its logic programming form is that the English sentence inserts the logical conditions "born in the United Kingdom after commencement" into the middle of the logical conclusion "a person shall be a British citizen". In logic programming style, the same sentence might be written in the form:

a person X acquires british citizenship by subsection 1.1 at time T
if *X is born in the uk at time T*
and *T is after commencement*
and *Y is a father of X or Y is a mother of X*
and *Y is a british citizen at time T or Y is settled in the uk at time T*

Here *X, Y* and *T* are *universally quantified variables*, meaning that the conditional holds for *all* values of the variables *X, Y* and *T.*

To use the conditional to determine whether a person acquires citizenship by subsection 1.1, it is necessary to provide additional conditionals defining such notions as being *settled in the uk* and to supply additional facts. A *fact* is an atomic sentence without variables and without qualifying conditions, such as *Mary is a person* and *John is the father of Mary*. Mathematically, it is convenient to regard a fact as a conditional in which there are no variables and the number of conditions is zero.

Logic programs are commonly characterized as a *declarative representation*, in which "knowledge" is expressed by declarative sentences. However, they can also be used as a *procedural representation*, to express goal-reduction procedures of the form ***to solve*** *conclusion* ***solve*** *condition(s).* Such goal-reduction procedures are obtained by applying *backward reasoning* to sentences of the form *conclusion* ***if*** *condition(s)*, attempting to show that the *conclusion* holds by showing that the *condition(s)* hold.

For example, backward reasoning applied to the logic programming form of the first sentence of the British Nationality Act turns it into the goal-reduction procedure:

to show *that X acquires british citizenship by subsection 1.1 at time T*
find *T such that X is born in the uk at time T*
and show *T is after commencement*
and find *Y such that Y is a father of X or Y is a mother of X*
and show *Y is a british citizen at time T or Y is settled in the uk at time T*

Such backward reasoning with conditionals gives logic programming the capabilities of a high-level, procedural programming language, making the use of other, more conventional programming languages theoretically unnecessary[1].

Backward reasoning is an *analytical* form of reasoning, which reduces concepts to sub-concepts. It contrasts with forward reasoning, which is s*ynthetic*, and which generates new concepts from existing concepts. We will see later in the paper that the distinction between backward and forward reasoning was an important issue in the development of WUENIC.

---

[1] This is why Prolog programs are written condition-first in the form *conclusion :- conditions*, because this syntax is neutral with respect to the declarative and procedural interpretations.

1.3 Rules and Exceptions

In normal logic programming, the conditions of a clause are a conjunction of atomic predicates and negations of atomic predicates. For example:

> *The Secretary of State may by order deprive a person of a citizenship status by subsection 40.-(2)* ***if*** *the Secretary of State is satisfied that deprivation is conducive to the public good,* ***and*** *he is* ***not*** *forbidden from depriving the person of citizenship status by subsection 40.-(2).*
> *The Secretary of State is forbidden from depriving a person of citizenship status by subsection 40.-(2) if he is satisfied that the order would make the person stateless.*

Here the first sentence represents a general rule, and the second sentence represents an exception to the rule. The negative condition of the first sentence qualifies the general rule, so that the rule does not apply to exceptions. Taken together, the rule and exception intuitively imply:

> *The Secretary of State may by order deprive a person of a citizenship status by subsection 40.-(2)* ***if*** *the Secretary of State is satisfied that deprivation is conducive to the public good,* ***and*** *he is* ***not*** *satisfied that the order would make the person stateless.*

This implication has the abstract form: From *A if B and **not** E* and *E if C* infer *A if B and **not** C.*

which is contrary to the laws of classical logic. In classical logic, it would be necessary to state the rule, exception and inference in the (arguably) less natural form: From *A if B and **not** E* and ***not** E if **not** C* infer *A if B and **not** C.*

The logic programming formulation, in contrast with classical logic, interprets negative conditions of the form ***not** E* as *negation as failure*:

> ***not** E* holds if and only if ***E*** fails to hold.

Hundreds - if not thousands – of research publications have sought to provide semantics for this form of negation. In the WUENIC project, we have found it useful to interpret negation as failure in terms of the argumentation semantics of Dung [4]:

> ***not** E* holds if and only if
> every attempted counter-argument to show that ***E*** holds fails.

Dung's semantics can also be applied directly to rules and exceptions expressed more informally in ordinary English. For example, in the British Nationality Act, the rule and exception are actually expressed in the potentially contradictory form:

> 40.-(2) The Secretary of State may by order deprive a person of a citizenship status if the Secretary of State is satisfied that deprivation is conducive to the public good.
>
> 40.-(4) The Secretary of State may not make an order under subsection (2) if he is satisfied that the order would make a person stateless.

In classical logic, these two sentences have the form *A if B* and ***not*** *A if C,* which is inconsistent if both *B* and *C* hold.

The argumentation semantics avoids the inconsistency by making it possible to assign different priorities to arguments constructed by means of general rules and to arguments constructed by means of exceptions. In the argumentation semantics, an argument supported by a general rule *A if B* can be attacked by a counter-argument supported by an exception ***not*** *A if C,* but an argument supported by the exception cannot be counter-attacked by an argument supported by the general rule.

In WUENIC, we found it easier to represent rules and exceptions using negation as failure with its argumentation semantics, than to use argumentation applied directly to ordinary, informal English. The use of negation as failure made it easy, in turn, to represent rules and exceptions directly in Prolog.

1.4 Argumentation

In Dung's argumentation, whether an argument succeeds or fails depends on whether or not it can defeat all counter-arguments by recruiting support from defending arguments. The following example is from [3]:

Rule 1: All thieves should be punished.
Rule 2: Thieves who are minors should not be punished.
Rule 3: Any thief who is violent should be punished.

Suppose that John is a thief who is a violent minor. Then by rule 1, there is an argument that he should be punished; by rule 2, a counter-argument that he should not be punished; and by rule 3, another argument that he should be punished.

In Dung's theory of argumentation, the second argument attacks the first, and the first and third arguments attack the second. Together, the first and third argument succeed, because for every counter-argument (the second one) that attacks one of them, there is at least one argument (in fact two arguments, the first and third) that counter-attacks and defeats it.

The second argument also succeeds, because for every counter-argument (the first and third) that attacks the second argument, there exists an argument (the second argument itself) that counter-attacks and defeats them both.

However, this analysis in terms of arguments fails to do justice to the intuition that the first rule is a general rule, the second rule is an exception to the general rule, and the third rule is an exception to the exception. Arguments constructed by means of an exception to a rule attack arguments constructed by means of the rule, but not vice versa. So in this case, only the first and third argument, taken together, should succeed, and the second argument should fail. Dung's theory accommodates this intuition, because it allows attack relations to be specified abstractly and arbitrarily.

In argumentation, the attack relations between arguments and the resulting success or failure of arguments is embedded both in the semantics and in the proof theory of the logic of argumentation. It is not represented explicitly in the representation of the arguments themselves. In contrast, in normal logic programming, the attack and defeats relations are represented explicitly by negative conditions in the rules. Here is one such representation, applied to the rules for punishing thieves:

a person should be punished
if *the person is a thief*
and *the person is* ***not*** *an exception to the punishment rule.*

a person is an exception to the punishment rule
if *the person is a minor*
and *the person is* ***no****t an exception to the exception to the punishment rule.*

a person is an exception to the exception to the punishment rule
if *the person is violent.*

This representation has the advantage that additional exceptions can be catered for by adding additional clauses, without changing the rule. For example:

a person is an exception to the punishment rule
if *the person is* ***not*** *of full mental capacity.*

Dung's argumentation semantics of logic programs with negative conditions [5, 6] has practical significance, because it means that explanations of conclusions derived by means of logic programs can be expressed in argumentation terms.

1.5 Production Rules

In Artificial Intelligence, the notion of *rule* has two different interpretations: as a conditional in logic programming, and as an expression of the form ***if*** *conditions* ***then*** *actions* in production systems. Both kinds of rules have a similar ***if-then*** syntactic form. But rules in logic programming have both a declarative, logical semantics and an "operational semantics" in terms of backward and forward reasoning. Production rules, on the other hand, do not have a declarative semantics, and have an "operational semantics" in terms of state transitions determined by means of *actions*.

Production rules of the form ***if*** *conditions* ***then*** *actions* are executed by checking whether the *conditions* hold in the current state, and, if they do, then performing the *actions* to transform the current state into a new one. This execution strategy resembles and is sometimes confused with forward reasoning.

The confusion between logic programming rules and production rules was a factor in our discussions about knowledge representation in WUENIC. It also helped to motivate the first author to investigate the relationship between the two kinds of rules in greater detail [8, 9]. In the case of WUENIC, we decided to focus our efforts on formalising the rules in normal logic programming form, which had the advantage that they could then be translated directly into pure Prolog.

2 WUENIC

The WHO/UNICEF working group had been applying an informal set of rules [2], before starting the project in the summer of 2009. At first, the most natural formalization seemed to be in terms of production rules. For example:

if *for a Country, Vaccine and Year, there are nationally reported data*
and *there is* ***no*** *survey data*
then *the WUENIC estimate* ***is*** *the reported data.*

if *for a Country, Vaccine and Year, there is nationally reported data*
and *there is survey data*
and *the survey data is within 10% of the reported data*
then *the WUENIC estimate* ***is*** *the reported data.*

if *for a Country, Vaccine and Year, there is nationally reported data*
and *there is survey data*
and *the survey data is* ***not*** *within 10% of the reported data*
then *the WUENIC estimate* ***is*** *the survey data.*

It was natural to think of the ***then*** part of the rule as an action. But it is also possible to think of the same rules in more logical terms, where ***is*** in the conclusion is equality, and the *WUENIC estimate*, the *reported data* and the *survey data* are functions that take *Country, Vaccine* and *Year* as input and return percentages as output.

2.1 Functions versus Relations

In logic programming, input-output functions are represented as relations, as in relational databases. For example, instead of representing the mother of a person as the output of a motherhood function:

mother-of(john) = mary.

the function is represented as a relation:

mother(john, mary).

Instead of using equality (or ***is***) to define a new function, as in:

parent(X)= {*Y* | *mother-of(X) = Y* or *father-of(X) = Y*}

a new relation is defined using the logical connective ***if***:

parent(X, Y) ***if*** *mother(X, Y)*
parent(X, Y) ***if*** *father(X, Y)*

In ordinary mathematical logic, definitions are usually represented by means of equivalences (***if and only if***). But in logic programming, definitions are represented by means of the ***if*** halves of equivalences, and the ***only-if*** halves are implicit.

Treating functions as relations in this way, the first production rule above becomes the logic programming rule:

if *the nationally reported data for a Country, Vaccine and Year is R*
and *there is* ***no*** *survey data S for the Country, Vaccine and Any-Year*
then *the WUENIC estimate for the Country, Vaccine and Year is R.*

In Prolog syntax, this is written in the form:

```
wuenic(Country, Vaccine, Year, R) :-
reported(Country, Vaccine, Year, R),
not(survey(Country, Vaccine, Any-Year, S)).
```

where the variables Country, Vaccine, Year, R are universally quantified, but the condition not(survey(Country, Vaccine, Any-Year, S)) means there does not exist Any-Year and S such that survey(Country, Vaccine, Any-Year, S).

Whereas true production systems execute production rules only in the forward direction, from condition to actions, Prolog executes logic programs by backward reasoning, from conclusion to conditions. However, the declarative semantics of logic programs is independent from the manner in which they are executed, and in theory logic programs can be executed by backward or forward reasoning.

2.2 The General Structure of the Rules

After we decided to formalize the WUENIC rules in logic programming form and to implement them in Prolog, we turned our attention to the overall structure of the rules.

The same rules apply to all countries, and are applied to individual countries without reference to other countries. A single run of the rules, for a given country, produces estimates for all years in the period from 1997 to the current year and for all vaccines for which the country is required to report coverage. Estimates are produced for the entire period, because the estimates for previous years might be influenced by new information, such as a new survey, obtained in the current year.

We decided to organize the rules into four levels. The first three levels produce preliminary estimates for the individual vaccines. The fourth level reconciles any discrepancies between vaccines.

Level one. The primary data, consisting of the nationally reported immunisation coverage and any nationally representative surveys, are evaluated separately, and if necessary are ignored or adjusted. For example, the reported data are adjusted downward if they are over 100%, which can happen if the reporting authorities use a number for the target population that is smaller than the actual number of children vaccinated. Survey data may be ignored if the sample size is too small.

Level two. If data are available from only a single source (national reports or surveys) for every year for the given country and vaccine, then the estimates are based on that source alone. Otherwise, the estimates are made at "anchor years", which are years in which there are both reported data and survey data. If the (possibly adjusted) survey data does not support the (possibly adjusted) reported data, then the (possibly adjusted) survey data overrides the reported data, and is used as the WUENIC estimate, unless there is a compelling reason to do otherwise. A compelling reason might be an exception represented by another, overriding rule, or as we will see later by a "fact" representing a working group decision.

Level three. Estimates at non-anchor years, for which there is only reported data, but not survey data, are influenced by the estimates at the nearest anchor years. If the estimates at the nearest anchor years are based on the (possibly adjusted) reported data, then the estimate at the non-anchor year is similarly based on the reported data. But if the estimates at the nearest anchor years are based on the (possibly adjusted) survey data, then the estimate at the non-anchor years is based on the reported data, but calibrated to the level of the estimates in the anchor years.

Level four. The resulting estimates for the different vaccines are cross-checked for consistency between related vaccines, and adjustments are made if necessary. For example, the estimates for the first and third doses of the same vaccine are compared, to ensure that the estimate for the third dose is not higher than the estimate for the first dose. This reflects the fact that every child who receives a third dose of a vaccine must have received a first dose.

2.3 Two Ways of Logically Representing the Problem of Going from *a* to *z*

It is natural to interpret the four levels as four phases, in which the goal of producing the estimates for a given country is reduced to the consecutive subgoals of:

first evaluating and possibly adjusting the data,
then producing preliminary estimates at anchor years,
then producing preliminary estimates at non-anchor years, and
finally making possible adjustments for estimates of related vaccines.

This goal-reduction procedure, in turn, has a natural interpretation as an application of backwards reasoning to a logical conditional:

if *for a Country, the data are evaluated and possibly adjusted*
and *the possibly adjusted data are used to produce*
preliminary estimates at anchor years
and *the preliminary estimates at anchor years are used to produce*
preliminary estimates at non-anchor years
and *the preliminary estimates of related vaccines are*
compared and possibly adjusted
then *the WUENIC estimates for the Country are*
the resulting possibly adjusted estimates

However, this representation clashes with the natural way of representing the individual rules within the different levels. It took us some time to diagnose the problem, and longer to solve it.

The problem, discussed at length in Chapter 4 of [12], is that there are two ways of logically representing the problem of going from *a* to *z*. The first representation is analogous to the treatment of levels as phases:

if *you can go directly from X to Y* ***then*** *you can go from X to Y*
if *you can go directly from X to Y* ***and*** *you can go from Y to Z*
then *you can go from X to Z*

With this representation, given a set of direct connections represented as facts, say, of the form *you can go directly from a to b*, the problem of going from *a* to *z* is represented by the goal of showing:

you can go from a to z

The second representation is analogous to the natural way of representing the individual rules. In this representation, a direct connection, say, from *a* to *b* is represented by a rule:

if *you can go to a* ***then*** *you can go to b*

With this representation, given the set of direct connections represented as rules, the problem of going from *a* to *z* is represented by:

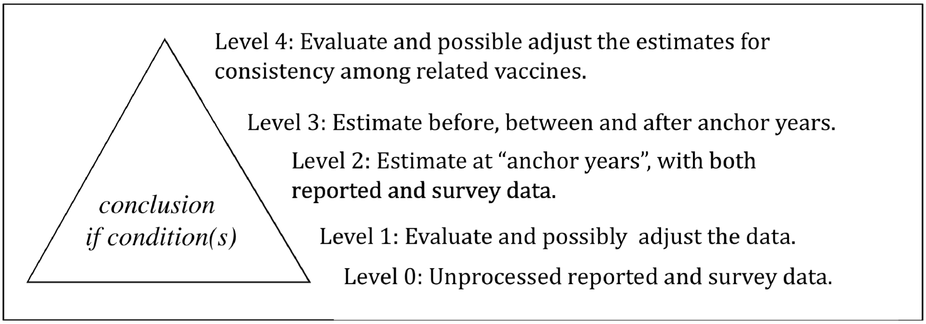
assuming that *you can go to a* ***and showing that*** *you can go to z*

The problem of going from *a* to *z* is typical of a wide class of problems, including the WUENIC problem of generating immunization estimates.

In general, the two representations have both a declarative and a procedural interpretation. But the first representation is lower-level, and easier for the programmer to control. The second representation is higher-level, and easier for non-programmers to understand. We decided to use the second representation, but did not anticipate the control problems that would arise with the Prolog implementation.

2.4 Forward versus Backward Reasoning

With the WUENIC program written in this way, the problem of generating the immunization estimates from the given initially reported and survey data can be viewed as the problem of filling in a triangle:



The conclusions of the rules represent predicates at higher levels of the triangle, and the conditions of the rules represent predicates at lower levels. Forwards reasoning with the rules fills in the triangle from the bottom up, treating the successive levels as phases. Backward reasoning, as in Prolog, fills in the triangle from the top down.

In theory, because we had written the rules purely declaratively, it should not have mattered whether the rules were used to reason forwards or backwards, and indeed that was the case. The rules gave the intended results even though they were used to reason backwards and to fill in the triangle top down. But they were impossibly inefficient, taking about 20 minutes to produce the estimates for a single country. This was longer than we could afford within the schedule of the working group meetings.

At this point, most novice Prolog programmers would probably have thrown up their hands in despair, discarded the Prolog implementation, and perhaps even reprogramed the rules in a production system language, which would run the rules more naturally and more efficiently in the forward direction. A more experienced Prolog programmer, on the other hand, might have rewritten the program at a lower level, analogous to the first representation of the problem of going from *a* to *z*. However, we were able to preserve the naturalness of the representation and reduce the execution time to about 30 seconds, by changing only a few lines of code.

The inefficiency of backward reasoning in this case is due to the repeated recalculation of the lower-level sub-goals. For example, if there are 10 years between two adjacent anchor years, backward reasoning recalculates the estimates at the two anchor years ten times. The Prolog solution is to cut the program in half, run the level two goals first, assert their solutions into the Prolog database, and then run the level four rules, accessing the asserted level two solutions without re-computing them.

The alternative and preferable solution, is either to execute the rules in the forward direction, or to execute them in the backward direction, but save the solution of sub-goals automatically, so they do not need to be re-solved later. The latter of these two alternatives is provided by Prolog systems, such as XSB Prolog, which use the technique of tabling [7] to generate the solution of every subgoal only once. We transported the WUENIC system to XSB Prolog soon after diagnosing the problem.

2.5 Rule Refinement and Working Group Decisions

The WUENIC rules are analogous both to a collection of legal rules and to an expert system. As with other expert systems, was useful to develop the rules by successive refinement. We developed the first collection of rules to cover the most commonly occurring cases, and then refined the rules when they did not deal satisfactorily with more complicated cases. In theory, this process of refinement is never-ending, since the rules can always be improved. But in practice, we needed to firm up the rules, to make it easier to communicate them and to apply them in practice.

Partly as a compromise between facilitating future refinement of the rules and of finalizing them as early as possible, we introduced the notion of working group decisions, represented by “facts, which can be changed from one run to another without changing the rules themselves. In theory, any rule can by overridden by a working group decision, similarly to the way in which an exception overrides a general rule. This is represented by adding an extra negative condition to the rule and by adding an extra rule if necessary. For example:

if *the nationally reported data for a Country, Vaccine and Year is R*
and *there is* ***no*** *survey data S for the Country, Vaccine and Any-Year*
and *there is* ***no*** *working group decision to assign an estimate W*
for the Country, Vaccine and Year
then *the WUENIC estimate for the Country, Vaccine and Year is R.*

if *there is a working group decision to assign an estimate W*
for a Country, Vaccine and Year
then *the WUENIC estimate for the Country, Vaccine and Year is W.*

Working group decisions are added as “facts” to the Prolog database. The Prolog implementation combines these facts with other facts, applies the rules, and derives the WUENIC estimates. In practice, we found it useful to use working group decisions interactively, judging the quality of the resulting estimates, and repeating the process with modified decisions, until we are satisfied with the results.

Working group decisions are a convenient way of representing decisions in cases where there are no well-defined rules, or where is not convenient to state the rules explicitly. In both cases, they are analogous to the powers of discretion given to the Secretary of State in certain provision of the British Nationality Act. For example:

> 6.-(1) If, on an application for naturalisation as a British citizen made by a person of full age and capacity, the Secretary of State is satisfied that the applicant fulfils the requirements of Schedule 1 for naturalisation as such a citizen under this subsection, he may, if he thinks fit, grant to him a certificate of naturalisation as such a citizen.

Here the word, “may” seems to suggest the need for a probabilistic or modal logic. In fact, it is just a way of emphasizing that the conclusion depends upon the decision of the Secretary of State. In logic programming form, the provision can be expressed without probability or modality in the form:

the secretary of state ***will*** *grant a certificate of naturalisation*
to a person by section 6.1
if *the person applies for naturalisation*
and *the person is of full age and capacity*
and *the secretary of state is satisfied that the person*
fulfils the requirements of schedule 1 for naturalisation by 6.1
and *the secretary of state thinks fit*
to grant the person a certificate of naturalisation.

The positive condition *the secretary of state thinks fit to* ***grant*** *the person a certificate of naturalization* can also be written as a negative condition *the secretary of state does* ***not*** *think fit to* ***withhold*** *a certificate of naturalization from the person*, analogously to the way we represent working group decisions in WUENIC.

3 Discussion and Conclusions

The WUENIC approach builds upon the well-established use of logic programming for representing and reasoning about rules in legal documents. It also has many features in common with rule-based expert systems and business rule applications [8].

The confusion between logic programming rules and production rules became an early issue in the development of WUENIC. This affected not only the representation of the rules and the choice of implementation language, but it also added to the motivation of the first author to attempt to clarify the relationship between the two [9]. It also contributed to the development of the logic-based production system language LPS [10], which combines logic programming and production systems in a logic-based framework. Roughly speaking, in LPS, logic programs are used, as in WUENIC, to define concepts that are needed for decision-making in an organisation, whereas production rules are used to generate associated actions and workflows.

In addition to considering different kinds of rules, we also considered rule-based versus argumentation-based representations and implementations. Here the relationship between logic programming and Dung's argumentation semantics is already well understood [5], and it is relatively easy to translate one representation into the other. As a consequence, it is possible to implement an argumentation approach using logic programming rules, and to use the relationship to explain the resulting conclusions in argumentation terms.

In Dung' argumentation, an argument succeeds if it can be extended to a set of defending arguments that collectively defeats every argument that attacks any argument in the defending set. It suffices for the defending set to contain only one such defeating argument for every attacking argument. However, in the WUENIC application, arguments can attack and defeat one another to varying degrees, and the more supporting and defeating arguments there are the better.

This broader kind of argumentation in the WUENIC application means that WUENIC estimates have varying degrees of confidence and uncertainty. We are currently exploring the problem of representing and reasoning with uncertainty.

In addition to the problem of understanding the relationships between different knowledge representation and problem solving paradigms, we had greater than expected problems with our chosen logic programming paradigm. We did not anticipate the difficulties of choosing between different logic programming representations and the problems of improving the efficiency of our preferred representation. On the one hand, this highlights the difficulties that other implementers may face with other applications. On the other hand, it draws attention to the potential of tabling [7] to overcome some of the problems of programming in Prolog, and to facilitate its wider application in the future.

Acknowledgments. We are pleased to acknowledge the support and contributions and of the other members of the Working Group: Marta Gacic-Dobo from WHO and David Brown and Rouslan Karimov from UNICEF. Thanks also to the anonymous referees for their comments, which helped to improve the paper.

References

1. Sergot, M.J., Sadri, F., Kowalski, R.A., Kriwaczek, F., Hammond, P., Cory, H.T.: The British Nationality Act as a Logic Program. CACM 29(5), 370–386 (1986)
2. Burton, A., Monash, R., Lautenbach, B., Gacic-Dobo, M., Neill, M., Karimov, R., Wolfson, L., Jones, G., Birmingham, M.: WHO and UNICEF Estimates of National Infant Immunization Coverage: Methods and Processes. Bull. World Health Organization 87, 535–541 (2009)
3. Prakken, H., Sartor, G.: A Dialectical Model of Assessing Conflicting Arguments in Legal Reasoning. Journal of Artificial Intelligence and Law 4(3-4) (1996)
4. Dung, P.M.: On the Acceptability of Arguments and its Fundamental Role in Non-monotonic Reasoning, Logic Programming and N-person Games. Journal of Artificial Intelligence 77, 321–357 (1995)
5. Bondarenko, A., Dung, P.M., Kowalski, R., Toni, F.: An Abstract Argumentation-theoretic Approach to Default Reasoning. Journal of Artificial Intelligence 93(1-2), 63–101 (1997)
6. Dung, P.M., Kowalski, R., Toni, F.: Dialectic proof procedures for Assumption-Based, Admissible Argumentation. Journal of Artificial Intelligence 170(2), 114–159 (2006)
7. Sagonas, K., Swift, T., Warren, D.S.: XSB as an Efficient Deductive Database Engine. SIGMOD Rec. 23(2), 442–453 (1994)
8. Grosof, B.N., Labrou, Y., Chan, H.Y.: A Declarative Approach to Business Rules in Contracts: Courteous Logic Programs in XML. In: Proceedings of the 1st ACM Conference on Electronic Commerce, pp. 68–77. ACM, New York (1999)
9. Kowalski, R.: Computational Logic and Human Thinking: How to be Artificially Intelligent. Cambridge University Press, Cambridge (2011)
10. Kowalski, R., Sadri, F.: Integrating Logic Programming and Production Systems in Abductive Logic Programming Agents. In: Polleres, A., Swift, T. (eds.) RR 2009. LNCS, vol. 5837, pp. 1–23. Springer, Heidelberg (2009)
11. Brown, D.W., Burton, A., Gacic-Dobo, M., Karimov, R.I.: A Summary of Global Routine Immunization Coverage through 2010. Open Infect. Dis. J. 5., 115–117 (2011)
12. Kowalski, R.: Logic for Problem Solving. Elsevier Science Publishing Company (1979), `http://www.doc.ic.ac.uk/~rak/`

Document Structure Analysis with Syntactic Model and Parsers: Application to Legal Judgments

Hirokazu Igari, Akira Shimazu, and Koichiro Ochimizu

School of Information Science
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292, Japan
{higari,shimazu,ochimizu}@jaist.ac.jp
http://www.jaist.ac.jp/

Abstract. The structure of a type of documents described in a common format like legal judgments can be expressed by and extracted by using syntax rules. In this paper, we propose a novel method for *document structure analysis*, based on a method to describe syntactic structure of documents with an *abstract document model*, and a method to implement a document structure parser by a combination of *syntactic parsers*. The parser implemented with this method has high generality and extensibility, thus it works well for a variety of document types with common description format, especially for legal documents such as judgments and legislations, while achieving high accuracy.

Keywords: document structure analysis, document model, document structure parser.

1 Introduction

The advances of information processing technologies and the high speed network in recent years have made a huge number of electronic documents, which are available for retrieval. To help readers quickly find documents and information in them depending upon purposes and interests, new technology is awaited, which automates extraction of relevant information from documents.

As an approach for realizing such technology, we are proceeding with research on methods for extracting structure and relevant information from documents. According to our empirical observations, for a type of documents described in a common format, the structures often have high similarity to each other, and the same type of information is often included in similar parts. For example, *legal judgments* (or simply called as *judgments*), written by the courts for describing the decision based on the facts, have almost the same structure if the court and the case type are the same. In addition, relevant information in them, such as claims by the parties or decision by the court, appears in similar parts in almost the same order, and similar section titles are assigned to them. With these facts,

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 126–140, 2012.

it is expected that, by using structure information, a new method for extracting relevant information from documents with high accuracy can be realized.

As the first step, we are researching methods for *document structure analysis*, which is a task for extracting structure information from documents. In this paper, we propose a novel method for document structure analysis. While most of existing document structure analysis methods make use of layout information for each document page [6,7,9,12,14,16], our method instead uses syntactic information for the entire document text. This method performs very well for types of documents with a common description format, and for legal documents such as judgments and legislations in particular. Until now, we have designed and implemented a *document structure parser* for legal judgments made by a Japanese court[1].

To make the document structure parser general and extensible to support a variety of document types, we have invented two novel and effective methods; one is an abstract document model called *block structure model,* and the other is a method to implement syntactic parsers for block structure called *block breaker*s. These methods work very well for increasing generality and extensibility of the parser, as well as for achieving high accuracy of document structure extraction. Fig. 1 illustrates an overview of the document structure parser implemented with these methods. A judgment text is broken by block breakers into hierarchical text segments according to the block structure denition, then the leaf text segments are parsed by syntactic parsers to extract the document structure.

In the remaining of this paper, related works on document structure analysis are discussed in Section 2, document structure expressions of judgments with syntax rules and document structure model are explained in Section 3, and implementation of document structure parser with syntactic parsers and block breakers are explained in Section 4, followed by performance evaluation and conclusion in Section 5 and 6.

2 Related Works on Document Structure Analysis

Methods for document structure analysis have been widely researched since the 1990s [9], when many documents started to be transformed to electronic and structured formats such as SGML and XML, so that the contents can be processed by computers. The information on the structure of a document includes; *physical structure* which consists of visual elements in each document page and their layout information; *logical structure* which consists of logical elements in the entire document text and their order and inclusive relation; *semantic structure* which consists of topics in the document and their semantic relation[2].

The main target of document structure analysis has been logical structure [6,7,12,14,16], which consists of information such as metadata and section structure relevant for structuring and transforming documents to electronic format.

---

[1] Important judgments made by Japanese courts are available as PDF files with text from the Courts in Japan web site (`http://www.courts.go.jp/`).

[2] Names and definitions for document structure information categories vary between researches. These categories are mentioned in researches such as [3] and [9].

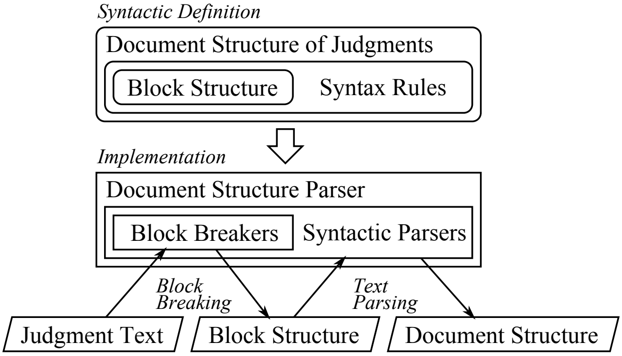


Fig. 1. Overview of Document Structure Parser

For most of the logical document structure analysis methods, physical structure is extracted first from document data such as OCR result or PDF file, and then logical structure is extracted by assigning logical roles to the physical elements and by analyzing their relations with layout information.

However, when documents are described in a common format, it is possible to express the logical structure by syntax rules for the document text. By taking advantage of this characteristic of such document types, the method we propose uses physical structure only for obtaining the document text, and extracts logical structure according to the syntactic document structure definition.

There have also been researches on document structure analysis for legal documents [1,10]. These methods are based on text grammar, similarly to our method, but the implemented parsers are specific to the target document types. Our method is more general and extensible based on an abstract document model and syntactic parsers, thus it works for a variety of document types.

Regarding semantic document structure, probabilistic methods for extracting topics in documents and analyzing their relation, such as *topic models* [3,8,15] and *content models* [2], have been proposed since the early 2000s. Most of these methods extract only document level topics, and few method considers logical document structure for extracting topics and their relation.

In this paper, we focus on logical document structure, and unless specifically mentioned, the term "document structure" refers logical document structure.

3 Document Structure of Legal Judgments

For legal judgments written by the same court for the same type of cases, their description formats have high commonality. For example, judgments made by the Japanese Intellectual Property High Court for patent cases are described in a format as Fig. 2. A judgment starts with case metadata consisting of judgment date, case number and name, and last debate date, followed by judgment type and parties. Sentence, fact and reason are described in a hierarchical section structure, and then court and judges are listed with any supporting papers.

(Judgment Date)	平成２１年１月２８日　判決言渡
(Case Numebr and Name)	平成１９年(行ケ)第１０２８９号　審決取消請求事件
(Last Debate Date)	平成２１年１月２８日　口頭弁論終結
(Judgment Type)	判　決
(Parties)	原告　レキシンジャパン株式会社
	被告　シコー株式会社
(Sentence)	主　文
(Section)	１　原告の請求を棄却する。
(Section)	２　訴訟費用は原告の負担とする。
(Fact and Reason)	事実及び理由
(Section)	第１　請求
	特許庁が…した審決を取り消す。
(Section)	第２　事案の概要及び判断
(Section)	１　被告らは，…。
	原告は，…。
(Section)	２　当裁判所の判断
	…よって，主文のとおり判決する。
(Court)	知的財産高等裁判所第３部
(Judges)	裁判長裁判官　飯村　敏明
	裁判官　中平　健

Fig. 2. Description Format of Legal Judgments by a Japanese Court

In addition, each element has unique prefix text pattern as Table 1, described by PEG parsing expressions which will be explained in the next section, thus the boundaries of elements can be recognized by text pattern matching. We take advantage of these characteristics of legal judgments for the document structure analysis method proposed in this paper.

3.1 Syntactic Document Structure Expressions

When a type of documents has a common description format, structure of documents can be formally expressed by syntax rules. For text parsing, syntax rules can be described by PEG (Parsing Expression Grammar) [4]. PEG is a derivative of CFG (context-free grammar), which is described by notations such as EBNF (Extended Backus-Naur Form), but PEG is different from CFG in that it is designed for text recognition and does not describe ambiguous syntax. Also, PEG is a formal description of *recursive descent parser* which performs top down syntax parsing. A PEG grammar G is defined as $G = (V_N, V_T, R, e_S)$, where:

- V_N is a finite set of *nonterminal symbols*
- V_T is a finite set of *terminal symbols*
- R is a finite set of *parsing rules*
- e_S is a *parsing expression* termed the start expression
- $V_N \cap V_T = \emptyset$
- Each parsing rule $r \in R$ is written as $A \leftarrow e$, where $A \in V_N$ and e is a parsing expression.
- For any nonterminal symbol A, there is exactly one parsing expression e for the parsing rule $A \leftarrow e \in R$.

Table 1. Prefix Text Patterns of Legal Judgments by a Japanese Court

Element	Prefix Text Pattern
Judgment Date	`Date "判決言渡"`
Case Number and Name	`CaseNumber CaseName`
Last Debate Date	`Date "口頭弁論終結"`
Judgment Type	`"判決"`
Parties	`(PartyTitle PartyName) | PartyAddress`
Sentence	`"主文"`
Fact and Reason	`"事実及び理由"`
Court	`CourtName CourtDivision`
Judges	`"裁判長"? "裁判官" JudgeName`

Terminal symbols are defined by the following expressions.

- '*abc*' or "*abc*", string literal
- $[abc]$ or $[a - c]$, character class
- ., any single character

Parsing expressions are defined as follows, where e, e_1 and e_2 are parsing expressions.

- ε, the empty string
- a, any terminal, where $a \in V_T$
- A, any nonterminal, where $A \in V_N$
- $e_1\, e_2$, a sequence
- $e_1 \,|\, e_2$, a prioritized choice
- $e*$, zero-or-more repetitions
- $e+$, one-or-more repetitions
- $e?$, an option
- $\&e$, an and-predicate - succeeds if e succeeds without consuming input
- $!e$, a not-predicate - succeeds if e fails without consuming input

A syntax rule for the top level document structure of the judgment presented in Fig. 2 can be described by PEG as follows, where nonterminals such as `JudgmentDate` and `CaseNumberNames` are defined separately.

```
Judgment <-
  (JudgmentDate | CaseNumberNames | LastDebateDate)+
  JudgmentType
  Parties
  Sentence
  ((Fact Reason) | FactReason)?
  CourtJudges
  AttachedPaper?
```

For example, syntax rules for `CaseNumberNames` can be described as follows.

```
CaseNumberNames <- CaseNumberName+
CaseNumberName  <-
  (CaseAbbName | CaseNumber | CaseName)+ |
  (LPar (CaseAbbNames | OriginalSentences) RPar)
```

At the bottom level of the document structure, syntax rules for `CaseNumber` can be described as follows, where nonterminals such as `CourtName` and `Year` are defined separately.

```
CaseNumber   <- CourtName? (Year Symbol? | '同')? SeqNumber
SeqNumber    <- '第' HyphenNumber '号'? | HyphenNumber '号'
HyphenNumber <- Number (Hyphen Number)*
Symbol       <- SymbolPrefix? LPar SymbolChar+ RPar
SymbolChar   <- SymbolPrefix | SymbolType | Kana
SymbolPrefix <- [刑合特]
SymbolType   <- [受許行家医収少人秩手日分甲]
```

By a combination of these syntax rules, the structure of the entire document can be defined. Note that document structure can be defined arbitrarily for its purpose. That is, multiple document structures can be defined for a single document type depending on the information to be extracted. For example, to extract only metadata, defining the document body as a single element is sufficient.

3.2 Block Structure Model

Although it is possible to implement a document structure parser according to syntax rules for the actual document elements, such implementation works only for a specific document type. To make the document structure parser implementation general and extensible, we have invented a novel method to describe document structure with an abstract document model, called *block structure model*, which expresses document structure with abstract document elements, called *blocks* and *nodes*. Blocks are hierarchical text segments corresponding to the actual document elements, and nodes are leaf text segments to be parsed for extracting information. There are two types of blocks, *composite block* and *node block*. A composite block contains a sequence of child blocks, and a node block contains a node. The boundaries of blocks are recognized by the prefix text pattern of nodes. To express block structure by syntax rules, the following nonterminal symbols are defined for block structure elements.

- `Block`, a text segment for a document element
- `CompositeBlock`, a block which contains a sequence of child blocks
- `NodeBlock`, a block which contains a node
- `Node`, a leaf text segment to be parsed

Since many documents including judgments have hierarchical section structure, extended block structure elements to describe the section structure are also defined.

- `SectionBlock`, a composite block for a section
- `SectionNodeBlock`, a node block for a section node
- `SectionNode`, a node which consists of section prefix and text to be parsed

The block structure model is expressed by syntax rules as follows.

```
Document         <- CompositeBlock
CompositBlock    <- Block+
Block            <- SectionBlock | CompositeBlock | NodeBlock
NodeBlock        <- Node
Node             <- &NodePrefix NodeText
SectionBlock     <- SectionNodeBlock SectionBlock*
SectionNodeBlock <- SectionNode
SectionNode      <- &SectionPrefix SectionText
SectionPrefix    <- (SenctionNumber+ SectionTitle?) | SectionTitle
```

To describe document elements with abstract block structure elements, actual element names such as `Judgment` and `JudgmentDate` are appended with parentheses as the following example.

```
Document(Judgment)
CompositeBlock(Judgment)
NodeBlock(JudgmentDate)
SectionBlock(Sentence)
```

With the block structure model, the syntax rules for the top level document structure of the judgment presented in Fig. 2 can be rewritten as follows, and the judgment text is broken into blocks and nodes as Fig. 3.

```
Document(Judgment) <- CompositeBlock(Judgment)
CompositeBlock(Judgment) <-
  (NodeBlock(JudgmentDate) | NodeBlock(CaseNumberNames) |
   NodeBlock(LastDebateDate))+
  NodeBlock(JudgmentType)
  NodeBlock(Parties)
  SectionBlock(Sentence)
  (SectionBlock(Fact) SectionBlock(Reason)) |
   SectionBlock(FactReason))?
  NodeBlock(CourtJudges)
  NodeBlock(AttachedPaper)?
```

4 Document Structure Parser

A document structure parser extracts document structure from a given text of a document according to the syntactic document structure definition described as syntax rules with the block structure model. It consists of *syntactic parsers* implemented according to the syntax rules and extracts document structure by the following two steps of tasks.

- *Block Breaking* - Break a document text into blocks and nodes with *block breakers*, which are syntactic parsers for block structure.

Fig. 3. Block Structure of Legal Judgments by a Japanese Court

- *Text Parsing* - Parse nodes in the extracted block structure with syntactic parsers to extract information in them.

The extracted document structure is available as a block structure containing the text parsing results at its nodes. It can be traversed and transformed to any output format such as XML.

4.1 Syntactic Parsers

A parser which parses a given text according to syntax rules described by PEG can be implemented while preserving the structure of the syntax rules, with *parser combinators* [5,11]. We call such parsers as *syntactic parsers*. Parser combinators is a method to implement *recursive descent parsers*, and an implementation is provided as a standard library in *Scala* program language [13]. The structure of parser code written with parser combinators exactly matches the structure of syntax rules, thus implementing and understanding parsers is straightforward and easy. In addition, new parsers can be implemented by combinations of existing parsers, thus parsers can be easily and flexibly reused. Especially, parsers implemented with the Scala parser combinators have operators corresponding to PEG parsing expressions, thus any syntax rules can be directly implemented while preserving the structure as it is. To take this advantage, we developed the document structure parser with Scala. With the Scala parser combinators, syntactic parsers for `CaseNumberNames` presented above can be implemented as follows.

```
def caseNumberNames = (caseNumberName)+
def caseNumberName  =
  ((caseAbbName | caseNumber | caseName)+) |
  (lPar ~ (caseAbbNames | originalSentences) ~ rPar)
```

Similarly, syntactic parsers for `CaseNumber` can be implemented as follows.

```
def caseNumber =
  (courtName?) ~ (((year ~ (symbol?)) | '同')?) ~ seqNumber)


def seqNumber =
  ('第' ~ hyphenNumber ~ (('同')?)) | (hyphenNumber ~ '号')
def hyphenNumber = number ~ ((hyphen ~ number)*)
def symbol = (symbolPrefix?) ~ lPar ~ (symbolChar+) ~ rPar
def symbolChar = symbolType | symbolPrefix | kana
def symbolPrefix = among("刑合特")
def symbolType = among("受許行家医収少人秩手日分甲")
```

4.2 Block Breakers

Since the end of a block is determined by the prefix of the subsequent block, it is not possible to implement a parser for block breaking by a combination of independent syntactic parsers. To resolve this issue, we have invented a novel method to implement syntactic parsers for block breaking, which extract block structure while recognizing the boundaries of blocks, called *block breakers.* A block breaker generates two syntactic parsers, called *first* and *blocks,* which are used for implementing a new block breaker by a combination of existing block breakers. The notations `first(next)` and `blocks(next)` denote syntactic parsers, which are dynamically generated depending on the parameter `next`.

– `BlockBreaker`, a block breaker
  - `first(next)`, a syntactic parser which recognizes the prefix of the first existing node depending on `next`
  - `blocks(next)`, a syntactic parser which extracts the target block structure depending on `next`
  - `next`, one of the following parsers is used
    * `first` of a subsequent block breaker and used when there is a subsequent block breaker
    * `EOF`, a parser which recognizes the end of input and used when there is no subsequent block breaker
    * `NEVER`, a parser always fails and used when there is no need to consider subsequent block breaker
  - `blocks(EOF)`, a standalone syntactic parser which extracts the target block structure

The prefix of the first node in the target block is recognized by `first`, but the target block can be optional, thus the first existing node may be contained in a subsequent block. To handle such cases, `next` is used to recognize the first existing node. As `next`, `first` of the subsequent block breaker is used when there is a subsequent block breaker. Otherwise, `EOF` is used when there is no subsequent block, or `NEVER` is used when there is no need to consider subsequent blocks. As a special case, `block(EOF)` generates a standalone syntactic parser which extracts the target block structure from a given text.

To describe the structure of block breakers by syntax rules, block breakers for block types and the syntax rules for `first` and `blocks` are defined as follows. Nodes are recognized and parsed in terms of *node types* associated with two parsers, one recognizes node prefix and the other parses node text respectively. In addition, since sibling sections have sequential section numbers with the same indent, section nodes are recognized in terms of *section context* containing node type and other context information. The notations such as `b.first(next)`, `nodeType.prefix` and `context.nodeType` denote members of block breaker, node type and section context respectively.

- `CompositeBlockBreaker(b)`, extracts a composite block where `b` is a block breaker for child blocks
    - `first(next)  <- b.first(next)`
    - `blocks(next) <- b.blocks(next)`
- `NodeBlockBreaker(nodeType)`, extracts a node block where `nodeType` is the target node type
    - `first(next)  <- nodeType.prefix`
    - `blocks(next) <- nodeType.prefix (!next textLine)*`
- `SectionBlockBreaker(context)`, extract a section block where `context` is the section context
    - `SectionBlockBreaker(context) <- SectionNodeBlockBreaker(context) SectionBlockBreaker(context)*`
- `SectionNodeBlockBreaker(context)`, extract a section node block where `context` is the section context
    - `SectionNodeBlockBreaker(context) <- NodeBlockBreaker(context.nodeType)`

For any block breaker `b`, `b1` and `b2`, a new block breaker can be created by using the same syntax expression as the target block structure.

- `b1 b2`, a sequence
    - `first(next)  <- b1.first(b2.first(next))`
    - `blocks(next) <- b1.blocks(b2.first(next)) b2.blocks(next)`
- `b1 | b2`, a prioritized choice
    - `first(next)  <- b1.first(NEVER) | b2.first(next)`
    - `blocks(next) <- b1.blocks(next) | b2.blocks(next)`

- `b*`, zero-or-more repetitions
  - `first(next)  <- b.first(NEVER) | next`
  - `blocks(next) <- b.blocks(first(next))*`
- `b+`, one-or-more repetitions
  - `b+ <- b b*`
- `b?`, an option
  - `first(next)  <- b.first(NEVER) | next`
  - `blocks(next) <- b.blocks(next)?`

Fig. 4 shows how a new block breaker is implemented by a combination of existing block breakers. In this example, a sequence of block breakers `b1` and `b2`, described as `b1 b2`, is implemented by using `first` and `blocks` of `b1` and `b2`. The parser `next` used by `b1 b2` is determined depending on whether it has subsequent block breaker. If it has subsequent block breaker `b3`, `first` of `b3` is used as `next`. Otherwise, `EOF` or `NEVER` is used depending on the context in which `b1 b2` is used in the syntax rules.

Structure of block breakers can be expressed by the same syntax rules as the target block structure as follows.

```
DocumentBreaker      <- CompositeBlockBreaker
CompositBlockBreaker <- BlockBreaker+
BlockBreaker         <-
  SectionBlockBreaker | CompositeBlockBreaker | NodeBlockBreaker
SectionBlockBreaker  <- SectionNodeBlockBreaker SectionBlockBreaker*
```

As a result, the top level block breaker structure of the judgment presented in Fig. 2 can be expressed by syntax rules as follows.

```
DocumentBreaker(Judgment) <- CompositeBlockBreaker(Judgment)
CompositeBlockBreaker(Judgment) <-
  (NodeBlockBreaker(JudgmentDate) | NodeBlockBreaker(CaseNumberNames) |
   NodeBlockBreaker(LastDebateDate))+
  NodeBlockBreaker(judgmentType)
  NodeBlockBreaker(Parties)
  SectionBlockBreaker(Sentence)
  (SectionBlockBreaker(Fact) SectionBlockBreaker(Reason)) |
   SectionBlockBreaker(FactReason))?
  NodeBlockBreaker(CourtJudges)
  NodeBlockBreaker(AttachedPaper)?
```

Block breakers for these syntax rules can be implemented by a combination of block breakers as it is, where `compositeBlockBreaker`, `sectionBlockBreaker` and `nodeBlockBreaker` denote methods to create the corresponding types of block breakers.

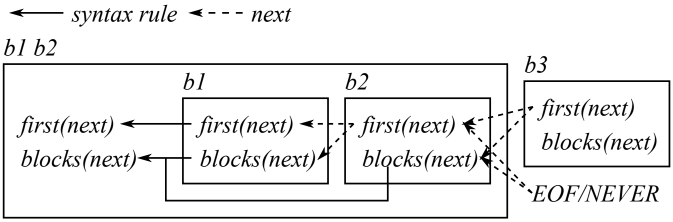


Fig. 4. Implementation of a Sequence of Block Breakers

```
def judgmentDocumentBreaker = judgmentCompositeBlockBreaker
def judgmentCompositeBlockBreaker = compositeBlockBreaker(
  ((nodeBlockBreaker(JudgmentDate) | nodeBlockBreaker(CaseNumberNames) |
    nodeBlockBreaker(LastDebateDate)+) ~
  nodeBlockBreaker(JudgmentType) ~
  nodeBlockBreaker(Parties) ~
  nodeBlockBreaker(Sentence) ~
  (((sectionBlockBreaker(Fact) ~ sectionBlockBreaker(Reason)) |
    sectionBlockBreaker(FactReason))?) ~
  nodeBlockBreaker(CourtJudges) ~
  (nodeBlockBreaker(AttachedPaper)?)
```

With these block breakers and the syntactic parsers for the nodes, the document structure parser extracts the document structure from a given text of a judgment.

5 Performance Evaluation

To evaluate the document structure parser implemented with the proposed method, performances are calculated for the current version of the parser and the baseline system. As a baseline system, a document structure parser based on non-syntactic method is used, which is actually used for creating legal case data for a Japanese legal information service. It extracts block structure by a program logic specific to judgments according to predefined rules. For 245 judgments made by the Japanese Intellectual Property Court for patent cases in 2009, extracted block structures are compared in terms of the node paths with the manually extracted block structures consisting of 25,213 nodes. The result is summarized in Table 2, which consists of three subtables.

The numbers of extracted nodes, their successes and errors, followed by precision, recall and f-score are listed in the first subtable. Errors are categorized in terms of whether nodes are extracted by only the parser (Type I) or by only manually (Type II). The current parser recognizes 24,922 nodes with 1,403 Type I and 1,694 Type II errors, and precision, recall and f-score are 0.944, 0.933 and 0.938. On the other hand, the baseline system recognizes 27,899 nodes with 7,368 Type I and 4,682 Type II errors, and precision, recall and f-score are 0.736, 0.814 and 0.773. This result shows significant improvement by the current method.

Table 2. Performance Evaluation of Document Structure Parser

Performance: Judgments=245 Manual Nodes=25,213

	nodes	success	TypeI err	TypeII err	precision	recall	f-score
Current	24,922	23,519	1,403	1,694	0.944	0.933	0.938
Baseline	27,899	20,531	7,368	4,682	0.736	0.814	0.773

F-Score Distribution: Judgments=245

	judgments		percentage		accumulated		percentage	
f-score	curr	base	curr	base	curr	base	curr	base
1.0	70	19	28.6%	7.8%	70	19	28.6%	7.8%
0.9 - 1.0	127	86	51.8%	35.1%	197	105	80.4%	42.9%
0.8 - 0.9	28	43	11.4%	17.6%	225	148	91.8%	60.4%
0.7 - 0.8	9	29	3.7%	11.8%	234	177	95.5%	72.2%
0.6 - 0.7	9	27	3.7%	11.0%	243	204	99.2%	83.3%
0.5 - 0.6	2	27	0.8%	11.0%	245	231	100.0%	94.3%
0.2 - 0.5	0	14	0.0%	5.7%	245	245	100.0%	100.0%

Error Detail: Judgments=10 Manual/Parser Nodes=1,736/1,599

	Type I				Type II			
Type	number	title	prev.err	total	number	indent	prev.err	total
Errors	18	11	131	157	23	24	244	294
Ratio	11.5%	7.0%	83.4%	100%	7.8%	8.2%	83.0%	100%

The numbers and percentages of judgments for f-score ranges, and their accumulated numbers and percentages are listed in the second subtable. The current parser recognizes nodes perfectly for 70 judgments (28.6%), with greater than 0.9 f-score for 197 judgments (80.4%), and with greater than 0.8 f-score for 225 judgments (91.8%). Although the structures and the node prefix patterns of the judgments should have only a small variance, since they are all made for patent cases within a year, this result still can be considered as fairly good, taking account of the size of judgments consisting of 102 nodes in average with hierarchical section structure. However, some judgments have a low f-score. According to the analysis, most of these judgments have quoted text of other judgments, which are not considered by the current document structure parser.

Error types, their numbers and ratios for 10 sampled judgments by the current parser are listed in the third subtable. For 1,736 manually extracted nodes, 1,599 nodes are extracted by the parser, and all errors are for section prefix recognition. Among 157 Type I errors, 18 (11.5%) section numbers and 11 (7.0%) titles which are not actually section prefixes are misrecognized. Among 294 Type II errors, 23 (7.8%) are by unexpected section number patterns, and 24 (8.2%) are by illegal indent greater than the predefined threshold. Other errors, 131 (83.4%) Type I and 244 (83.0%) Type II errors, are caused by previous prefix recognition errors. According to this result, it is expected that the parser performance can be further improved by adding more section prefix patterns and introducing new methods to improve section prefix recognition.

6 Conclusion

In this paper, we proposed a method for document structure analysis, which is based on a method to describe syntactic document structure with an abstract model, and a method to implement a document structure parser by a combination of syntactic parsers. The parser implemented with this method extracts the document structure with high accuracy from documents described in a common format and the structure can be expressed by syntax rules. In addition, the parser has high generality and extensibility, thus works well for a variety of document types, especially for legal documents such as judgments and legislations.

However, there are still some elements which can not be recognized correctly by the current method. For example, it is difficult to determine by the deterministic syntactic rules if the text in the first line of a section is title or part of a paragraph. In the current document structure parser implementation, this is determined by a program logic considering features in the line such as suffix pattern and length; however, many titles are still not recognized correctly.

To realize a new method for extracting relevant information from documents using structure information, it is essential that correct document structure is provided. Thus, to further improve the accuracy of document structure extraction, we are working to enhance the current method to support ambiguous syntax rules and to make use of machine learning methods. By selecting the most probable result among the alternatives returned for ambiguous syntax rules with machine learning methods, many elements not recognized by the current method are expected to be recognized more correctly.

References

1. Bacci, L., Spinosa, P., Marchetti, C., Battistoni, R., Senate, I.: Automatic mark-up of legislative documents and its application to parallel text generation. IDT, 45 (2009)
2. Barzilay, R., Lee, L.: Catching the drift: Probabilistic content models, with applications to generation and summarization. In: Proceedings of HLT-NAACL, vol. 2004 (2004)
3. Blei, D.M., Lafferty, J.D.: Topic models. Text Mining: Classification, Clustering, and Applications 10, 71 (2009)
4. Ford, B.: Parsing expression grammars: a recognition-based syntactic foundation. ACM SIGPLAN Notices 39, 111–122 (2004)
5. Hutton, G., Meijer, E.: Monadic parsing in haskell. Journal of Functional Programming 8(4), 437–444 (1998)
6. Klink, S., Dengel, A., Kieninger, T.: Document structure analysis based on layout and textual features. In: Proc. of International Workshop on Document Analysis Systems, DAS 2000, pp. 99–111. Citeseer (2000)
7. Lee, K.H., Choy, Y.C., Cho, S.B.: Logical structure analysis and generation for structured documents: a syntactic approach. IEEE Transactions on Knowledge and Data Engineering, 1277–1294 (2003)
8. Li, W., McCallum, A.: Pachinko allocation: Scalable mixture models of topic correlations. J. of Machine Learning Research (2008) (submitted)

9. Mao, S., Rosenfeld, A., Kanungo, T.: Document structure analysis algorithms: a literature survey. In: Proc. SPIE Electronic Imaging, vol. 5010, pp. 197–207. Citeseer (2003)
10. Moens, M.F., Uyttendaele, C.: Automatic text structuring and categorization as a first step in summarizing legal cases. Information Processing & Management 33(6), 727–737 (1997)
11. Moors, A., Piessens, F., Odersky, M.: Parser combinators in scala. CW Reports, vol. CW491. Department of Computer Science, KU Leuven (2008)
12. Namboodiri, A., Jain, A.: Document structure and layout analysis. In: Digital Document Processing, pp. 29–48 (2007)
13. Odersky, M., Altherr, P., Cremet, V., Emir, B., Maneth, S., Micheloud, S., Mihaylov, N., Schinz, M., Stenman, E., Zenger, M.: An overview of the scala programming language. Technical report. Citeseer (2004)
14. Rangoni, Y., Belaïd, A.: Document Logical Structure Analysis Based on Perceptive Cycles. In: Bunke, H., Spitz, A.L. (eds.) DAS 2006. LNCS, vol. 3872, pp. 117–128. Springer, Heidelberg (2006)
15. Steyvers, M., Griffiths, T.: Probabilistic topic models. In: Handbook of Latent Semantic Analysis, vol. 427(7), pp. 424–440 (2007)
16. Summers, K.: Automatic discovery of logical document structure (1998)

Recall-Oriented Evaluation Metrics for Consistent Translation of Japanese Legal Sentences

Yasuhiro Ogawa, Masaki Mori, and Katsuhiko Toyama

Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8603 Japan
{yasuhiro,toyama}@is.nagoya-u.ac.jp
http://www.kl.i.is.nagoya-u.ac.jp/

Abstract. We propose new evaluation metrics for statutory sentences translated from Japanese to English. Since translation variety is unacceptable and consistency is crucial in legal translation, a new metric called CIEL has been proposed that evaluates translation consistency. That metric is based on precision with a recall strategy. However, since we believe that recall-oriented metrics are more suitable for consistent evaluation, we propose recall-oriented metrics to evaluate the consistency of legal translations and illustrate their performances by comparing them with other metrics.

Keywords: ROSSO, translation evaluation metric, legal translation, consistency, ROUGE.

1 Introduction

Recently, the social demand for the translation of Japanese statutes into foreign languages has been increasing to conduct international transactions more smoothly, to promote international investment in Japan, to support legal reform in developing countries, and so on. Since Japanese statutes have been individually translated by government ministries or private publishing companies, translation equivalents may be inconsistent among translated documents. For example, in various legal documents, the Japanese legal term "弁護士 (*bengoshi*)" was translated as "*attorney*," "*barrister*," or "*lawyer*," all of which have different meanings in English. Therefore if "*attorney*" is used in one document and "*lawyer*" is used in another document for the same Japanese term, readers may be confused. For this reason, the same translation equivalent must be used for the same term; consistent translation is required.

To solve this problem, the Japanese government has compiled a *Japanese-English Standard Bilingual Dictionary*[1] [16,17] for legal technical terms in Japanese statutes. It currently includes about 4,400 Japanese entries and about 5,750 English equivalents. Japanese statutes are being translated by the government in compliance with this dictionary. The next task is quality evaluation of the translations in compliance with the dictionary.

Since a term sometimes has several translation equivalents, a suitable one in context should be used in a translation. For example, in the Standard Bilingual Dictionary,

---

[1] http://www.japaneselawtranslation.go.jp/

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 141–154, 2012.

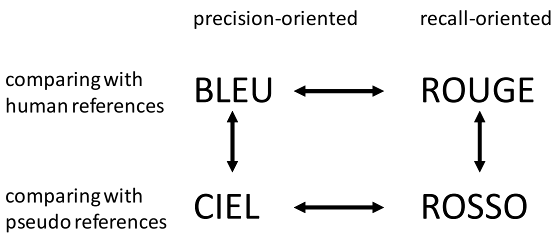


Fig. 1. Relations among four metrics

the term “免除する (*menjo-suru*)” has six equivalents: “*release*,” “*exempt*,” “*waive*,” “*exculpate*,” “*remit*,” and “*immunize*.” We should choose the most suitable one depending on the context. Although notices for the choice might be roughly given to some equivalents in the dictionary, registering every detailed criterion for the choice in the dictionary is not easy. Thus it is insufficient to only rely on the dictionary for consistent translations.

To overcome this problem, a new automatic metric called CIEL [14] is proposed. It is a derivative of the BLEU metric [15] and evaluates translation consistency. Since BLEU’s basic idea is to compare machine translations with human reference translations that are considered correct, it requires such reference translations. In other words, the BLEU metric can evaluate machine translations as long as the human reference translations are given. On the other hand, it is impossible to prepare human reference translations for the evaluation of legal translations. In fact, if there exist *correct* reference translations, we no longer need other translations. Therefore, the CIEL metric prepares pseudo reference translations (PRTs), which are the translations of similar sentences with a source sentence. The CIEL metric is modified from BLEU to use PRTs and successfully evaluated the translation consistency.

Here, note that BLEU is a metric based on precision. Since the same fixed expressions in the source sentences must have the same translation form the viewpoint of translation consistency, expressions that frequently appear in the reference translations must occur in a candidate translation. Therefore the evaluation metrics for consistency should also consider recall, which motivated CIEL’s design.

However, an experiment compared CIEL with other metrics and found that the ROUGE metrics [10,11], which are recall-oriented, also achieved good results [14]. This suggests that evaluation metrics for consistency should be basically designed as recall-oriented from the beginning. Thus, in this paper, we consider possibility of such evaluation metrics and propose a new metric called ROSSO, which is recall-oriented and a derivative of ROUGE tailored with PRTs. Figure 1 shows the relations among BLEU, ROUGE, CIEL, and ROSSO.

We applied ROUGE, ROSSO, CIEL, and other metrics to three kinds of translations of the Labor Standard Act (Act No. 49 of 1947): by the Japanese government, a publishing company, and the Google translation tool as well as Ogawa et al.[14]. The ROSSO and CIEL metrics successfully distinguished the translations by the government from those by the publishing company, but the ROUGE metrics also did. Therefore, we confirmed that, although the modification on the ROSSO metrics tailored with PRTs is ineffective, the recall-oriented strategy is useful for the consistent evaluation of legal translations.

This paper is organized as follows. In Section 2 we introduce the BLEU, CIEL, and ROUGE metrics. Next, we propose our evaluation metrics called ROSSO in Section 3. Then we describe some evaluation experiments in Section 4. Finally, Section 5 is a conclusion.

2 Previous Metrics

In this section, we summarize previous metrics: BLEU, CIEL, and ROUGE.

2.1 BLEU

The BLEU metric [15] is an automatic evaluation metric for machine translation. It compares n-grams in the candidate translation, which is a machine translation sentence for a given source sentence, with n-grams in the human reference translations. Since several translations are possible for one source sentence, the BLEU metric prepares multiple human translations as references. For comparison, the following precision score p_N is calculated:

$$p_N = \frac{\sum_{S \in CT} \sum_{gram_N \in Grams_N(S)} Count_{clip}(gram_N, S)}{\sum_{S \in CT} \sum_{gram_N \in Grams_N(S)} Count(gram_N, S)}, \tag{1}$$

$$Count_{clip}(gram_N, S) = \min\left(Count(gram_N, S), \max_{R \in RT(S)} (Count(gram_N, R)) \right), \tag{2}$$

where CT is a candidate translation document and S is a sentence in CT. $Grams_N(S)$ is a set of n-grams with length N in S. $Count(gram_N, S)$ is the number of occurrences of n-gram $gram_N$ in S. $Count_{clip}(gram_N, S)$ is also the number of the occurrences of $gram_N$ in S, but if it is greater than the maximum number of occurrences of $gram_N$ that occur in any single reference translation R, $Count_{clip}(gram_N, S)$ equals the maximum number. $RT(S)$ is a set of reference translations for S. If $gram_N$ does not occur in any reference translations, $Count_{clip}(gram_N, S)$ is 0. The external sum ranges over all candidate translations in the document, which means that the BLEU metric evaluates the whole translation document.

Next, if the candidate translation is shorter than its reference translations, the denominator of the above formula becomes smaller so that p_N becomes larger. To penalize this situation, the BLEU metric computes brevity penalty (BP):

$$\mathrm{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{1-r/c} & \text{if } c \le r \end{cases}, \tag{3}$$

where c is the length of the candidate translation and r is its effective reference length.

Finally, introducing positive weights w_N based on the value of N, the BLEU score is defined as follows:

$$\mathrm{BLEU} = \mathrm{BP} \cdot \exp\left(\sum_{N=1}^{M} w_N \log p_N\right) . \tag{4}$$

Usually, the upper of N is set to be $M = 4$ and uniform weights $w_N = 1/M$. Using n-grams up to length M, the BLEU metric evaluates both the adequacy and the fluency of the candidate translations, where adequacy indicates how much information is retained in the translation and fluency indicates to what extent the translation reads like good English.

2.2 CIEL

The CIEL metric [14] is based on the BLEU metric and evaluates the human translations of statutory sentences from the viewpoint of translation consistency. For this purpose, the CIEL metric uses a legal parallel corpus and compares candidate translations with it.

The first difference between the BLEU and CIEL metrics is that CIEL cannot prepare human reference translations that are considered *correct*. In fact, if there exists a *correct* translation of a statutory sentence, it does not need to evaluate other translations any more. Therefore, metrics for consistent evaluation need to evaluate human translations without *correct* references.

To overcome this problem, the CIEL metric focused on the fact that Japanese statutory sentences have many fixed expressions. This is because the Cabinet Legislation Bureau reviews most Japanese statutes and controls the use of legal terms and expressions in the statutes during their drafting. That is, evaluation consistency can be judged by referring to existing translations of statutory sentences.

The CIEL metrics use a parallel corpus of Japanese statutes to evaluate consistency, retrieve similar sentences to a given source sentence, and collect their translations from the corpus. Since such translations may not be exact translations of source sentences, they are called *pseudo reference translations* (PRTs).

Acquisition of Pseudo Reference Translations. Ogawa et al. [14] used a hierarchical clustering method [7] to obtain a set of PRTs. They divided the source sentences in the corpus into clusters and selected the closest one to a given source sentence. Since such clusters contain similar sentences to the source, their translations were collected as PRTs. The following shows the details of the clustering method in Ogawa et al. [14].

First, they split a set of source sentences since the cost of clustering tasks for all sentences is considered to be too high. They used the peculiarity of Japanese language, that is, main predicates occur at the end of sentences and play an important role in sentences. So they split the source sentences by their last morphemes and reduced the clustering cost.

Next, they deleted all *bunsetsus*,[2] except the last one, those depending on the last one, and those depending on them. This is to delete non-fixed expressions from the sentences.

For example, consider the following sentence:

---

[2] A *bunsetsu* is the smallest coherent components in a Japanese sentence[6].

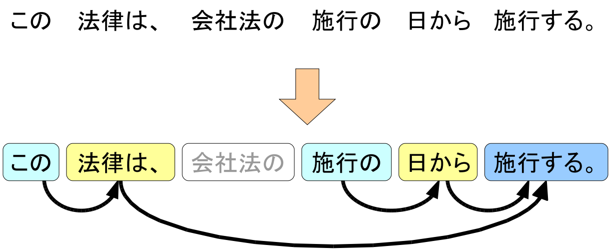


Fig. 2. *Bunsetsu* deletion

Sentence 1: この 法律は、 会社法の 施行の 日から 施行する。
(*This Act shall come into force as from the date of enforcement of the Companies Act*).

While Sentence 1 consists of six *bunsetsus*, the following *bunsetsus* are left after the deletion:

1. "施行する (*shall come into force*)";
this is the last *bunsetsu*.
2. "法律は (*Act*)" and "日から (*as from the date*)";
these depend on the last *bunsetsu* "施行する."
3. "この (*This*)" and "施行の (*of enforcement*)";
"この" depends on "法律は" and "施行の" depends on "日から."

This result is illustrated in Fig. 2. Sentence 1 is transformed into

この法律は、施行の日から施行する。
(*This Act shall come into force as from the date of enforcement*).

In order to analyze dependency relations between *bunsetsus*, CaboCha[8] was used, which is a Japanese dependency/syntactic parser based on machine learning and achieves about 90% accuracy.

After transforming the source sentences as above, Ogawa et al. [14] applied hierarchical clustering. They used the group average method and the morpheme-based edit distance. The distance between two sentences is defined as the minimum number of operations needed to transform one sentence into the other, where an operation is the one of the insertion, deletion, or substitution of a single morpheme. However this distance is sensitive to the sentence length, so it was normalized into interval $[0,1]$ by dividing by the sentence length.

Ogawa et al. [14] used the resulting clusters as PRTs except those containing only one sentence since such clusters are unreliable for evaluation.

Furthermore, fixed sentences are used in many statutes. For example, the sentence

Sentence 2: この法律は、公布の日から施行する。
(*This Act shall come into force as from the day of promulgation.*)

occurs in many statutes. From the viewpoint of consistent translation, the same source sentences should be translated into the same translation. If the same sentence was included more than once in the corpus, the translations of the sentence were used as reference translations instead of the cluster.

Modifying BLEU Metric. Since PRTs may not be exact translations of source sentences, some n-grams occurring in the candidate translation may not occur in the PRTs, reducing the BLEU score. The CIEL metric resolved this problem by introducing a weight $w(gram_N)$ that indicates the ratio of sentences containing $gram_N$:

$$w(gram_N) = \frac{\text{\# of sentences with } gram_N \text{ in PRTs}}{\text{\# of sentences in PRTs}} . \tag{5}$$

Although introducing $w(gram_N)$ successfully removed the negative effects of n-grams that only occur in the candidates, it causes another problem; wrong translations do not reduce the BLEU score since they do not occur in any reference translations. Therefore the CIEL metric introduced a recall-oriented strategy. It defines $TopGrams_N(\alpha, PRTs)$ as the set of n-grams occurring more often than the ratio $\alpha (0 \le \alpha \le 1)$ in the PRTs as follows:

$$TopGrams_N(\alpha, PRTs) = \{gram_N \in Grams_N(PRTs) \mid w(gram_N) > \alpha\} , \tag{6}$$

Using $TopGrams_N(\alpha, PRTs)$, the CIEL metric is defined as follows:

$$p_N = \frac{\sum_{gram_N \in S \cup TopGrams_N(\alpha, PRTs)} Count_{clip}(gram_N, S) \cdot w(gram_N)}{\sum_{gram_N \in S \cup TopGrams_N(\alpha, PRTs)} \max(Count(gram_N, S), 1) \cdot w(gram_N)} , \tag{7}$$

$$\text{CIEL} = \exp(\sum_{N=1}^{M} w_n \log p_N) . \tag{8}$$

Note that the CIEL metric does not include the multiplication of BP, which is used in the BLEU metric (4) as a penalty for shorter candidate translations. This is because the length of PRTs has nothing to do with the candidate translations. The original BLEU metric evaluates the whole translation document, while the CIEL metric evaluates each sentence.

The CIEL metric is basically a derivative of the BLEU metric with a recall-oriented strategy that came from the ROUGE metric described in the next subsection.

2.3 ROUGE

The ROUGE metric is a standard evaluation measure in automatic text summarization. Among its variations, our idea is based on the basic ROUGE-N metric and the LCS-based ROUGE-L, whose details are described in [10,11].

ROUGE-N: N-gram Co-Occurrence Statistics. The ROUGE-N metric [10] compares n-grams of two summaries and counts the matches. When $n = 1, 2, 3, \cdots$, the metric is called ROUGE-1, ROUGE-2, ROUGE-3, respectively. The measure is defined by the following equation:

$$\text{ROUGE-N} = \frac{\sum\limits_{R \in RS} \sum\limits_{gram_N \in Grams_N(R)} Count_{match}(gram_N, R)}{\sum\limits_{R \in RS} \sum\limits_{gram_N \in Grams_N(R)} Count(gram_N, R)}, \tag{9}$$

where RS is a set of reference summaries and $Count_{match}(gram_N, R)$ is the maximum number of n-grams co-occurring in a candidate summary and a set of reference summaries.

Comparing the BLEU and ROUGE metrics, (1) resembles (9). The BLEU metric counts n-grams in a candidate translation, so that it is a precision-oriented approach. On the other hand, the ROUGE metric counts n-grams in a reference, so that it is a recall-oriented approach.

ROUGE-L: Longest Common Subsequence. The ROUGE-L metric [11] is based on the longest common subsequence (LCS). Given two sequences, X and Y, the LCS of X and Y is a common subsequence with maximum length [3]. To apply LCS in a translation evaluation, a translation is considered a sequence of words. Intuitively, the longer the LCS of two translations is, the more similar they are. The ROUGE-L metric is a LCS-based F-measure that estimates the similarity between two translations, R of length M and S of length N, assuming R is a reference translation and S is a candidate translation, as follows:

$$R_{lcs} = \frac{LCS(R,S)}{M}, \tag{10}$$

$$P_{lcs} = \frac{LCS(R,S)}{N}, \tag{11}$$

$$\text{ROUGE-L} = F_{lcs} = \frac{(1+\beta^2) R_{lcs} P_{lcs}}{R_{lcs} + \beta^2 P_{lcs}}, \tag{12}$$

where β is a non-negative real. When β is 1, F_{lcs} is a harmonic mean of R_{lcs} and P_{lcs}. If $\beta < 1$, ROUGE-L is weighted toward precision; otherwise it is weighted toward recall. Usually $\beta \to \infty$ is used, that is, ROUGE-L $= R_{lcs}$, considering only recall.

Notice that ROUGE-L is 1 when R and S are identical since $LCS(R,S) = M$ or N; ROUGE-L is zero when $LCS(R,S) = 0$, i.e. R and S share nothing.

3 ROSSO: Proposed Metric

Although the ROUGE metrics are evaluation methods for summarization, they have been applied to translation evaluation [11,12]. In fact, the IQ$_{MT}$ framework[3] [5], which

---

[3] http://www.lsi.upc.edu/~nlp/IQMT/

is an open source framework for automatic machine translation evaluation, adopts the ROUGE metrics as well as the BLEU, METEOR [1], and NIST [4] metrics.

As mentioned in Section 1, it is expected that recall-oriented metrics achieve good results for consistency evaluation. Thus, we proposed a new recall-oriented consistency evaluation metric considering the ROUGE metrics. First, applying the ROUGE metric with PRTs, we point out its problems. Next we modify it and propose our new metric: ROSSO.

3.1 Problems with ROUGE-N Metric

Since our evaluation does not target machine translation but the consistency of human translation, we use the ROUGE-N metric with PRTs instead of reference translations as in the case of the CIEL metric. However, this approach causes problems. For example, consider the following two candidate translations:

Source: 次に掲げる者は、監督委員 となることができない。
Candidate 1: The following persons may not act as supervisors:
Candidate 2: The following persons may not act as *supervising* committee members:

For comparison, we prepared the following two PRTs:

Pseudo Reference 1: The following persons may not act as directors:
(次に掲げる者は、取締役 となることができない。)
Pseudo Reference 2: The following persons may not act as *supervising* immigration inspectors:
(次に掲げる者は、主任審査官 となることができない。)

Both Candidates 1 and 2 obviously resemble each other, and the only difference is the translation equivalent of “監督委員”: “supervisors” and “supervising committee members.” We cannot evaluate which equivalent is better since “監督委員” does not occur in the references. Therefore, only the underlined parts of the sentences should be evaluated and both candidates should have the same scores, even though their ROUGE scores are different. As shown in (9) of Section 2.3, the ROUGE-1 score is calculated by dividing the number of the unigrams in both a candidate and any of its references by the number of the unigrams in the references. In fact, the ROUGE-1 score of Candidate 1 is $(7+7)/(8+10) = 14/18 = 0.778$. In the same way, the score of Candidate 2 is $(7+8)/(8+10) = 15/18 = 0.833$. This difference is caused by “*supervising*” in both Candidate 2 and the Pseudo Reference 2.

3.2 ROSSO-N

To overcome this problem, we ignore the low frequent expressions in PRTs and introduce a weight $tw(\alpha, gram_N)$ that indicates the ratio of the occurrence of $gram_N$ in PRTs:

$$tw(\alpha, gram_N) = \begin{cases} w(gram_N) & \text{if } w(gram_N) > \alpha \\ 0 & \text{if } w(gram_N) \leq \alpha \end{cases}, \tag{13}$$

where we set $tw(\alpha, gram_N)$ as 0 if it is less than threshold value $\alpha (0 \leq \alpha \leq 1)$. Our proposed metric ROSSO is defined as follows:

$$\text{ROSSO-N} = \frac{\sum\limits_{R \in PRTs} \sum\limits_{gram_N \in Grams_N(R)} Count_{match}(gram_N, R) \cdot tw(\alpha, gram_N)}{\sum\limits_{R \in PRTs} \sum\limits_{gram_N \in Grams_N(R)} Count(gram_N, R) \cdot tw(\alpha, gram_N)}. \tag{14}$$

Notice that the original ROUGE-N metric uses reference translations instead of reference summaries when it evaluates translation. Here R indicates a certain reference document and the ROUGE-N metric evaluates the whole translation document. Contrary to this, we consider that R is a certain sentence in the reference, that is, we make the ROUGE-N and ROSSO-L metrics evaluate each sentence.

When we set parameter α in (14) to 0.5, the ROSSO-1 score of Candidate 1 is $(7+7)/(7+7) = 1.000$ and the same with the score of Candidate 2.

3.3 ROSSO-L

We also modify the ROUGE-L metric and propose the ROSSO-L metric. We prepared multiple PRTs but (11) only considers one reference. Therefore we redefine the ROUGE-L metric as follows in order to calculate the score for all references.

$$\text{ROUGE-L} = \frac{\sum\limits_{R \in PRTs} LCS(R, S)}{\sum\limits_{R \in PRTs} \sum\limits_{gram_1 \in Grams_1(R)} Count(gram_1, R)}, \tag{15}$$

where the length of a sentence in the PRTs is calculated using unigram $gram_1$.

Similar to the ROSSO-N metric, we ignore infrequent expressions in PRTs and introduce $tw(\alpha, gram_N)$. Therefore we define a new metric ROSSO-L as follows:

$$\text{ROSSO-L} = \frac{\sum\limits_{R \in PRTs} LCS_{tw}(R, S)}{\sum\limits_{R \in PRTs} \sum\limits_{gram_1 \in Grams_1(R)} Count(gram_1, R) \cdot tw(\alpha, gram_1)}, \tag{16}$$

where $LCS_{tw}(R, S)$ is the maximum value of $tw(\alpha, CS)$ and CS is a common subsequence between PRT R and candidate S. Notice that we make the ROUGE-L and ROSSO-L metrics evaluate each sentence as well as the ROUGE-N metric.

4 Evaluation Experiment

We evaluated our proposed metrics ROSSO-N and ROSSO-L by experimentally comparing them with previously proposed metrics.

4.1 Experimental Targets

We calculated the ROSSO-N, ROSSO-L, ROUGE-N, ROUGE-L, BLEU, CIEL, Word Error Rate (WER) [9], Position independent Word Error Rate (PER) [9], METEOR [1], and NIST [4] scores. We used the same data as Ogawa et al. [14]. Our evaluation target was translations of the Labor Standards Act (Act No. 49 of 1947) that contains 242 sentences as main provisions and three kinds of translations: by the Japanese government, a publishing company [13], and a Google translation tool.

The government translation, which was done by legal specialists using the Japanese-English Standard Bilingual Dictionary (SBD)[1] [16], was provided on the web[1] by the Japanese government. The company translation was done by legal specialists without using the SBD. The government translation is expected to be more consistent than the company one due to the SBD. The Google translation was the result of a machine translation system created by Google[4], so it can be considered inferior to the others. We expect the proposed metric to rank them in this order.

For the compilation of the PRTs, we used a parallel corpus of Japanese statutes translated by the Japanese government[1], including 34,873 Japanese sentences of 162 acts and bylaws, where we excluded the Labor Standards Act and deleted duplicated sentences. We divided the 34,873 Japanese sentences into clusters and selected the closest cluster to each source sentence as mentioned in Section 2.2. Notice that we used more sentences than Ogawa et al. [14] for the PRTs since the Japanese government has currently provided more translations than were available when Ogawa et al. did their work.

They used 220 sentences in the Labor Standards Act in their first experiment, but they pointed out that the scores of the metrics were unreliable when the distance is large between a source sentence of a candidate translation and its closest cluster. Therefore we used the top sentences with close clusters for our evaluations.

In fact, it is a matter to decide how many sentences we should use. Given a sufficiently large sample size, a statistical comparison always show a significant difference unless the difference of populations means is exactly zero. Thus we need to determine an appropriate sample size. In power analysis [2], the sample size is determined by significance level, power and effect size, which is a statistical term that represents how powerful or large a difference is. In this experiment, we used the top 51 sentences since we set a medium effect size (i.e., 0.5), 0.01 significance level, and 0.80 power.

In the CIEL metric, we set parameter α of $TopGrams_N$ to 0.5, since we consider that n-grams appeared in more than half of PRTs are fixed expressions. We also set parameter α to 0.5 in the ROSSO-N and ROSSO-L metrics.

4.2 Experimental Results

We calculated each metric score and its average as shown in Table 1. We marked the values with an asterisk (*) that are significantly greater ($p < 0.01$) than those in the immediately right column. All metrics showed significant differences between the Google translation and the others. In contrast, only the CIEL, ROSSO-N, and ROUGE-N metrics showed significant differences between the government translation and the company one; only these metrics can distinguish which translation is better.

---

[4] `http://www.google.com/translate_t/`

Table 1. Average scores of evaluation metrics

metric	government	company	Google
ROUGE-1	* 0.372	* 0.343	0.232
ROUGE-2	* 0.168	* 0.139	0.042
ROUGE-3	* 0.116	* 0.083	0.009
ROUGE-4	* 0.088	* 0.055	0.002
ROSSO-1	*** 0.514**	*** 0.479**	**0.325**
ROSSO-2	*** 0.334**	*** 0.282**	**0.094**
ROSSO-3	*** 0.268**	*** 0.202**	**0.031**
ROSSO-4	*** 0.218**	*** 0.139**	**0.009**
ROUGE-L	0.335	* 0.315	0.198
ROSSO-L	**0.486**	*** 0.465**	**0.308**
BLEU	0.210	* 0.184	0.070
CIEL	* 0.425	* 0.328	0.099
1-WER	0.171	* 0.164	0.094
1-PER	0.291	* 0.271	0.166
NIST	1.845	* 1.726	1.071
METEOR	0.325	* 0.321	0.235

4.3 Discussion

Table 1 shows that the ROUGE-N metric can evaluate consistency without the modification. Thus introducing a weight is ineffective, which differs from the case of the BLEU metric.

In order to analyze the result, we also counted the number of desirable cases, where the score of the government translation exceeds the company one, and the number of undesirable cases, where the score of the government translation is inferior to the company one as shown in Table 2. The even cases indicate that both the government and company translations have the same scores; two of 51 sentences have identical translations between the government and the company.

The several undesirable cases are not unsuitable since some company translations are more consistent than the government's, as described in Ogawa et al. [14]. However, some of the undesirable cases are unsuitable, meaning the metrics scored more consistent translations as lower. Although we have expected that such undesirable cases in the ROUGE-N metric would be eliminated in the ROSSO-N metric, the numbers of undesirable cases are similar between them. This result implies that the problem with ROUGE-N metric considered in Section 3.1 seldom occurred. From this, we conclude that the ROUGE-N metric is enough to evaluate the consistency without the weight $tw(\alpha, gram_N)$. However, since all the three high-performing metrics use the recall-oriented strategy, it is effective for consistency evaluation.

On the other hand, the ROUGE-N and ROSSO-N metrics have a defect. While the numbers of undesirable cases are small in the BLEU, CIEL, ROUGE-N, and ROSSO-N metrics, the ROUGE-N and ROSSO-N metrics have more even cases than the BLEU and CIEL metrics. One reason is that the ROUGE-N and ROSSO-N metric confines the count to frequent n-grams occurring in PRTs. Therefore, counted n-grams in the

Table 2. Desirable and undesirable cases

metric	desirable	even	undesirable
ROUGE-1	34	7	10
ROUGE-2	31	10	10
ROUGE-3	30	14	7
ROUGE-4	24	20	7
ROSSO-1	**32**	**8**	**11**
ROSSO-2	**27**	**14**	**10**
ROSSO-3	**25**	**20**	**6**
ROSSO-4	**20**	**27**	**4**
ROUGE-L	24	11	16
ROSSO-L	**27**	**8**	**16**
BLEU	38	4	9
CIEL	38	4	9
1-WER	18	19	14
1-PER	31	7	13
NIST	21	10	20
METEOR	29	3	19

candidate translation are sometimes similar between the government and the company translation. In fact, for the ROUGE-3, ROUGE-4, ROSSO-3, and ROSSO-4 metrics, both the scores of the government and company translations are sometimes 0; no frequent 3- or 4-grams in the PRTs occur in the candidate translations, which increased the number of even cases.

Tables 1 and 2 show that the ROSSO-L metric result was no good. $LCS_{tw}(R,S)$ may be unsuitable for consistency evaluation.

5 Conclusion

In this paper, we proposed two consistency evaluation metrics for legal translations: the ROSSO-N and ROSSO-L metrics, which are derivatives of ROUGE. The ROSSO-N metric is based on n-gram alignment scoring and is recall-oriented. We confirmed that the ROSSO-N metrics can evaluate several translations of one source sentence from the viewpoint of consistent translation, but the ROUGE-N metric also can. Although we failed to show the effectiveness of modification in the ROSSO-N metric, we confirmed that the recall-strategy is effective for consistency evaluation and the ROUGE-N metrics is applicable with PRTs. Since this result would possibly depend on construction method of PRTs, we intend to try other construction methods.

Our experiments showed that the ROUGE-N, ROSSO-N, and CIEL metrics are useful but failed to determine which is the best. Therefore, future work includes a greater comparison between them to examine how they correlate to intuitive evaluations by human experts.

We intend to apply our proposed metrics to the Japanese Law Translation Database Systems[1] [18] to determine whether the first versions of the translation statutes provided by the Japanese government are appropriate for its database that aims for consistent and reliable translations.

The ROSSO-N and CIEL metrics are designed for evaluation of legal translation and they use the peculiarity of Japanese statutory sentences that have many fixed expressions to compile PRTs. However, these metrics can be applied to other domains such as translation of technical manuals, which we want to investigate in future work.

Acknowledgments. The authors thank Professor Yoshiharu Matsuura, Associate Professors Tokuyasu Kakuta and Designated Assistant Professor Makoto Nakamura, Japan Legal Information Institute (JaLII), Graduate School of Law, Nagoya University for their helpful discussions and suggestions. This research project was partly supported by a Grant-in-Aid for Scientific Research (S) and (B) from the Japan Society for the Promotion of Science.

References

1. Banerjee, S., Lavie, A.: Meteor: An automatic metric for mt evaluation with improved correlation with human judgements. In: Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, pp. 65–72 (2005)
2. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences, 2nd edn. Lawrence Erlbaum (1988)
3. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms, 2nd edn. MIT Press (2001)
4. Doddington, G.: Automatic evaluation of machine translation quality using n-gram cooccurrence statistics. In: Proceedings of the Second International Conference on Human Language Technology Research, pp. 138–145 (2002)
5. Giménez, J., Amigó, E.: IQ$_{MT}$: A Framework for Automatic Machine Translation Evaluation. In: Proceedings of the 5th International Conference on Language Resources and Evaluation, pp. 685–690 (2006)
6. Hashimoto, S.: Kokugo-hou Kenkyuu. Iwanami Shoten (1948) (in Japanese)
7. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Computing Surveys 31(3), 264–323 (1999)
8. Kudo, T., Matsumoto, Y.: Japanese Dependency Analysis using Cascaded Chunking. In: Proceedings of the International Conference on Computational Linguistics, pp. 1–7 (2002)
9. Leusch, G., Ueffing, N., Ney, H.: A novel string-to-string distance measure with applications to machine translation evaluation. In: Proceedings MT Summit IX, pp. 240–247 (2003)
10. Lin, C.: ROUGE: a package for automatic evaluation of summaries. In: Proceedings of Text Summarization Branches Out, Workshop at the ACL 2004, pp. 74–81 (2004)
11. Lin, C., Och, F.J.: Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Main Volume, pp. 605–612 (2004)
12. Lin, C., Och, F.J.: ORANGE: a Method for Evaluating Automatic Evaluation Metrics for Machine Translation. In: Proceedings of the 20th International Conference on Computational Linguistics, pp. 501–507 (2004)
13. Nakane, F.: EHS Law Bulletin Series: Japan, vol. VIII. Eibun-Horei-Sha, Inc. (2002)

14. Ogawa, Y., Imai, K., Toyama, K.: Evaluation Metrics for Consistent Translation of Japanese Legal Sentences. In: Francesconi, E., Montemagni, S., Peters, W., Tiscornia, D. (eds.) Semantic Processing of Legal Texts. LNCS, vol. 6036, pp. 235–248. Springer, Heidelberg (2010)
15. Papineni, K., Roukos, S., Ward, T., Zhu, W.: Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 311–318 (2002)
16. Study Council for Promoting Translation of Japanese Laws and Regulations into Foreign Languages: Final report (2006), `http://www.cas.go.jp/jp/seisaku/hourei/report.pdf`
17. Toyama, K., Ogawa, Y., Imai, K., Matsuura, Y.: Application of word alignment for supporting translation of Japanese statutes into English. In: van Engers, T.M. (ed.) Legal Knowledge and Information Systems JURIX 2006: The Nineteenth Annual Conference, pp. 141–150 (2006)
18. Toyama, K., Saito, D., Sekine, Y., Ogawa, Y., Kakuta, T., Kimura, T., K., Matsuura, Y.: Design and Development of Japanese Translation Database System. Law via the Internet 2011, 12 pages (2011)

Deliberation Process Support System for Citizen Judge Trial Based on Structure of Factors

Takanori Sato, Shogo Okada, and Katsumi Nitta

Interdisciplinary Graduate School of Science and Engineering,
Tokyo Institute of Technology, Japan
{tsato,okada,nitta}@ntt.dis.titech.ac.jp

Abstract. In 2009, the Japanese government adopted the citizen judge system. In this system, three professional judges and six citizen judges listen to the arguments between the prosecutor and the attorney, and decide the judgment through the discussion in deliberation. However the presiding judges have not been trained for moderation sufficiently, so that their skills of moderation affect the performance and quality of the discussion. Therefore, in this paper, we propose a deliberation process support system. The system assists the presiding judge to facilitate the deliberation by some functions. First it visualizes an argument structure graph, representing summary of the arguments between the prosecutor and the attorney. Next it recommends topics which need argument and their order. We propose a novel algorithm to select these topics based on preliminary research. Finally it also recommends speakers who did not have the opportunity to make remarks.

Keywords: Argument Visualization, Argumentation, Moderator Support.

1 Introduction

In May 2009, the Japanese government adopted the citizen judge system. Citizen judges chosen from ordinary people started to participate in trials as judges. At first in the trial, the prosecutor argues. Next, the attorney counters the prosecutor's argument. After that, professional judges and citizen judges have a discussion based on the arguments. It is called deliberation. In deliberation, they decide whether the accused is guilty or innocence. If guilty, they also decide the punishment. The presiding judge plays a key role as the moderator.

In deliberation, several problems exist. At first, many topics are intricately related to other ones [1]. Therefore, the citizen judges are confused about the topics under discussion in deliberation because they deal with the huge quantity of information. Second, the presiding judge needs to moderate the discussion that many individuals participate in [2]. The presiding judges have not been trained for moderating discussion, so that their skills of moderation affect the performance and quality of the discussion. Next, the time for deliberation is limited. Therefore, the presiding judge needs to insure that the discussion is effective

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 155–169, 2012.

during the limited time. Finally, the citizen judges have very little knowledge of the law. Therefore the presiding judge needs to inform them of what the legal knowledge is required.

To solve the above problems, it is necessary for the presiding judges to select topics properly. In addition, the presiding judge has to give a fair chance to remark and proper advice to each participant. In the situation described above, a system to support the presiding judge in deliberation is needed.

For developing this kind of the system, it is promising that it has functions of argument visualization and moderator navigation. In related studies, Reed et al. proposed Araucaria [3]. This system analyzes arguments and visualizes them as a diagram. It is used for education and intended for use as argument analysis. In addition, Loukis et al. proposed 'Computer Supported Argument Visualization' (CSAV) [4]. It is a system that was for remote support of legislation debate. It has focused on argument visualization. Nohara et al. proposed an method of argument using the "chart method" in deliberation [1]. The "chart method" is a series of methods, making "a chart" and processing the deliberation using it. This approach allows the participants to share information, so that they can grasp which topic is discussed. Hotta quantitatively analyzes simulated deliberations in the citizen judge system [2]. He uses the quantity of utterances in some simulated deliberations and compares between the characteristics of the participants and it. In addition, Anzai et al. developed an annotation system for the citizen judge system. It can annotate records of the deliberation and visualize information about the deliberation [5]. This system is for analysis as to what is good deliberation. As mentioned above, there have been extensive researches done regarding argument analysis and visualization. However systems with the ability to make navigation to moderator are rare.

In this paper, we propose a deliberation process support system for the presiding judge to carry out the deliberation smoothly. The system visualizes the argument summary as a graph. In addition, it gives some recommended information to the presiding judge to moderate discussion smoothly.

In section 2, we introduce the system outline. In section 3, we show factor registration editor. In section 4, we show deliberation process support system. In section 5, we show results of evaluation of the proposed system. In section 6, we give our conclusion.

2 Overview of Proposed System

As our deliberation process support system uses a factor based approach, we define a factor at first. Then, we show the overview of our system.

2.1 Factor

A factor is a proposition representing a fact, an opinion, or a claim. We define it on the bases of the factor in [6].

A factor has information such as "ID", "state", "meaning", "type", "support", and "conflict". "ID" is the ID to identify each factor. "State" refers to the

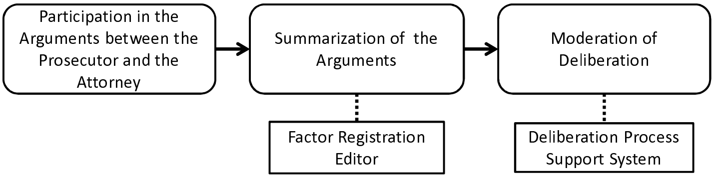


Fig. 1. Step Flow of the Citizen Judge Trial Using the Deliberation Process Support System

position of the person claiming an issue, taking “k”, “b”, or “o” as the public prosecutor’s claims, the attorney’s claims, or the other state. “Meaning” is a description of the factor. “Type” refers to the type of the factor and is registered as one of the following three types.

- Penalty factor
  This factor is related to the legal issues. It is what the prosecutor and the attorney claim as to what punishment is appropriate.
  e.g. “The accused is to be in prison in five years.”, “The accused should be declared innocent.”
- Main factor
  This factor is needed to think whether the accused is guilty or not, or whether the punishment is serious or not. It assists penalty factors.
  e.g. “The accused had a motivation.”, “The accused regrets having committed the crime.”
- Evidence factor
  This factor represents the fact, the testimony of witnesses, the evidence of the case, and so on. It assists main factors.
  e.g. “The fingerprints of the accused were found on the knife left at the crime scene.”

“Support” refers to support factors, which current factor assists as the evidence or the cause. “Conflict” refers to conflict factors, with which the current factor conflicts.

2.2 Overview of Citizen Judge Trial Using Deliberation Process Support System

We propose a deliberation process support system. The target user is the presiding judge. To use the system, there are three steps. Fig.1 shows the step flow.

At first, the judges participate in the arguments between the prosecutor and the attorney. Next, the presiding judge uses a factor registration editor to summarize the arguments. Then, the system makes a factor list, a collection of factors. After creating the factor list, the editor outputs an argument structure graph. In deliberation, the presiding judge moderates the deliberation using the graph. He/she inputs factors which appeared in the participant’s remark in remark record table. The system visualizes old remarks on the table. In addition, it analyzes the graph and the table to notify the user of two types of recommendation. One recommendation is the factors which need arguments and their order.

The other is the speakers who have not had the opportunity to speak about each factor. The user decides the next factor or speaker with the information and facilitates the deliberation. After the participant make a remark, the phase of inputting the remarks is again used and the cycle repeats.

The editor and the system were developed using Java and prefuse [7][1].

3 Factor Registration Editor

3.1 Overview

This editor helps the user to register factor to summarize the arguments between the prosecutor and the attorney. The editor has the following functions.

- Registration of factors and creation of argument structure graph
  The user registers factors based on records of the arguments between the prosecutor and the attorney. A collection of the registered factors is called a factor list. The system makes a graph structure, called a argument structure, from the factor list.
- Visualization of argument structure graph
  It visualizes the argument structure graph, representing the relationship of the factors based on the factor list.

Fig.2 shows system architecture of the editor. The editor displays records of the arguments and helps the user register factors. Then it creates argument structure. In registering factors, it displays the graph. After the user registers the factors, the editor outputs the graph data.

Fig.3 shows the editor. The editor shows records of the arguments as text, factor configuration space, and factor list. In this space, the user can register or modify factors and the factor list, which is a collection of registered factors.

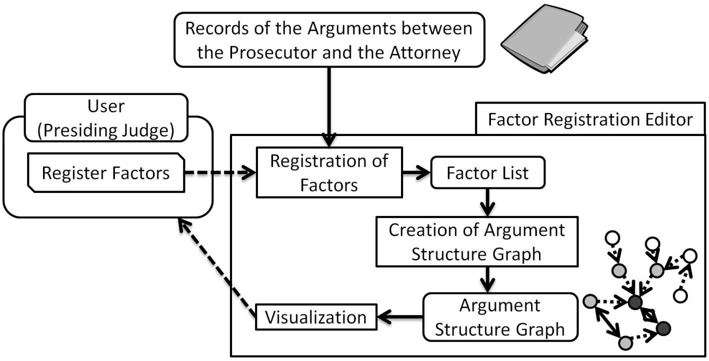
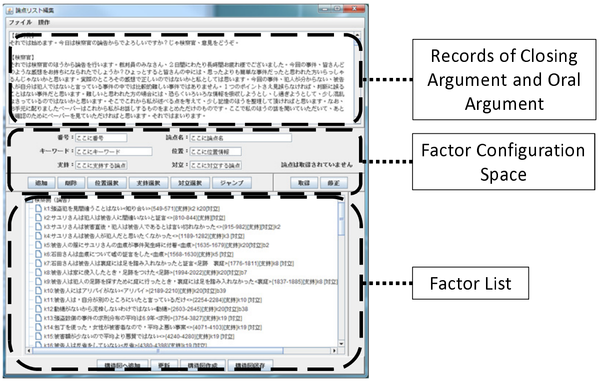


Fig. 2. System Architecture of Factor Registration Editor

---

[1] For more information and download, access to `http://prefuse.org/`



それでは始めます。今日は検察官の論告からでよろしいですか？じゃ検察官、意見をどうぞ。

【検察官】
それでは検察官のほうから論告を行います。裁判員のみなさん、２日間にわたり長時間お疲れ様でございました。今回の事件、皆さんどのような感想をお持ちになられたでしょうか？ひょっとすると皆さんの中には、思ったよりも簡単な事件だったと思われた方いらっしゃるんじゃないかと思います。実際のところその感想で正しいのではないかと私としては思います。今回の事件、犯人が分からない、被告人が自分は犯人ではないと言っている事件の中では比較的難しい事件ではありません。１つのポイントさえ見誤らなければ、判断に誤ることはない事件だと思います。難しいと思われた方の場合には、恐らくいろいろな情報を吸収しようとし、し過ぎようとして、少し混乱なさっているのではないかと思います。そこでこれから私が述べる点を考えて、少し記憶のほうを整理して頂ければと思います。なお、お手元に配りましたペーパーはこれから私がお話しするものをまとめただけのものです。ここで私のほうの話を聞いていただいて、あと確認のためにペーパーを見ていただければと思います。それではまいります。

番号：ここに番号
論点名：ここに論点名
キーワード：ここにキーワード
位置：ここに位置情報
支持：ここに支持する論点
対立：ここに対立する論点
論点は取得されていません

検察側（論告）
- k1:強盗犯を見間違うことはない<知り合い>{549-571}[支持]k2 k20[対立]
- k2:サユリさんは犯人は被告人に間違いないと証言<>{810-844}[支持][対立]
- k3:サユリさんは被害直後，犯人は被告人であるとは言い切れなかった<>{915-982}[支持][対立]k2
- k4:サユリさんは被告人が犯人だと思いたくなかった<>{1189-1282}[支持]k3 [対立]
- k5:被告人の服にサユリさんの血痕が事件発生時に付着<血痕>{1635-1679}[支持]k20[対立]b2
- k6:石田さんは血痕について嘘の証言をした<血痕>{1568-1630}[支持]k5 [対立]
- k7:石田さんは被告人は裏庭には足を踏み入れなかったと証言<足跡　裏庭>{1776-1811}[支持]k8 [対立]
- k8:被告人は家に侵入したとき，足跡をつけた<足跡>{1994-2022}[支持]k20[対立]b7
- k9:被告人は犯人の足跡を探すために庭に行ったとき，裏庭には足を踏み入れなかった<裏庭>{1837-1885}[支持]k8 [対立]
- k10:被告人にはアリバイがない<アリバイ>{2189-2210}[支持]k20[対立]b39
- k11:被告人は，自分が別のところにいたと言っているだけ<>{2254-2284}[支持]k10 [対立]
- k12:動機がないから泥棒しないわけではない<動機>{2603-2645}[支持]k20[対立]b38
- k13:強盗致傷の事件の求刑分布の平均は6.9年<求刑>{3754-3827}[支持]k19 [対立]
- k14:包丁を使った，女性が被害者なので，平均より悪い事案<>{4071-4103}[支持]k19 [対立]
- k15:被害額が少ないので平均より悪質ではない<>{4240-4280}[支持]k19 [対立]
- k16:被告人は反省をしていない<反省>{4380-4398}[支持]k19 [対立]

Fig. 3. Snapshot of Factor Registration Editor

3.2 Registration of Factors and Creation of Argument Structure Graph

The user inputs records of the arguments to the editor. The user registers factors by referring to the records. Specifically to register factors, the user inputs information such as “ID”, “state”, “meaning”, “type”, “support”, and “conflict” described in section 2.1. The registered factors are included in factor list.

After the user finishes registering factors, the system makes the factor list into a graph. This graph is called the argument structure graph. It consists of nodes and edges. Each node represents a factor. It has information, node ID, which contains “state” and “ID”, and “meaning”. On the other hand, each edge represents a relationship between factors. It has information of the type of relationship. The edge registered “support” represents a directed dotted arrow. In addition, the edge registered “conflict” represents a bidirected solid arrow. If looking at the visual graph when creating factor list, the user can consider the type of relationship between a factor being registered and factors that have already been registered.

The argument structure graph can be thought to summarize the arguments between the prosecutor and the attorney. The user can use the graph to proceed with arguments in the deliberation later. To do this, the editor has the function to output data to be input to the deliberation process support system. This data is represented by the GraphML. It has information required to produce the graph.

4 Deliberation Process Support System

4.1 Overview

The system supports the user to facilitate moderation during deliberation. The system functions are as follows.

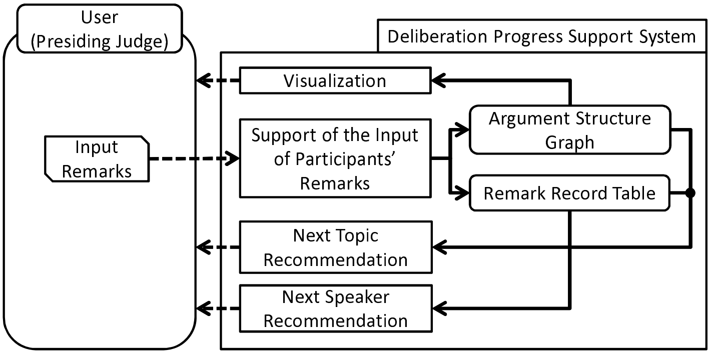


Fig. 4. System Architecture of Deliberation Process Support System

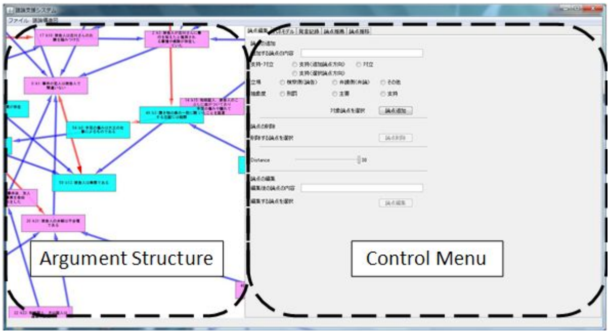
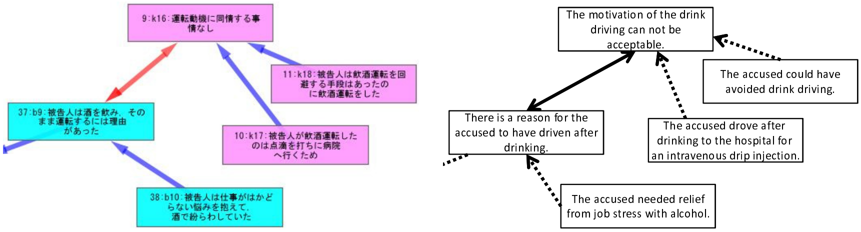


Fig. 5. Snapshot of Deliberation Process Support System

- Visualization of argument structure graph
  It gets a file of argument structure graph which is output from the factor registration editor and visualizes the graph.
- Support of the input of participants' remarks
  It provides an table in which the user inputs factors which appeared in the participants' remarks.
- Next topic recommendation
  It indicates which factors should be discussed in deliberation.
- Next speaker recommendation
  It indicates who should make remarks.

Fig.4 shows the system architecture. Before deliberation, the system gets the data of the argument structure graph and displays the graph. In addition, the user sets the deliberation limit time as remaining time. During deliberation, the user inputs the participants' remarks in a table, called a remark record table. Using some data, the system decides factors to be discussed and speakers required to make remarks. After that, it notifies the user of the factors and the speakers.

Fig.5 shows the system. It displays the argument structure graph and the menu to input various operations.



(a) Original View (b) Translation in English of (a)

Fig. 6. Argument Structure Graph Displayed in Deliberation Process Support System

4.2 Visualization of Argument Structure Graph

The system receives the data created from the factor registration editor and displays the graph.

Fig.6 shows a diagram of the graph when the data is input in the system. The input data is created from moot court. The number of factors is 49. As well as the factor registration editor, the system allows the user to add or remove a factor on the menu during deliberation. Furthermore, if it is difficult to see the graph, it is possible to control it for easier viewing. For example, when factors overlap because the edges are too short, the user can solve the problem by lengthening the edges. The user can moderate the deliberation by confirming which factor is being discussed by using this function.

4.3 Support of the Input of Participants' Remarks

In deliberation, the presiding judge listens to participants' opinions and summarizes them. The summarized opinions are basis of adjudication. This is called fact-finding. In Japan, presiding judges are required to do so by the law. Hence presiding judges have a heavy burden from this work. So the system offers a record table for summarizing opinions.

The user inputs the remarks or the opinions of factors in the remark record table. The table consists of rows of the factors and columns of the participants. The user fills in each participant's remarks or opinions about each factor on a corresponding cell. More specifically, the user clicks the node from the graph and fills in a displayed cell where the participant's remark meets the corresponding factor. When confirming remarks of the factor, the user clicks the factor, and then the system displays the remark record table of the clicked factor.

Each cell of the table has an "approval" or a "denial" tag. Each tag represents the standpoint that the participant takes. An "approval" tag represents that the participant supports the factor. A "denial" tag represents that the participant opposes the factor. "approval" cells are painted blue, "denial" cells are painted red. Furthermore, cells which are not filled in are colored in yellow, and cells which are filled in with something are colored white. The user can visually recognize the conditions of the cells.

4.4 Next Topic Recommendation

The system has a function to recommend factors which are to be discussed. Thus, the user complies with the recommended factors so as to moderate deliberation smoothly.

The recommended factor is selected by the factor selection algorithm. This algorithm uses the argument structure graph. It outputs recommended factors and their order.

Preliminary Research. When making the algorithm, during preliminary research, we examined the difference of factor selection between a professional of the law and an amateur in the moot court.

At first, we collected the simulated deliberation records. The one moderated by a law professional acquired from the records of a moot court of citizen judge system. The records included arguments between the prosecutor and the attorney, and the deliberation. In order to compare the deliberation record, we held simulated deliberation moderated by an amateur and got the records. In the real deliberation of citizen judge trial, three professional judges (including a presiding judge) and six citizen judges participated. However, in the simulated deliberation moderated by an amateur, only a presiding judge and two citizen judges participated. The role of the presiding judge was to moderate the deliberation and to argue his/her opinions. The role of the citizen judge was to only argue his/her opinions. They summarized their opinions and decided which the accused was guilty or not and the punishment if guilty. A flow of the simulated deliberation by amateurs is described below. At first, we explained the summary of citizen judge system, the basic mechanism of the citizen judge trial, and the basic method of deliberation to the participants. Then, they read records of the arguments between the prosecutor and the attorney. The records were parts of the moot court mentioned above. Next they discussed the simulated case and decided whether the accused was guilty or not. If guilty, they continued the discussion and decided the punishment. We recorded the discussion using a microphone and a video camera. In the deliberation done by amateurs, nine subjects (graduate students, age 22 to 24 years, eight males and one female) participated. Then they were divided into three groups and a record was taken from the deliberation of each group.

After collecting the deliberation records, we analyzed the difference between deliberations moderated by the law professionals and by the amateurs. In deliberation, the participants discussed some "topics". A topic is defined as a set of factors corresponding a specific issue. We made a collection of topics of the case. The collection had 10 topics. In analysis, we check what time the participants discussed the topics in deliberation. If the participants discussed a specific topic more than one, we classified the topic as a "multiple discussed topic (MDT)". If the participants did not discuss a specific topic, we classified the topic as a "no discussed topic (NDT)". We counted totals of MDT and NDT.

Table.1(a) shows the result. We checked MDT as $\triangle$, and NDT as $\times$. In addition, table.1(b) shows MDT rate and NDT rate. They are calculated as follow.

Table 1. Deliberation Analysis (Preliminary Research)

(a) Matches of MDT and NDT in Deliberation

Topic No.	Law Professional	Amature
1		
2		
3	△	×
4		△
5		△
6		△
7		△
8		
9		△
10	×	×
Total of MDT	1	5
Total of NDT	1	2

(Legend) △: FMDT ×: FNDT

(b) MDT Rate and NDT Rate

	Law Professional	Amature
MDT Rate	0.100	0.500
NDT Rate	0.100	0.200

$$\text{MDT Rate} = \frac{\text{MDT}}{\text{total of topics}} \tag{1}$$

$$\text{NDT Rate} = \frac{\text{NDT}}{\text{total of topics}} \tag{2}$$

As to the deliberation moderated by the law professional, the moderator avoided dealing with almost all topic more than once. In addition, almost all topics were dealt with. The moderator confirmed factors to be discussed before the deliberation, so that he/she could show the factors one by one and listened to the participants' opinions. By contrast, as to the deliberation moderated by the amateur, the moderator dealt with some topics several times. It is wasteful to moderate deliberation.

From the above, it is important for factor selection to deal with topics only once in deliberation. From now on, using the above knowledge, we show an algorithm to select factors to be discussed, called factor selection algorithm.

Factor Selection Algorithm. This algorithm operates based on the following hypotheses and the result of the preliminary research. First, factors related to many conflict factors are topical in the arguments between the prosecutor and the attorney, so that they are to be discussed first. Among them, factors related to the many support factors are the major and core topics of the case, so that they are to be discussed early. On the other hand, with regard to factors not related to conflict factors, the factors which are supported by fewer factors have less impact on the graph, so that they are easier to handle in the deliberation. Therefore they are picked up early.

Factor selection algorithm is described as below.

- Initial Input
  Graph $G = (V, E)$ consisting of a set of factor node $V \ni v_i, v_i = (nid_i, sug_i, abst_i)$ and a set of edge $E \ni e_j, e_j = (eid_j, vs_j, ve_j, rel_j)$, where we define nid_i, sug_i, $abst_i$, eid_j, vs_j, ve_j, and rel_j. nid_i is the node ID and equal to i. sug_i is the state of the factor, taking "*prose*", "*attor*", and "*other*" as the public prosecutor, the attorney, and the other. $abst_i$ is the level of abstraction, taking "*punish*", "*main*", and "*support*" as penalty factor, main factor, and evidence factor. eid_j is the edge ID and equal to j. vs_j is the head factor node ID of the edge. ve_j is the tail factor node ID of the edge. rel_j is the relationship between vs_j and ve_j, taking "s" and "c" as support and conflict.
- Algorithm
  1. $V_m = \{v_i \in V | \exists i, abst_i = \text{“}main\text{”}\}$ is takenD
  2. The parameters $f_s(v_i)$C$f_a(v_i)$ from each $v_i \in V_m$ are calculated.
     - Overall support factor number $f_s(v_i)$
       $f_s(v_i)$ is the number of the counted factors supporting the intended factor, including $f_s(v_i)$ of the factors supporting the intended factor. It is defined by (3).

$$f_s(v_i) = \begin{cases} 0 & (|V_{v_i}| = 0) \\ 1 + \sum_{v_j \in V_{v_i}} f_s(v_j) & (|V_{v_i}| \geq 1) \end{cases} \tag{3}$$

$$V_{v_i} = \{v_i | \exists i, j, ve_j = v_i, rel_j = \text{“}s\text{”}\} \tag{4}$$

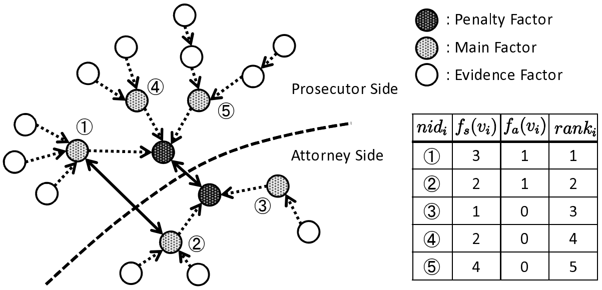
     - Conflict factor number $f_a(v_i)$
       $f_a(v_i)$ is the number of the counted factors conflicting with the intended factor. It is defined by (5).

$$f_a(v_i) = |\{e_j | \exists i, j, rel_j = \text{“}c\text{”}, sug_{vs_j} \neq sug_{ve_j}\}| \tag{5}$$

     Furthermore, the element $rank_i$ is added to each $v_i \in V_m$, so that $v_i = (nid_i, sug_i, abst_i, rank_i)$. Each $rank_i$ will be substituted for either number $1, 2, ..., |V_m|$, which represents argument order.
  3. $1, 2, ..., |V_m|$ is substituted ascending for $rank_i$ of v_i in order of larger $f_a(v_i)$.
     - If $f_a(v_i) = f_a(v_j)$ $(i \neq j)$,
       The factor of larger f_s in v_i and v_j is selected and substituted first.
       - ∗ If $f_s(v_i) = f_s(v_j)$ $(i \neq j)$,
         The order of v_i and v_j is decided randomly.
- Output
  V_m consisting of v_i having $rank_i$

After factors are ordered in the algorithm, time constraint information is also added to them. It is named discussion time. Discussion time $t(v_i)$ is defined by (6) and assigned to each factor.

$$t(v_i) = \frac{T_s}{|V_m|} \tag{6}$$



nid_i	$f_s(v_i)$	$f_a(v_i)$	$rank_i$
①	3	1	1
②	2	1	2
③	1	0	3
④	2	0	4
⑤	4	0	5

Fig. 7. Example of Factor Selection Algorithm Operation

In (6), T_s is the remaining time, the whole time that is available to be spent on the deliberation. The user inputs this time.

Finally the system recommends $v_i \in V_m$ in order by $rank$. Moreover, if the discussion time of each factor is more than $t(v_i)$, the system notifies the user that the recommendation moves on to the next factor.

Fig.7 shows an example of factor selection algorithm operation using an example of the argument structure graph. The table in Fig.7 represents parameters of main factors ①, ②, ③, ④, and ⑤. First, the main factors are taken. Second, the two parameters $f_s(v_i)$, $f_a(v_i)$ for each of the taken factors are calculated. The result of the value is shown in the table in Fig.7. When considered visually, $f_s(v_i)$ is the number of all subsequent nodes that support v_i, including indirect support nodes. On the other hand, $f_a(v_i)$ is the number of nodes that conflict with v_i. After the calculation of the parameters, ① and ② is selected because their $f_a(v_i)$ are larger. Among them, ① is selected because its $f_s(v_i)$ is larger, and 1 is substituted for its $rank_i$. Next, 2 is substituted for ②'s $rank_i$. Then, there are factors ③, ④, and ⑤ which do not relate to conflicting factors. Among them, their $rank_i$ are arranged in ascending order of the lesser $f_s(v_i)$. Hence 1, 2, and 3 are substituted for $rank_i$ of factors ③, ④, and ⑤.

4.5 Next Speaker Recommendation

In deliberation, it occurs that the discussion proceeds while the participants can not make remarks about his/her opinions. This situation should be avoided as much as possible. The presiding judge should give a chance to listen to the opinion of the appropriate participants in this situation. The deliberation process support system has a function to make recommendations to give an opportunity to make remarks to the participants who were previously unable to give their opinions.

The function supervises remark record table $R \in r_{i,j}$, where $r_{i,j}$ is a remark record cell of the participant j related to a factor i. If the user inputs notification to the system the end of argument about factor i', the function confirms the status of remark record table. If $R' = \{r_{i,j} \in R | i = i', r_{i,j} = \text{“”}\}$ is taken, where “$r_{i,j}$ = “”” represents that $r_{i,j}$ has no remark record, the system lets the user confirm the remarks by the participants j corresponding to $r_{i,j} \in R'$. By giving proper opportunities to make remarks in this way, it is possible to have the discussion while listening to not only the participants who make many remarks, but also the ones who do not.

5 Evaluation

We evaluated the proposed system. At first, we evaluated factor selection algorithm. Next, we used the system in actual mock trials moderated by an amateur and evaluated it.

5.1 Evaluation of Factor Selection Algorithm

To verify that factor selection algorithm works effectively, we evaluated the algorithm. The evaluation method was similar to preliminary research. We compared the treatment for topics of the case by factor selection algorithm with one in deliberation moderted by a law professional and one in deliberation moderated by an amateur.

At first, we got the record of a mock trial. This record is different from one used in preliminary research. The record contains the arguments between the prosecutor and the attorney, then we made a collection of topics of the case from the arguments. The collection had 7 topics. On the other hand, we made argument structure graph from the arguments. Next, we input the graph into factor selection algorithm and got the order of factors to be discussed. Then, we checked what time the order of factors dealt with a specific topic in the collection. Meanwhile, we got the deliberation record moderated by a law professional from the record of a mock trial. In addition the deliberation record moderated by an amateur was gotten in a similar way to preliminary research. We check what time the participants discussed the topics in the two deliberations. Finally we counted the totals of MDT and NDT.

Table 2(a) shows the result. In a similar way to preliminary research, we checked MDT as △, and NDT as × . In addition, table 2(b) shows the rate of MDT and NDT.

Factor selection algorithm dealt with more than half of topics once. Therefore the moderator can moderate deliberation smoothly if using the order by factor selection algorithm. By contrast, the moderator of law professional dealt with the specific topics several times. It spent much time. The moderator of amateur did not deal with 3 topics, namely topic 2, 4, and 7. It is not good to moderate deliberation.

Accordingly, factor selection algorithm works well when it is used to select factors to be discussed.

5.2 Evaluation of Deliberation Process Support System

To verify that the system effectively supports the user, we evaluated the system.

Table 2. Deliberation Analysis (Evaluation of Factor Selection Algorithm)

(a) Matches of MDT and NDT in Deliberation

Topic No.	Algorithm	Law Professional	Amature
1		△	
2		△	×
3		△	
4	△	△	×
5	△	△	
6			△
7			×
Total of MDT	2	5	1
Total of NDT	0	0	3

(Legend) △ : FMDT × : FNDT

(b) MDT Rate and NDT Rate

	Algorithm	Law Professional	Amature
MDT Rate	0.286	0.714	0.143
NDT Rate	0.000	0.000	0.429

Table 3. Deliberation Analysis (Evaluation of Deliberation Process Support System)

(a) Matches of MDT and NDT in Deliberation

Topic No.	Group A	Group B
1		
2	×	△
3		
4	△	△
5	△	
6	△	
7		
Total of MDT	3	2
Total of NDT	1	0

(Legend) △ : MDT, × : NDT

(b) MDT Rate and NDT Rate

	Group A	Group B
MDT Rate	0.429	0.286
NDT Rate	0.143	0.000

We held two mock trials moderated by amateurs with the system. The user was not the presiding judge, but an man, who was not participated in the deliberation. The participants were amateurs. They discussed the same case used in evaluation of topic selection algorithm. Argument structure graph of the case was made in advance and used in the deliberations. In this time of the mock trials, they played roles of a presiding judge and 4 citizen judges in a mock trial. In addition, to compare the proposed functions, we made a difference what functions could be used in deliberations. One group (Group A) only used a function of visualization of argument structure in deliberation. The another group (Group B) used it, and additionally used functions of next factor recommendation and next speaker recommendation in deliberation. Evaluation method was same as evaluation of topic selection algorithm. We counted the totals of MDT and NDT. In addition, we carried out a questionnaire survey for the participants. We asked some questions about the system.

The participants could figure out the abstract of the case by argument structure graph. Therefore, group A discussed almost all topics, so that NDTR was low. However, the presiding judge did not avoid dealing with the discussed topics, then MDT rate was high. On the other hand, MDT rate and NDT rate of group B were low. Group B also could use argument structure graph, so that they could discuss all topics. In addition, they discussed the topics in the order created by next topic recommendation. Therefore they could avoid dealing with the discussed topics, then MDT rate was low. The result shows that the system can supports the user to moderate deliberation efficiently. Especially, the user can avoid dealing with the discussed topics and finishing deliberation without dealing with some topics.

The questionnaire survey shows good result. The participants answered some questions by 1-to-7 rating scale (1="I disagree" and 7="I agree"). The average score of the question "remark record table is useful" was 6.8. It suggests that remark record table works efficiently. In free description about remark record table, the participants answered "I could easily figure out whose opinion was lack" and "I could discuss topics as I figured out remarks of mine and others."

6 Conclusion

In this paper, we proposed a deliberation process support system for the presiding judge to facilitate discussion in the deliberations of citizen judge trial. The system uses an argument structure graph based on the arguments between the prosecutor and the attorney. Then the presiding judge uses a factor registration editor. The editor supports the user to register factors and their relations from records of the arguments between the prosecutor and the attorney. After that, it creates a factor list and the argument structure graph. The deliberation process support system uses the graph and supports the user. The system has some functions, such as visualizing the graph, supporting the input of remarks by participants, recommending the topics to be discussed and the participants that are required to make remarks. In addition, we conducted evaluation about the system. Evaluation of factor selection algorithm shows good result. Furthermore, evaluation of the system also shows good result. The evaluation suggests that the system will avoid dealing with the discussed topics and finishing deliberation without dealing with some topics. Therefore the user can moderate deliberation smoothly.

Acknowledgment. The authors thank Prof. S. Hotta from Meiji University for his generous support.

References

1. Nohara, K., Morimoto, I., Mishima, S., Takeuchi, K.: Expansion of Deliberation Design Theory in the Citizen Judge System (Final 3.) Towards Realization of Deliberation Based on Closing Argument Analysis - Concrete Methods and Challenges and Prospects in Designing Deliberation. Horitsujiho 81(10), 84–95 (2009) (in Japanese)
2. Hotta, S.: Linguistic Theories in Legal Settings. Hituzi Syobo (2010) (in Japanese)
3. Reed, C., Rowe, G.: Araucaria: Software for argument analysis, diagramming and representation. International Journal on Artificial Intelligence Tools 13(4), 961–979 (2004)
4. Loukis, E., Xenakis, A., Tseperli, N.: Using Argument Visualization to Enhance e-Participation in the Legislation Formation Process. In: Macintosh, A., Tambouris, E. (eds.) ePart 2009. LNCS, vol. 5694, pp. 125–138. Springer, Heidelberg (2009)
5. Aanzai, K., Nitta, K.: Analyzing system of conference record of citizen judge trial based on annotation. In: Proceedings of the SIG-KBS Conference, vol. 90, pp. 31–36. Japanese Society for Artificial Intelligence (2010)
6. Aleven, V., Ashley, K.D.: Teaching CaseBased Argumentation Through a Model and Examples: Empirical Evaluation of an Intelligent Learning Environment. In: Proceedings of AIED 1997 World Conference, pp. 87–94. IOS Press (1997)
7. Heer, J., Card, S.K., Landay, J.A.: Prefuse: A Toolkit for Interactive Information Visualization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2005, pp. 421–430. ACM, New York (2005)

An Intuitionistic Investigation of Prerequisite-Effect Structure

Katsuhiko Sano, Shingo Hagiwara, and Satoshi Tojo

School of Information Science,
Japan Advanced Institute of Science and Technology,
1–1 Asahidai, Nomi, Ishikawa 923-1292, Japan
`{v-sano,s-hagiwa,tojo}@jaist.ac.jp`

Abstract. This paper studies the prerequisite-effect structures in legal documents from a logical viewpoint. First, we distinguish a written prerequisite-effect structure in legal documents and an application-instance of the written structure to a particular case, and formalize the latter as $\rightsquigarrow$. Second, we specify its semantics as 'immediately after' on intuitionistic Kripke model. Third, we establish semantic properties of $\rightsquigarrow$ including the undefinability of $\rightsquigarrow$ in intuitionistic logic, and provide a cut-free and complete labelled sequent calculus with the intuitionistic logic with $\rightsquigarrow$. Finally, we illustrate a formal representation of Articles 7 and 9 of Japanese Income Tax Act with the help of $\rightsquigarrow$.

1 Introduction

To translate legal documents into a formal language is quite useful; for example if they can be represented in XML format, they can be utilized electronically on the web. Furthermore, if they could be translated into logic, they could be applied to electronic law consulting system and automatic deduction. This kind of attempt has a long history and was actively studied as so-called expert systems. However, one of the most salient problems of the translation concerns if-then structure in legal documents (for the rich variety of normative conditionals, see, e.g. [1, Sec.2] or [2]). Since these if-then structures have various meanings, we cannot translate them into one logical implication in predicate or first-order logic uniformly.

Law, or so-called 'normative knowledge,' is written in the prerequisite-effect structure (PE-structure, in short). PE-structure plays the most important role in legal decision under incomplete information, as has been intensively studied in law community, where PE-structure is mentioned as 'the theory of presupposed ultimate facts' [3]. Our objective in this paper lies also on such *incomplete* but gradually accumulating knowledge. Roughly speaking, the prerequisite corresponds to 'if'-part, and the conditions for the application of the normative knowledge are written here. The effect corresponds to 'then'-part, and the expected result for the application of the law is mentioned. For example, in the following statement:

> *Income Tax Act: Article 7* (*Scope of Taxable Income*)
> Income tax shall be imposed with respect to income specified in each of

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 170–183, 2012.

> the following items for the category of person listed in the relevant item:
> (i) A resident other than a non-permanent resident: All income

the law is to be applied to a resident other than a non-permanent resident ('if'-part), and the effect would be imposition of taxes with respect to his/her all income ('then'-part). Therefore, our intended if-then structure here is: If x is a permanent Japanese, then income tax shall be imposed with respect to all income of x. In order to construct a legal reasoning system as our ultimate objective, this paper focuses on a formalization of this prerequisite-effect structure (cf. [4]).

We proceed as follows. Section 2 introduces a distinction between static and dynamic prerequisite structures, and chooses the dynamic one as our target of formalization. Section 2 also specifies an underlying semantic idea of the dynamic prerequisite-effect structure over intuitionistic Kripke semantics and compares it with an approach by Priorian temporal logic. Section 3 introduces the expanded syntax of intuitionistic (propositional) logic with a new connective $\rightsquigarrow$ for the dynamic prerequisite structure and give a rigorous formulation of Kripke semantics. Then, we also review some semantic properties of $\rightsquigarrow$. Section 4 provides a cut-free and complete labelled sequent calculus with the intuitionistic logic with $\rightsquigarrow$ (this axiomatization problem was one of the open problems of the earlier version of this paper [5], which was asked by one of the reviewers of [5]). Section 5 illustrates a formal representation of Articles 7 and 9 of Japanese Income Tax Act with the help of $\rightsquigarrow$.

2 A Formalization of Prerequisite-Effect Structure

2.1 Static and Dynamic PE-Structures

Here, we would like to introduce a distinction between *static* and *dynamic* PE-structures as our own terminology. By *static PE-structure*, we mean the written or static description of law, which can be applied to all the relevant events e. For example, the description of Income Tax Act in the introduction can be regarded as a static PE-structure. Let us rewrite the article above as follows:

$$\forall\, e : \text{event}.\, \forall\, x : \text{individual}.\, (\text{if } P(e,x) \text{ then } I(e,x)), \tag{1}$$

where $P(e,x)$ and $I(e,x)$ correspond to the prerequisite and the effect part, respectively. On the other hand, when we would like to apply the above static Article 7 (i) to some particular case or event e and some particular individual x, we need to instantiate the static structure to e and x. We call such an instance of the static structure the *dynamic PE-structure*. Suppose that Koichi earned one hundred and ten million yen in 2002. Let us denote this event by e_1. Then, we obtain

$$\text{if } P(e_1, \mathsf{Koichi}) \text{ then } I(e_1, \mathsf{Koichi}). \tag{2}$$

One important difference between static and dynamic PE-structures is that a dynamic PE-structure, derived from the static one, is open to add more parameters than the static one, e.g., time, space, temperature, etc. In particular, when we need to apply the specific law to some particular case and individual, however, we cannot avoid using temporal parameter in dynamic PE-structures, because the plaintiff and defendant contest, *within time*, each other for the confirmation of $P(e_1, \mathsf{Koichi})$, i.e., when the prerequisite will be confirmed or proved. If we take a more radical view, we could regard a process of accumulating the confirmed facts itself as a flow of time. Note that we do not always assume such an explicit reference to time in static PE-structures in general. One way to realize this aspect in (2) is to add the temporal parameter t as a new argument of $P(e_1, \mathsf{Koichi})$ and $I(e_1, \mathsf{Koichi})$, i.e., we add one more sort to our intended syntax for dynamic PE-structures. This is a plausible option but this makes our syntax complicated. Instead of this, however, we would like to keep the same number of sorts as in the static PE-structures (i.e., keep the simple syntax), but introduce a new connective $\leadsto$ to reflect the additional temporal dimension of the dynamic PE-structure into its *semantics* (we will specify its semantics later):

$$P(e_1, \mathsf{Koichi}) \leadsto I(e_1, \mathsf{Koichi}). \tag{3}$$

In what follows in this paper, we concentrate on the formalization and its semantics of the *dynamic* PE-structures and we simply regard $P(e_1, \mathsf{Koichi})$ and $I(e_1, \mathsf{Koichi})$ as proposition letters p, q, etc.

2.2 A Semantic Idea for Dynamic PE-Structure

Now, we specify how we should give a semantics to the dynamic prerequisite-effect structure $A \leadsto B$. Our guiding intuition is the following: the dynamic PE-structure would be defined in the augmentation of known facts (or, evidences); no sooner the knowledge increases to the point the prerequisite is confirmed, than the effect follows. Therefore, our requirement to the semantics of $A \leadsto B$ can be summarized as the following two items:

(R1). Set up the semantic structures that can reflect the accumulation of confirmed facts.

(R2). Read '$A \leadsto B$' as 'immediately after A is confirmed, B will become effective'.

Remark that we should distinguish, e.g., the actual time t_1 (say, 2002) that Koichi is a permanent Japanese from the time t_2 that it is confirmed, in the court, that Koichi is a permanent Japanese at t_1. In general, we can assume that $t_1 < t_2$, but, when we formalize the dynamic PE-structures, we are always concerned with t_2. Similarly, we should also distinguish between the time t_3 that the effect of Article 7 of Income Tax Act becomes effective and the time t_4 that Koichi actually pays the tax imposed to all his income (say, at 2002). In the formalization, we focus on t_3, which is immediately after t_2. Therefore, our intended temporal information in this example is *not* (t_1, t_4) but (t_2, t_3). Otherwise, our semantic idea 'immediately after' for $\varphi \leadsto \psi$ is meaningless. In this sense, we can say that our model is concerned with 'time in the court'.

Then, we cannot formalize '$A \leadsto B$' as the *material implication* of classical logic, because it does not satisfy our requirement (R2), i.e., it allows the possibility that B hold at the same time as A is confirmed and we should prohibit such a possibility. It is also obvious that the truth table semantics for the material implication does not satisfy the requirement (R1).

Classical logic assumes only one single actual world and all the propositions must refer to the same time. If we allow the multiplicity of world and set up the accessibility relation between possible worlds as a transitive and reflexive relation, we can obtain a *Kripke frame* for intuitionistic logic (if the reader is unfamiliar with intuitionistic logic, he/she can refer to [6, Ch.5]). As for the truth (or, valuation) of proposition letters, we add one restriction called *hereditary condition* or *persistence condition*, which means that, once p becomes true at any given world w, p will continue to be true at any worlds accessible from w. We define that a *Kripke model* is a pair of Kripke frame and a valuation satisfying hereditary condition. Here, we model a change of knowledge states as a sequence of temporal states; thus, instead of physical time parameter, we employ *intuitionistic* Kripke model. Then, our requirement (R1) is satisfied by the hereditary condition. Moreover, the notion of '*immediately after*' implies, over intuitionistic Kripke model, that there cannot be any other change in the world, i.e., there is no extraneous information besides the prerequisite and the effect, to be free from the annoying *frame problem.*

In Kripke models for intuitionistic logic, the truth condition of intuitionistic implication $A \to B$ is defined as follows:

$$w \models A \to B \iff \forall\, w' \in W.\ ((w \leq w' \text{ and } w' \models A) \text{ implies } w' \models B)$$

where '$\leq$' is a transitive and reflexive accessibility relation on possible worlds W. Can we represent our $A \leadsto B$ as the intuitionistic implication $A \to B$? The answer is negative. Suppose that we could formalize '$A \leadsto B$' as the intuitionistic $A \leadsto B$. Then, we may consider the situation that B becomes effective, e.g., *two years after* when the prerequisite A is proved (recall that we concentrate on the pair (t_2, t_3) rather than (t_1, t_4)). We, however, should exclude such a situation in the case of the dynamic PE-structures by our requirement (R2).

These considerations lead us to the following semantic formulation of '$A \leadsto B$' over intuitionistic Kripke model.

Definition 1 (Dynamic PE-structure). *Given any Kripke model* $(W, \leq, V)$ *for intuitionistic logic, a dynamic PE-structure* $A \leadsto B$ *holds at* w *iff, for any future state* w' *of* w*, if* w' *is the first state satisfying the prerequisite* A*, then the effect* B *does not hold at* w' *but* B *hold immediately after* w'*, i.e.,*

$$\forall\, w' \geq w.\ \Big((w' \models A \textit{ and } \forall\, z < w'.\, z \not\models A) \textit{ implies } (w' \not\models B \textit{ and } \forall\, y > w'.\, y \models B)\Big),$$

where $w' \geq w$ *means* $w \leq w'$*,* $x < y$ *is defined as* $x \leq y$ *and* $x \neq y$*, and* $y > x$ *means* $x < y$*, respectively.*

We emphasize again that the semantics of $A \leadsto B$ does not allow any change other than B after when A is confirmed. Therefore, we can regard the confirmation of A as a direct 'cause' of the effect B in $A \leadsto B$.

2.3 Why Do We Employ Intuitionistic Kripke Model?

In this subsection, we explain one reason why we use the notion of intuitionistic Kripke model, and then give an answer to the rejection of (R1) within Priorian temporal logic [7].

First, one of our reasons to prefer intuitionistic Kripke model is to reject the law of excluded middle (LEM): $A \vee \neg A$. Even if we discuss about the confirmation of the prerequisite p in the court, it might be difficult to confirm p and also uneasy to *dis*confirm p, because we have some indirect evidences alone. This means that there is no need for our aim to keep $A \vee \neg A$ as a logical validity. As is well-known, we can invalidate LEM in terms of intuitionistic Kripke model (it suffices to consider two points linear Kripke model such that p is true only at the top element [6, p.165]).

Let us move to the second point. An alternative to realize (R2) is to employ Priorian temporal logic. In [5, Sec.2.3], we have already shown how to formalize the dynamic PE-structures in terms of Priorian temporal logic. Our formalization was $(\mathrm{H} \sim A \wedge (\mathrm{G}A \wedge A)) \supset ((\mathrm{H} \sim B \wedge \sim A) \wedge \mathrm{G}B)$, where G and H mean 'all the future' and 'all the past', respectively, $\supset$ is the material implication of classical logic, and $\sim$ is the classical negation. However, one might reject (R1) by the following reason (note that we can reject (R1) within Priorian temporal logic): Suppose that it was confirmed that p (Koichi is a permanent Japanese) in the court at a moment t_0 and so the effect of q (imposition of tax to all income of Koichi) became effective at the next moment t_1 of t_0. However, suppose also that the effect q was *cancelled* at some future moment t_2 of t_1, because of another prerequisite-effect structure and new evidences. Therefore, the accumulation of the confirmed facts fails.

Let us give a reply to this rejection of (R1). First of all, we should emphasize that the cancellation does not imply that we could change the past and that the confirmation of the effect q at t_1 itself still holds at t_2. What the cancellation of q at t_2 does is to *regard or presume from the future perspective of* t_2 that the effect q was *not* confirmed at t_1 in the court. This does not contradict the confirmation of q at t_1 at all. Moreover, the cancellation of t_2 itself can be seen as another confirmed fact, and so, the cancellation of q at t_2 is also true at any later moment than t_2.

In the actual lawsuits, some decisions in the past may be overturned. In such cases, we track back to some past and can regard that the time axis bifurcates into two different futures from that point; one of them reflects the real situation while the other may be virtual rendering. This branching time is exactly modeled by hereditary Kripke semantics.

3 Dynamic Prerequisite-Effect Structure on Intuitionistic Kripke Model

3.1 Syntax and Semantics

Let us introduce our syntax handling the dynamic PE-structure over intuitionistic Kripke models. Our vocabulary $\mathcal{L}_{\mathrm{PE}}$ consists of a (countable) set Prop of proposition letters, the symbol $\rightsquigarrow$ for the dynamic PE-structure, and the logical connectives of intuitionistic propositional logic (i.e., $\bot$, $\vee$, $\wedge$, and $\to$). We denote by $\mathcal{L}_{\mathrm{IL}}$ the result of dropping $\rightsquigarrow$ from the vocabulary $\mathcal{L}_{\mathrm{PE}}$. Then, the set of formulas of $\mathcal{L}_{\mathrm{PE}}$ are defined inductively as:

$$A ::= p \mid \bot \mid A \vee B \mid A \wedge B \mid A \to B \mid A \rightsquigarrow B,$$

where $p \in \mathsf{Prop}$. We define $\neg A := A \to \bot$. Let us move to the semantics of $\mathcal{L}_{\mathrm{PE}}$. Given any Kripke frame $(W, \leq)$ (recall that $\leq$ is reflexive and transitive at least), $V : \mathsf{Prop} \to \mathcal{P}(W)$ is a *valuation* if it satisfies the *hereditary condition*: $w \leq w'$ and $w \in V(p)$ jointly imply $w' \in V(p)$ (i.e., $V(p)$ is upward closed with respect to $\leq$), for any $w, w' \in W$ and any $p \in \mathsf{Prop}$. A *Kripke model* (written: $\mathfrak{M}$) is a pair of Kripke frame and a valuation. Given any Kripke model $\mathfrak{M} = (W, \leq, V)$, any $w \in W$, and any formula A, we define the satisfaction relation $w \models A$ as follows (recall Definition 1 for the dynamic PE-structure $A \rightsquigarrow B$):

$$\begin{array}{lcl}
w \models p & \iff & w \in V(p) \\
w \models \bot & & \text{Never} \\
w \models A \vee B & \iff & w \models A \text{ or } w \models B \\
w \models A \wedge B & \iff & w \models A \text{ and } w \models B \\
w \models A \to B & \iff & \forall w' \geq w.\, (w' \models A \text{ implies } w' \models B) \\
w \models A \rightsquigarrow B & \iff & \forall w' \geq w.\, \Big((w' \models A \text{ and } \forall z < w'.\, z \not\models A) \text{ implies} \\
& & \qquad\qquad\qquad (w' \not\models B \text{ and } \forall y > w'.\, y \models B)\Big)
\end{array}$$

Define the *semantic consequence relation* $\{ A_i \mid i \in I \} \models B$ as follows: for any Kripke model $\mathfrak{M}$ and any w in $\mathfrak{M}$, $w \models A_i$ $(i \in I)$ implies $w \models B$. In what follows, we write $A \models B$ instead of $\{ A \} \models B$.

3.2 Some Semantic Properties

In this subsection, we review some semantic properties of $\rightsquigarrow$ from [5] (all the proofs except Proposition 4 (vii) and (viii) can be found in [5]). Some of the reader might consider that we can express $\rightsquigarrow$ by combining some connectives of $\mathcal{L}_{\mathrm{IL}}$. However, this is impossible.

Proposition 1. *$\rightsquigarrow$ is undefinable in the syntax $\mathcal{L}_{\mathrm{IL}}$ of intuitionistic logic.*

Therefore, we have expanded the syntax $\mathcal{L}_{\mathrm{IL}}$ of intuitionistic logic with an additional symbol $\rightsquigarrow$. Our addition of $\rightsquigarrow$ does not break the following *hereditary condition* over Kripke models.

Proposition 2. *Let $(W, \leq, V)$ be a Kripke model for intuitionistic logic. For any formula A of $\mathcal{L}_{\mathrm{PE}}$, if $w \leq u$ and $w \models A$, then $u \models A$.*

This hereditary condition implies the *reverse hereditary condition*, i.e., $w \leq v$ and $v \not\models A$ implies $w \not\models A$. By this, we can demonstrate that our formalization $A \rightsquigarrow B$ of the dynamic PE-structures can exclude the situation that the effect B holds *before* the prerequisite A is confirmed, since we prohibit, in $w \models A \rightsquigarrow B$, the possibility that the effect B hold at the same time as the prerequisite A is confirmed.

Proposition 3. *Let* $(W, \leq, V)$ *be a Kripke model for intuitionistic logic. If* $w \models A \rightsquigarrow B$*, then, for any future state* w' *of* w*, if* w' *is the first state satisfying the prerequisite* A*, then the effect* B *does not hold in any past state of* w'*, i.e.,*

$$\forall w' \geq w. \Big((w' \models A \text{ and } \forall z < w'.\, z \not\models A) \text{ implies } (\forall y \leq w'.\, y \not\models B)\Big).$$

We have the following logical properties about $\rightsquigarrow$ [1].

Proposition 4. *(i)* $(A_1 \rightsquigarrow B) \wedge (A_2 \rightsquigarrow B) \models (A_1 \vee A_2) \rightsquigarrow B$.
(ii) $((A_1 \vee A_2) \rightsquigarrow B) \wedge \neg A_1 \models A_2 \rightsquigarrow B$.
(iii) $(A \rightsquigarrow B_1) \wedge (A \rightsquigarrow B_2) \models A \rightsquigarrow (B_1 \wedge B_2)$.
(iv) $(A \rightsquigarrow (B_1 \vee B_2)) \wedge \neg B_1 \models A \rightsquigarrow B_2$.
(v) $(A \rightsquigarrow (B_1 \vee B_2)) \wedge (A \rightarrow \neg B_1) \models A \rightsquigarrow B_2$.
(vi) $(A \rightsquigarrow (B \rightarrow C)) \models A \rightsquigarrow C$.
(vii) $\neg A \models A \rightsquigarrow B$.
(viii) $A \wedge (B \rightsquigarrow C) \models (A \wedge B) \rightsquigarrow C$.

Proof. We only give proofs of (ii) and (vii). First, let us prove (ii). Assume that $w \models (A_1 \vee A_2) \rightsquigarrow B$ and $w \models \neg A_1$. Consider any $w' \geq w$ with $w' \models A_2$ and $\forall y < w'.\, y \not\models A_2$. We show that w' is the first state satisfying $A_1 \vee A_2$. It is clear that $w' \models A_1 \vee A_2$. So, let us establish $\forall y < w'.\, y \not\models A_1 \vee A_2$. Fix any $y < w'$. We can assume that $w \leq y$, because $w \not\models A_1 \vee A_2$ implies $y' \models A_1 \vee A_2$ for any $y' < w$ by the reverse hereditary condition. Trivially, $y \not\models A_2$. Moreover, we deduce from $w \models \neg A_1$ that $y \not\models A_1$, which implies $y \not\models A_1 \vee A_2$. We have shown that w' is the first state satisfying $A_1 \vee A_2$. Then, we can demonstrate the desired conclusion by $w \models (A_1 \vee A_2) \rightsquigarrow B$.

Second, let us establish (vii) Assume that $w \models \neg A$. Consider any $w' \geq w$ such that $w' \models A$ and $\forall y < w'.\, y \not\models A$. By assumption, we get $w \not\models A$, and so, we can obtain the conclusion trivially. □

In Section 5, we use (ii) of Proposition 4 to illustrate our example of Japanese Income Tax Act.

Remark 1. One might think that $\neg A \models A \rightsquigarrow B$ shows that $\rightsquigarrow$ is not appropriate to analyze the dynamic PE-structures, since this seems to mean that, if something (A) has been confirmed in the court, then the opposite ($\neg A$) causes anything else (B). We, however, claim that such derived $A \rightsquigarrow B$ is a *trivial* dynamic PE-structure. Since A has been already confirmed, it is impossible to find a first future state satisfying A. This means that an arbitrary B *never* becomes effective. Moreover, we can regard any given law as a set of *non-trivial* dynamic PE-structures. Our intention is to use our $\rightsquigarrow$ to calculate effects from the law. In this sense, $\neg A \models A \rightsquigarrow B$ is harmless for our aim.

Proposition 5. *(i)* $\not\models A \rightsquigarrow A$.
(ii) $A \rightarrow B \not\models A \rightsquigarrow B$.

---

[1] (vii) was pointed out by one of the reviewers.

(iii) $A \rightsquigarrow B \not\models A \to B$.
(iv) $(A \rightsquigarrow B) \wedge (B \rightsquigarrow C) \not\models A \rightsquigarrow C$.
(v) $A \rightsquigarrow B \not\models (A \wedge A') \rightsquigarrow B$.
(vi) $(A_1 \vee A_2) \rightsquigarrow B \not\models (A_1 \rightsquigarrow B) \wedge (A_2 \rightsquigarrow B)$.
(vii) $A \rightsquigarrow (B_1 \wedge B_2) \not\models (A \rightsquigarrow B_1) \wedge (A \rightsquigarrow B_2)$.

4 Tree-Sequent Calculus for Dynamic PE-Structures

In this section, we provide a tree-sequent calculus [8, 9, 10], a variant of labelled sequent calculus, with our intuitionistic logic with $\rightsquigarrow$.

Let us introduce some terminology. A *label* is a finite sequence $\langle n_1, ..., n_l \rangle$ of natural numbers. We use letters α, β, etc. for labels and denote the empty sequence $\langle\, \rangle$ by ε. Given any label $\alpha = \langle n_1, ..., n_l \rangle$ and any natural number m, $\alpha \cdot m$ means $\langle n_1, ..., n_l, m \rangle$, i.e., the concatenation of α with m. β is a *child* of α (notation: $\alpha \prec \beta$) if $\beta = \alpha \cdot m$ for some natural number m. β is a *descendant* of α if there are finite $\alpha_1, \ldots, \alpha_k$ such that $\alpha \prec \alpha_1 \prec \cdots \prec \alpha_n \prec \beta$. A *tree* is a set $\mathcal{T}$ of labels such that $\varepsilon \in \mathcal{T}$ and for each $\alpha \cdot m \in \mathcal{T}$, $\alpha \in \mathcal{T}$. A *labelled formula* $\alpha : A$ is a pair of a label α and a formula A of $\mathcal{L}_{\mathrm{PE}}$. A *tree-sequent* is an expression $\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$ where Γ and Δ are finite set of labelled formulas, $\mathcal{T}$ is a *finite* tree, and each label in Γ and Δ is an element of $\mathcal{T}$.

Given any tree sequent $\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$, we can associate a sequent $\Gamma_\alpha \Rightarrow \Delta_\alpha$ with any label α of $\mathcal{T}$, where Γ_α (or, Δ_α) is a set of formulas A such that $\alpha : A \in \Gamma$ (or, $\alpha : A \in \Delta$, respectively). In this sense, a tree sequent can be displayed as structured ordinary sequents.

Now let us introduce the tree-sequent calculus **TPE**. This system defines inference schemes which allow us to manipulate tree-sequents. The axioms of **TPE** are of the following forms:

$$\alpha : A, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta, \alpha : A \quad (\mathrm{Ax}) \qquad \alpha : \bot, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta \quad (\bot)$$

The inference rules of **TPE** are the following:

$$\frac{\beta : A, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta}{\alpha : A, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta}\ (\text{Move}) \text{ where } \alpha \prec \beta \qquad \frac{\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta, \alpha : A \quad \alpha : A, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta}{\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta}\ (\text{Cut})$$

$$\frac{\alpha : A, \alpha : B, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta}{\alpha : A \wedge B, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta}\ (\wedge L) \qquad \frac{\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta, \alpha : A \quad \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta, \alpha : B}{\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta, \alpha : A \wedge B}\ (\wedge R)$$

$$\frac{\alpha : A, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta \quad \alpha : B, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta}{\alpha : A \vee B, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta}\ (\vee L) \qquad \frac{\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta, \alpha : A, \alpha : B}{\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta, \alpha : A \vee B}\ (\vee R)$$

$$\frac{\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta, \alpha : A \quad \alpha : B, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta}{\alpha : A \to B, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta}\ (\to L) \qquad \frac{\alpha \cdot n : A, \Gamma \overset{\mathcal{T} \cup \{\alpha \cdot n\}}{\Rightarrow} \Delta, \alpha \cdot n : B}{\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta, \alpha : A \to B}\ (\to R)$$

where $\alpha \cdot n$ does not occur in $\mathcal{T}$.

$$\frac{\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta, \alpha : A \quad \beta_1 : A, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta \quad \cdots \quad \beta_n : A, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta \quad \gamma_1 : B, \ldots, \gamma_p : B, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta, \alpha : B}{\alpha : A \rightsquigarrow B, \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta} \ (\rightsquigarrow L)$$

where $\varepsilon = \beta_1 \prec \cdots \prec \beta_n \prec \alpha$ and $\gamma_1, \ldots, \gamma_p$ are all the children of α in $\mathcal{T}$.

$$\frac{\alpha \cdot n : B, \alpha \cdot n : A, \Gamma \overset{\mathcal{T} \cup \{\alpha \cdot n\}}{\Rightarrow} \Delta, \alpha : A \quad \alpha \cdot n : A, \Gamma \overset{\mathcal{T} \cup \{\alpha \cdot n, \alpha \cdot n \cdot m\}}{\Rightarrow} \Delta, \alpha : A, \alpha \cdot n \cdot m : B}{\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta, \alpha : A \rightsquigarrow B} \ (\rightsquigarrow R)$$

where $\alpha \cdot n$ and $\alpha \cdot n \cdot m$ do not occur in $\mathcal{T}$.

The tree-sequent calculus $\mathbf{TPE}^-$ is obtained by dropping (Cut) from $\mathbf{TPE}$. Whenever a tree-sequent $\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$ is provable in $\mathbf{TPE}$ or $\mathbf{TPE}^-$, we write $\mathbf{TPE} \vdash \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$ or $\mathbf{TPE}^- \vdash \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$, respectively.

Example 1. The tree sequent $\varepsilon : A_1 \vee A_2 \rightsquigarrow B, \varepsilon : \neg A_1 \overset{\{\varepsilon\}}{\Rightarrow} \varepsilon : A_2 \rightsquigarrow B$ is provable in $\mathbf{TPE}^-$. Let us give a sketch of the derivation. A basic idea is to rewrite the proof of Proposition 4 (ii) proof-theoretically. Remark that

$$\begin{array}{c} \overset{\mathcal{T}}{\Rightarrow} \Delta, \alpha : A \\ \hline\hline \alpha : \neg A, \overset{\mathcal{T}}{\Rightarrow} \Delta \end{array} \ (\neg L)$$

is a derived inference rule of $\mathbf{TPE}^-$.

$$\frac{\overset{\mathcal{D}_1}{1 : B, 1 : A_2, \varepsilon : A_1 \vee A_2 \rightsquigarrow B, \varepsilon : \neg A_1 \overset{\{\varepsilon, 1\}}{\Rightarrow} \varepsilon : A_2} \quad \overset{\mathcal{D}_2}{1 : A_2, \varepsilon : A_1 \vee A_2 \rightsquigarrow B, \varepsilon : \neg A_1 \overset{\{\varepsilon, 1, \langle 1, 1 \rangle\}}{\Rightarrow} \varepsilon : A_2, \langle 1, 1 \rangle : B}}{\varepsilon : A_1 \vee A_2 \rightsquigarrow B, \varepsilon : \neg A_1 \overset{\{\varepsilon\}}{\Rightarrow} \varepsilon : A_2 \rightsquigarrow B} \ (\rightsquigarrow R)$$

where we concentrate only on $\mathcal{D}_1 :=$

$$\begin{array}{c} \dfrac{\dfrac{\dfrac{1 : B, 1 : A_2 \overset{\{\varepsilon, 1\}}{\Rightarrow} \varepsilon : A_2, \varepsilon : A_1, 1 : A_1, 1 : A_2}{1 : B, 1 : A_2 \overset{\{\varepsilon, 1\}}{\Rightarrow} \varepsilon : A_2, \varepsilon : A_1, 1 : A_1 \vee A_2} \ (\vee R) \qquad \mathcal{D}_3 \quad 1 : B, 1 : A_2 \overset{\{\varepsilon, 1\}}{\Rightarrow} \varepsilon : A_2, \varepsilon : A_1, 1 : B}{1 : B, 1 : A_2, 1 : A_1 \vee A_2 \rightsquigarrow B, \overset{\{\varepsilon, 1\}}{\Rightarrow} \varepsilon : A_2, \varepsilon : A_1} \ (\rightsquigarrow L)}{1 : B, 1 : A_2, \varepsilon : A_1 \vee A_2 \rightsquigarrow B, \overset{\{\varepsilon, 1\}}{\Rightarrow} \varepsilon : A_2, \varepsilon : A_1} \ (\mathrm{Move}) \\ \hline\hline 1 : B, 1 : A_2, \varepsilon : A_1 \vee A_2 \rightsquigarrow B, \varepsilon : \neg A_1 \overset{\{\varepsilon, 1\}}{\Rightarrow} \varepsilon : A_2 \end{array} \ (\neg L)$$

and $\mathcal{D}_3$ is:

$$\frac{\varepsilon : A_1, 1 : B, 1 : A_2 \overset{\{\varepsilon, 1\}}{\Rightarrow} \varepsilon : A_2, \varepsilon : A_1 \quad \varepsilon : A_2, 1 : B, 1 : A_2 \overset{\{\varepsilon, 1\}}{\Rightarrow} \varepsilon : A_2, \varepsilon : A_1}{\varepsilon : A_1 \vee A_2, 1 : B, 1 : A_2 \overset{\{\varepsilon, 1\}}{\Rightarrow} \varepsilon : A_2, \varepsilon : A_1} \ (\vee L)$$

Theorem 1. *If* $\mathbf{TPE} \vdash \varepsilon : A_1, \ldots, \varepsilon : A_n \overset{\{\varepsilon\}}{\Rightarrow} \varepsilon : B$, *then* $\{A_1, \ldots, A_n\} \models B$.

Proof. Let us introduce one terminology. Given any $\mathfrak{M} = (W, \leq, V)$ and any tree-sequent $\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$, $\mathfrak{M}$ is *faithful* to $\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$ if there is a $f : \mathcal{T} \to W$ such that if $\alpha \prec \beta$ then $f(\alpha) \leq f(\beta)$; if $\alpha : A \in \Gamma$ then $f(\alpha) \models A$; if $\alpha : A \in \Delta$ then $f(\alpha) \not\models A$. Given any $\mathfrak{M}$, it is easy to observe that the axioms of TPE are not faithful to $\mathfrak{M}$ and that each inference rule of TPE preserves non-faithfulness to $\mathfrak{M}$, i.e., if the conclusion of the rule is faithful to $\mathfrak{M}$, then one of the premises is faithful to $\mathfrak{M}$ (if we regard our system as a tableau system, i.e., we read each of the rules from bottom to top, then it should preserve faithfulness. However, since we are concerned with a sequent-calculus formulation and read each rule from top to bottom, we need to consider the preservation of *non*-faithfulness).

We establish the soundness of TPE as follows: Assume that $\varepsilon : A_1, \ldots, \varepsilon : A_n \overset{\{\varepsilon\}}{\Rightarrow} \varepsilon : B$ is derivable in TPE. Suppose for contradiction that $w \models A_i$ $(1 \leq i \leq n)$ and $w \not\models B$ for some $\mathfrak{M}$ and some w in $\mathfrak{M}$. By assumption and the observation above, $\varepsilon : A_1, \ldots, \varepsilon : A_n \overset{\{\varepsilon\}}{\Rightarrow} \varepsilon : B$ is not faithful to $\mathfrak{M}$. However, it is clear that $\varepsilon : A_1, \ldots, \varepsilon : A_n \overset{\{\varepsilon\}}{\Rightarrow} \varepsilon : B$ is faithful to $\mathfrak{M}$, a contradiction. □

Theorem 2. *If* $\{ A_1, \ldots, A_n \} \models B$, *then* $\mathsf{TPE}^- \vdash \varepsilon : A_1, \ldots, \varepsilon : A_n \overset{\{\varepsilon\}}{\Rightarrow} \varepsilon : B$.

An outline of the proof of this theorem can be found in Appendix A. By Theorem 1 and Theorem 2, we obtain the following.

Corollary 1. *The following are all equivalent:* (i) $\mathsf{TPE} \vdash \varepsilon : A_1, \ldots, \varepsilon : A_n \overset{\{\varepsilon\}}{\Rightarrow} \varepsilon : B$, (ii) $\mathsf{TPE}^- \vdash \varepsilon : A_1, \ldots, \varepsilon : A_n \overset{\{\varepsilon\}}{\Rightarrow} \varepsilon : B$, (iii) $\{ A_1, \ldots, A_n \} \models B$.

This corollary tells us that TPE enjoys the following cut elimination theorem: $\mathsf{TPE} \vdash \varepsilon : A_1, \ldots, \varepsilon : A_n \overset{\{\varepsilon\}}{\Rightarrow} \varepsilon : B$ implies $\mathsf{TPE}^- \vdash \varepsilon : A_1, \ldots, \varepsilon : A_n \overset{\{\varepsilon\}}{\Rightarrow} \varepsilon : B$.

5 A Formal Representation of Articles 7 and 9 of Japanese Income Tax Act

In our motivating example, we apply Article 7 (i) of Income Tax Act to the following specific event e_1: Koichi earned one hundred and ten million yen in 2002. Then, our dynamic PE-structure is $p \rightsquigarrow q$, where

p : Koichi is a permanent Japanese,
q : Income Tax shall be imposed to all income of Koichi.

Let us add one more assumption to our story: Koichi won a Nobel Prize in 2002 (this is why he earned so much money in this year). Does he need to pay income tax to the money obtained from Nobel Foundation as Nobel Prize? The answer is negative, since we can find the following description in Article 9 (xiii) of Japanese Income Tax Act:

> *Income Tax Act: Article 9* (xiii)
> Income tax shall not be imposed with respect to money and/or goods delivered as Nobel Prize from Nobel Foundation.

If we prepare the following propositions:

a : Koichi obtained money from Nobel Foundation;
b : Koichi obtained money as Nobel Prize;
$\neg p$: Koichi is not a permanent Japanese,

then we can obtain another dynamic PE-structure of Article 9 (xiii):

$$(a \wedge b) \rightsquigarrow \neg p.$$

Then, our dynamic PE-structure $p \rightsquigarrow q$ for Article 7 (i) is not adequate if we take such exception $(a \wedge b) \rightsquigarrow \neg p$ into consideration. Let us explain this point. Suppose that we have already confirmed $\Theta_1 = \{ a, b, p \}$ at the same time. Immediately after when p is confirmed, the effect of q appears. On the other hand, immediately after when both a and b are confirmed, the effect of $\neg p$ appears. However, we do not want to have the effects of both p and $\neg p$.

This argument means that we need to add to the prerequisite of $p \rightsquigarrow q$ an additional clause to avoid such an unintended case. For example, our additional prerequisite will deal with the cases: i) Koichi did not obtain money from Nobel Foundation; ii) Koichi obtained money from Nobel Foundation *not as* Nobel Prize (e.g., Koichi got money from the foundation because he contributed to select the candidates of Nobel Prize). We can formalize these cases by $\neg a \vee (a \wedge \neg b)$. Then, we can obtain our new formalization of Article 7 (i) in relationship with Article 9 (xiii) as: $(p \wedge (\neg a \vee (a \wedge \neg b))) \rightsquigarrow q$. By distributivity of $\wedge$ over $\vee$, this is equivalent to:

$$((p \wedge \neg a) \vee (p \wedge a \wedge \neg b)) \rightsquigarrow q$$

Now, we can obtain our desired formalization of Article 7 (i) of Income Tax Act:

$$\Gamma_2 := \{ ((p \wedge \neg a) \vee (p \wedge a \wedge \neg b)) \rightsquigarrow q, (a \wedge b) \rightsquigarrow \neg p \}.$$

Then, when we have already confirmed $\Theta_1 = \{ p, a, b \}$, the effect derived from Γ_2 and Θ_1 becomes q alone.

Let us consider a different scenario: we have confirmed $\Theta_2 = \{ p, \neg a \}$. It is clear that $\neg a \to \neg(p \wedge a \wedge \neg b)$ is a theorem of intuitionistic logic. Thus, $\Theta_2 \models \neg(p \wedge a \wedge \neg b)$. By Proposition 4 (ii), we can get:

$$\Gamma_2 \cup \Theta_2 \models (p \wedge \neg a) \rightsquigarrow q,$$

and so, the effect derived from Γ_2 and Θ_2 is to be q.

We admit that such exceptional precondition should be treated in *default logic* [11], or other non-monotonic reasoning formalisms (e.g. [12]), however, in this paper, we simply append such exceptions to prerequisite to avoid logical complication.

6 Concluding Remarks

6.1 Related Works

Governatori, etc. [1] gave a rich variety of normative conditionals and treated them in terms of computationally oriented non-monotonic multi-modal logic.

Our treatment of prerequisite-effect structures also results in a non-monotonic behavior of $\rightsquigarrow$ (see Proposition 5 (v)). However, our study is different from [1] in the following two respects. First, we keep our syntax simple and do not introduce temporal parameter in it. Rather, temporality is built in intuitionistic Kripke semantics. Second, we do not employ multi-modal operators and stay in a 'modality-free zone', which is suitable to logic-programming. Our syntax is just a propositional syntax of intuitionistic logic expanded with a new binary connective. Since the pages are very limited, this paper mainly focuses on the semantic aspect of prerequisite-effect structures. However, this does not mean that our study is immediately useless in non-monotonic reasoning of the legal context ([12]). We could combine our study in this paper with a framework of adaptive logic [13], which keeps a given premise set 'as normally as possible' with respect to some standard of normality and so can be applicable to handing inconsistency, etc. Finally, the reader can find another application of intuitionistic logic to juris-informatics legal ontologies in [14].

6.2 Conclusion and Further Directions

In this paper, we have formalized the dynamic prerequisite-effect structures in terms of expanded intuitionistic logic. To construct a formal reasoning system of law, we must mix various kinds of if-then relations, including classical implication, subsumption relation, temporal relation, and other causal relations. Since our ultimate objective is to represent legal documents in logic programming, those various kinds of logical relations should be represented in one and unique logic formalism. If we adopt modal operators for each relation such as temporal, deontic, and epistemic operators, the legal system would be a complicated product of polymodal logic, which is not realistic in implementation. This is because the method of product usually leads us to an undecidability result [15]. Thus, we avoided introducing modality.

For further research, we would like to suggest possible directions of this paper here. The first direction is concerned with a combination of $\rightsquigarrow$ with some other type of if-then relations, e.g., subsumption relation (cf. [4]), e.g., the notion of Japanese man subsumes the notion of Japanese. Assume the following: eIf Taro is a permanent Japanese, then income tax shall be imposed with respect to all income of Taro. Taro is a Japanese man.' Intuitively, we can conclude that 'if Taro is a permanent Japanese man, then income tax shall be imposed with respect to all income of Taro.' The reader can find another example in [4]. Can we formalize such legal inferences containing both subsumption relation and prerequisite-effect structure? To this aim, a first-order extension of $\mathcal{L}_{\mathrm{PE}}$ would be a plausible syntax for our formalization. Several studies [8, 9] of first-order intuitionistic or intermediate logic would be useful to this direction. Second, we do not know if the semantic consequence relation $\{ A_1, ..., A_n \} \models B$ is decidable in this stage. By Corollary 1, the question boils down to the decidability of $\mathsf{TPE}^- \vdash \varepsilon : A_1, \ldots, \varepsilon : A_n \overset{\{\varepsilon\}}{\Rightarrow} \varepsilon : B$. We conjecture that it is decidable. Third, we may apply the idea of $A \rightsquigarrow B$ to the other contexts. We have mentioned that $A \rightsquigarrow B$ does not allow any change other than B after when A is confirmed. It seems natural to investigate the connective $\rightsquigarrow$ also in the context of logic of change and/or causality. Finally, we hope to give an implementation to a legal reasoning system via our syntax with $\rightsquigarrow$ in terms of logic-programming[2].

References

[1] Governatori, G., Rotolo, A., Sartor, G.: Temporalised normative positions in defeasible logic. In: Gardner, A. (ed.) Procedings of the 10th International Conference on Artificial Intelligence and Law, pp. 25–34. ACM Press (2005)

[2] Gordon, T.F., Governatori, G., Rotolo, A.: Rules and Norms: Requirements for Rule Interchange Languages in the Legal Domain. In: Governatori, G., Hall, J., Paschke, A. (eds.) RuleML 2009. LNCS, vol. 5858, pp. 282–296. Springer, Heidelberg (2009)

[3] Satoh, K., Asai, K., Kogawa, T., Kubota, M., Nakamura, M., Nishigai, Y., Shirakawa, K., Takano, C.: PROLEG: An Implementation of the Presupposed Ultimate Fact Theory of Japanese Civil Code by PROLOG Technology. In: Onoda, T., Bekki, D., McCready, E. (eds.) JSAI-isAI 2010. LNCS (LNAI), vol. 6797, pp. 153–164. Springer, Heidelberg (2011)

[4] Hagiwara, S.: Multiple relations in legal documents based on intuitionistic logic. In: Proceedings of the 4th International Workshop on Juris-informatics (JURISIN 2010), pp. 41–50 (2010)

[5] Sano, K., Hagiwara, S., Tojo, S.: Prerequisite-effect structure on intuitionistic kripke model. In: Proceedings of the Fifth International Workshop on Juris-informatics (JURISIN 2011), pp. 50–61 (2011)

[6] Van Dalen, D.: Logic and Structure. Springer (2004)

[7] Prior, A.N.: Past, Present and Future. Clarendon Press, Oxford (1967)

[8] Kashima, R.: Sequent calculi of non-classical logics - Proofs of completeness theorems by sequent calculi. In: Proceedings of Mathematical Society of Japan Annual Colloquium of Foundations of Mathematics, pp. 49–67 (1999) (in Japanese)

[9] Ishigaki, R., Kikuchi, K.: Tree-Sequent Methods for Subintuitionistic Predicate Logics. In: Olivetti, N. (ed.) TABLEAUX 2007. LNCS (LNAI), vol. 4548, pp. 149–164. Springer, Heidelberg (2007)

[10] Sano, K.: Sound and Complete Tree-Sequent Calculus for Inquisitive Logic. In: Ono, H., Kanazawa, M., de Queiroz, R. (eds.) WoLLIC 2009. LNCS (LNAI), vol. 5514, pp. 365–378. Springer, Heidelberg (2009)

[11] Reiter, R.: A logic for default reasoning. Artificial Intelligence 13, 81–132 (1980)

[12] Governatori, G., Rotolo, A.: Changing legal systems: legal abrogations and annulments in defeasible logic. Logic Journal of the IGPL 18(1), 157–194 (2010)

[13] Batens, D.: A universal logic approach to adaptive logics. Logica Universalis 1(1), 221–242 (2007)

[14] Haeusler, E.H., De Paiva, V., Rademaker, A.: Intuitionistic logic and legal ontologies. In: Winkels, R.G. (ed.) Legal Knowledge and Information Systems - JURIX 2010: The Twenty-Third Annual Conference. Frontiers in Artificial Intelligence and Applications, vol. 223, pp. 155–158 (2010)

---

[2] The first author would like to thank Katsuhiko Toyama, Fumihiko Takahashi, Ken Satoh and Robert Kowalski for their discussions at JURISIN 2011. He also wishes to thank three reviewers for their careful and constructive suggestions and comments. All errors, however, are mine.

[15] Gabbay, D.M., Kurucz, A., Wolter, F., Zakharyaschev, M.: Many-Dimensional Modal Logics: Theory and Applications. Studies in Logic and the Foundations of Mathematics. Elsevier, North-Holland (2003)

A Proof of Completeness of **TPE** for Intuitionistic Kripke Semantics

This section gives an outline of our proof of Theorem 2. In the following, Γ, Δ and $\mathcal{T}$ are possibly infinite in the expression $\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$ of a tree-sequent. In the case where Γ, Δ and $\mathfrak{T}$ are all finite, $\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$ is said to be *finite*. A (possibly infinite) tree-sequent $\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$ is *provable* in TPE^- if there is a finite tree sequent $\Gamma' \overset{\mathcal{T}'}{\Rightarrow} \Delta'$ such that $\Gamma' \subseteq \Gamma$, $\Delta' \subseteq \Delta$, and $\mathcal{T}' \subseteq \mathcal{T}$. In what follows, we extend our notation $\mathsf{TPE}^- \vdash \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$ to cover any possibly infinite tree-sequent in the above sense.

Definition 2. *A tree-sequent $\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$ is* saturated *if it satisfies the following:*

(consistency) (i) *If $\alpha : A \in \Gamma$, then $\alpha : A \notin \Delta$;* (ii) *$\alpha : \bot \notin \Gamma$.*
(hereditary condition) *If $\alpha : A \in \Gamma$ and $\alpha \prec \beta$, then $\beta : A \in \Gamma$.*
($\wedge$l) *If $\alpha : A \wedge B \in \Gamma$, then $\alpha : A \in \Gamma$ and $\beta : A \in \Gamma$.*
($\wedge$r) *If $\alpha : A \wedge B \in \Delta$, then $\alpha : A \in \Delta$ or $\beta : A \in \Delta$.*
($\vee$l) *If $\alpha : A \vee B \in \Gamma$, then $\alpha : A \in \Gamma$ or $\beta : A \in \Gamma$.*
($\vee$r) *If $\alpha : A \vee B \in \Delta$, then $\alpha : A \in \Delta$ and $\beta : A \in \Delta$.*
($\to$l) *If $\alpha : A \to B \in \Gamma$, then $\alpha : A \in \Delta$ or $\alpha : B \in \Gamma$.*
($\to$r) *If $\alpha : A \to B \in \Delta$, then $\beta : A \in \Gamma$ and $\beta : B \in \Delta$ for some $\beta \succ \alpha$.*
($\rightsquigarrow$l) *If $\alpha : A \rightsquigarrow B \in \Gamma$, then $\alpha : A \in \Delta$, $\beta : A \in \Gamma$ for some descendant β of α, or ($\gamma : B \in \Gamma$ and $\alpha : B \in \Delta$) for all $\gamma \succ \alpha$.*
($\rightsquigarrow$r) *If $\alpha : A \rightsquigarrow B \in \Delta$, then*
  - *($\beta : B \in \Gamma$, $\beta : A \in \Gamma$ and $\alpha : A \in \Delta$) for some $\beta \succ \alpha$, or*
  - *($\beta : A \in \Gamma$, $\alpha : A \in \Delta$ and $\gamma : B \in \Delta$) for some β and γ with $\alpha \prec \beta \prec \gamma$*

Lemma 1. *If a finite tree-sequent $\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$ is not provable in TPE^-, then there exists a saturated tree-sequent $\Gamma^+ \overset{\mathcal{T}^+}{\Rightarrow} \Delta^+$ such that $\Gamma \subseteq \Gamma^+$, $\Delta \subseteq \Delta^+$, $\mathcal{T} \subseteq \mathcal{T}^+$, and $\Gamma^+ \overset{\mathcal{T}^+}{\Rightarrow} \Delta^+$ is not provable in TPE^-.*

We can show this lemma by the standard argument as in [9, 10]. Let us give a proof of Theorem 2.

Proof. We show the contrapositive implication. Assume that $\mathsf{TPE}^- \nvdash \varepsilon : A_1, \ldots, \varepsilon : A_n \overset{\{\varepsilon\}}{\Rightarrow} \varepsilon : B$. By Lemma 1, there exists some saturated tree-sequent $\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$ such that $\varepsilon : A_i \in \Gamma$ $(1 \le i \le n)$, $\varepsilon : B \in \Delta$, $\varepsilon \in \mathcal{T}$, and $\mathsf{TPE}^- \nvdash \Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$. Let us define a valuation V on a Kripke frame $(W, \prec^*)$ ($\prec^*$ is the reflexive and transitive closure of $\prec$) by $V(p) := \{\alpha \in \mathcal{T} \mid \alpha : p \in \Gamma\}$. By the saturation of $\Gamma \overset{\mathcal{T}}{\Rightarrow} \Delta$, V satisfies the hereditary condition. By induction on A, we can establish the following claim: $\alpha : A \in \Gamma$ implies $\alpha \models A$, and $\alpha : A \in \Delta$ implies $\alpha \not\models A$. It follows from this claim and our assumption that $\varepsilon \models A_i$ $(1 \le i \le n)$ and $\varepsilon \not\models B$. Therefore, we can conclude that $\{A_1, \ldots, A_n\} \not\models B$. □

Second Workshop on Algorithms for Large-Scale Information Processing in Knowledge Discovery (ALSIP)

Koji Tsuda[1,3] and Shin-ichi Minato[2,3]

[1] Computational Biology Research Center,
National Institute of Advanced Industrial Science and Technology
[2] Graduate School of Information Science and Technology, Hokkaido University
[3] JST ERATO Minato Discrete Structure Manipulation System Project

Preface

The Second Workshop on Algorithms for Large-Scale Information Processing in Knowledge Discovery (ALSIP 2011) is held on December 1-2 at Takamatsu Sunport Hall, Takamatsu, Japan. This workshop is a part of JSAI Symposium Series and succeeds the 1st ALSIP held in Osaka, Japan in 2008.

Information created by people has increased rapidly and now we are in a time which we could call the information-explosion era. To cope with such a large-scale information space, novel algorithms and data structures are desired for solving various problems in the area of knowledge discovery. This workshop aims to exchange fresh ideas on large-scale data processing in the problems such as data mining, clustering, machine learning, statistical analysis, and other computational aspects of knowledge discovery problems.

This year, we put special emphasis on succinct data structures and invited three prominent invited speakers: Rajeev Raman (University of Leicester, UK), Kunihiko Sadakane (National Institute of Informatics, Tokyo), Daisuke Okanohara (Preferred Infrastructure, Tokyo). In this proceedings, we accepted three papers out of ten submissions based on the reviews of the program committee members.

We would like to express our gratitude to the financial support from JSAI and JST ERATO Minato Project.

Workshop Co-chairs
Koji Tsuda
Shin-ichi Minato

Organizing Committee

- Koji Tsuda (AIST, Workshop Co-chair)
- Shin-ichi Minato (Hokkaido University, Co-chair)
- Jun Kawahara (JST, Local chair)

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 184–185, 2012.

Program Committee

- Hiroki Arimura (Hokkaido University)
- Marco Cuturi (Kyoto University)
- Koichi Hirata (Kyushu Institute of Technology)
- Michael Houle (National Institute of Informatics)
- Hisashi Kashima (University of Tokyo)
- Tetsuji Kuboyama (Gakushuin University)
- Hiroshi Sakamoto (Kyushu Institute of Technology)
- Takashi Washio (Osaka University)
- Osamu Watanabe (Tokyo Institute of Technology)
- Akihiro Yamamoto (Kyoto University)

An A* Algorithm for Computing Edit Distance between Rooted Labeled Unordered Trees[⋆]

Shoichi Higuchi[1], Tomohiro Kan[1], Yoshiyuki Yamamoto[1], and Kouichi Hirata[2]

[1] Graduate School of Computer Science and Systems Engineering
[2] Department of Artificial Intelligence
Kyushu Institute of Technology
Kawazu 680-4, Iizuka 820-8502, Japan
{syou_hig,kan,yamamoto,hirata}@dumbo.ai.kyutech.ac.jp

Abstract. In this paper, we design an A* algorithm for computing the edit distance between rooted labeled unordered trees. First, we introduce some *lower bounding functions* that provide the constant factor lower bounds on the edit distance. Then, by using the lower bounding functions as a heuristic function, we design the A* algorithm as the best-first search for the *edit distance search tree*. Finally, we give experimental results for the A* algorithm.

1 Introduction

Rooted labeled unordered trees (*trees*, for short) are rooted trees whose nodes are labeled and in which only ancestor relationship are significant. Such trees arise naturally in many fields such as glycan data or phylogenetic trees in bioinformatics, chemical compounds and their properties in chemistry, object representation and recognition in computer vision, and so on (*cf.*, [6]). For many such applications, it is necessary to compare tree by some meaningful distance measure.

The most famous distance measure between trees is the *edit distance* [1,7,8,11]. The edit distance between trees (that we sometimes call the *unordered tree edit distance*) is formulated as the minimum cost to transform a tree to another tree by applying *edit operations* of *substitutions*, *deletions* and *insertions* to trees. However, it is known that the problem of computing the unordered tree edit distance is intractable, that is, NP-hard [8,10] and MAX SNP-hard [3,9].

In order to compare two phylogenetic trees, Horesh *et al.* [4] have developed an A* *algorithm* for computing the edit distance between rooted *unlabeled* unordered trees, which is also an intractable problem. This A* algorithm uses three *lower bounding functions* that provide constant factor lower bounds on the unordered tree edit distance, including the *degree histogram L_1-distance* introduced by Kailing *et al.* [5]. Note that Kailing *et al.* [5] have introduced not only the

---

[⋆] This work is partially supported by Grand-in-Aid for Scientific Research 20500126, 21500145 and 22240010 from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 186–196, 2012.

degree histogram L_1-distance but also the *label histogram L_1-distance* as lower bounding functions.

Motivated by this A* algorithm [4], in this paper, we design an A* algorithm for computing the edit distance between rooted *labeled* unordered trees. First, we use not only three lower bounding functions introduced by Horesh *et al.* [4], that is, the difference of the number of nodes [4], the degree histogram L_1-distance [4,5] and the degree histogram L_∞-distance [4] but also additionally two lower bounding functions, that is, the label histogram L_1-distance [5] and the label histogram L_∞-distance. Next, we introduce the *edit distance search tree* to transform the problem of finding the shortest path in a connected graph for the standard A* algorithm to one of computing the unordered tree edit distance. Then, by setting the heuristic function to the maximum value of the above 5 lower bounding functions, we design the A* algorithm as the best-first search for the edit distance search tree with constructing just necessary branches.

Finally, we implement the A* algorithm and give experimental results for glycan data. Then, we compare the running time for computing the unordered tree edit distance by the A* algorithm with one by the exhaustive search algorithm designed by Shasha *et al.* [6] and the clique-based algorithm designed by Fukagawa *et al.* [2]. Furthermore, we evaluate the effect of 5 lower bounding functions in the execution of the A* algorithm.

2 Preliminaries

A *tree* is a connected graph without cycles. For a tree $T = (V, E)$, we denote V and E by $V(T)$ and $E(T)$, respectively. Also the *size* of T is $|V|$ and denoted by $|T|$. We sometimes denote $v \in V(T)$ by $v \in T$. A *rooted tree* is a tree with one node r chosen as its *root*. Here, we denote the root of a rooted tree T by $r(T)$. We denote the (complete) subtree of T rooted at $v \in T$ by $T(v)$.

For each node v in a rooted tree T with the root r, let $UP_T(v)$ be the unique path from v to r. The *parent* of $v(\neq r)$ is its adjacent node on $UP_T(v)$. The parent of the root r is undefined. We say that u is a *child* of v if v is the parent of u. Two nodes with the common parent are called *siblings*. A *leaf* is a node having no children.

Furthermore, the *depth* of v is the number of edges in the path from v to $r(T)$, that is, $|UP_T(v)| - 1$, and the *depth* of T is the maximum depth for every node in T, which we denote by $d(T)$. Also the *degree* of v is the number of the children of v, and the *degree* of T is the maximum degree for every node in T.

We say that a rooted tree is *ordered* if a left-to-right order among siblings is given; *Unordered* otherwise. Also we say that a tree is *labeled* over Σ if each node is assigned a symbol from a fixed finite alphabet Σ, where we denote the label of a node v by $l(v)$. We sometimes identify v with $l(v)$. In this paper, we call a rooted labeled unordered tree over Σ a *tree*, simply.

Definition 1 (Edit operations). Let T be a tree. Then, we call the following three operations *edit operations*. Also see Figure 1.

1. *Substitution*: Change the label of the node v in T (from l_1 to l_2).
2. *Deletion*: Delete a non-root node v in T (labeled by l_1) with a parent v' (labeled by l'), making the children of v become the children of v'. The children are inserted in the place of v as a subset of the children of v'.
3. *Insertion*: The complement of deletion. Insert a node v (labeled by l_2) as a child of v' (labeled by l') in T making v the parent of a subset of the children of v'.

For a special *blank* symbol $\varepsilon \notin \Sigma$, let $\Sigma_\varepsilon = \Sigma \cup \{\varepsilon\}$. Then, we represent each edit operation by $l_1 \mapsto l_2$, where $(l_1, l_2) \in (\Sigma_\varepsilon \times \Sigma_\varepsilon - \{(\varepsilon, \varepsilon)\})$. The operation is a substitution if $l_1 \neq \varepsilon$ and $l_2 \neq \varepsilon$, a deletion if $l_2 = \varepsilon$, and an insertion if $l_1 = \varepsilon$.

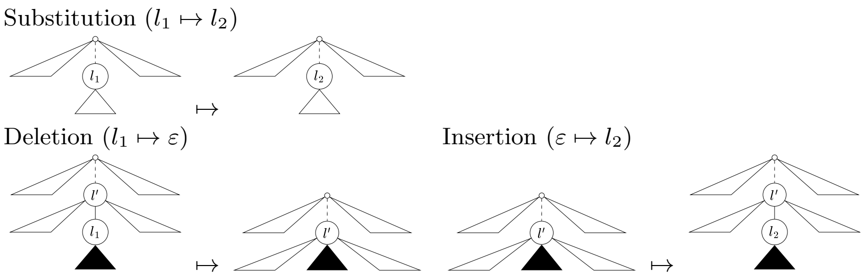


Fig. 1. Edit operations for trees

We define a *cost function* $\gamma : (\Sigma_\varepsilon \times \Sigma_\varepsilon - \{(\varepsilon, \varepsilon)\}) \mapsto \mathbf{R}$ on pairs of labels. We constrain a cost function γ to be a *metric*, that is, $\gamma(l_1, l_2) \geq 0$, $\gamma(l_1, l_1) = 0$, $\gamma(l_1, l_2) = \gamma(l_2, l_1)$ and $\gamma(l_1, l_3) \leq \gamma(l_1, l_2) + \gamma(l_2, l_3)$.

For a cost function γ, we define the *cost* of an edit operation by setting $\gamma(l_1 \mapsto l_2) = \gamma(l_1, l_2)$. The *cost* of a sequence $\mathbf{s} = s_1, \ldots, s_k$ of edit operations is given by $\gamma(S) = \sum_{i=1}^{k} \gamma(s_i)$.

Definition 2 (Edit distance). Let T and S be trees and γ a cost function. Then, the *edit distance* $\tau(T, S)$ between T and S under γ is defined as follow:

$$\tau(T, S) = \min \left\{ \gamma(\mathbf{s}) \;\middle|\; \begin{array}{l} \mathbf{s} \text{ is a sequence of edit operations} \\ \text{transforming from } T \text{ to } S \end{array} \right\}.$$

The edit distance is closely related to the following mapping [7].

Definition 3 (Mapping). For trees T and S, we say that the triple (M, T, S) is a *mapping* between T and S if $M \subseteq V(T) \times V(S)$ and every pair (v_1, w_1) and (v_2, w_2) in M satisfies the following conditions.

1. $v_1 = v_2$ iff $w_1 = w_2$ (one-to-one condition).
2. $v_1 \leq v_2$ iff $w_1 \leq w_2$ (ancestor condition).

We will use M instead of (M, T, S) when there is no confusion.

Let M be a mapping between T and S. Also let I_T (*resp.*, I_S) be the set of nodes in T (*resp.*, S) but not in M. Then, the *cost* $\gamma(M)$ of M is given as $\sum_{(v,w)\in M} \gamma(l(v), l(w)) + \sum_{v\in I_T} \gamma(l(v), \varepsilon) + \sum_{w\in I_S} \gamma(\varepsilon, l(w))$. Hence, it is known the following relationship between $\tau(T, S)$ and $\gamma(M)$ [7].

$$\tau(T, S) = \min\{\gamma(M) \mid M \text{ is a mapping between } T \text{ and } S\}.$$

3 A* Algorithm and Lower Bounding Functions

The A* *algorithm* is an algorithm to find the shortest path from the start node a to the goal node z in a connected graph by using an estimated length smaller than the actual length from a current node n to z.

For a current node n, suppose that $g(n)$ and $h(n)$ are the actual length of the shortest path from a to n and one from n to z, respectively. Then, for the length $f(n)$ of the shortest path from a to z through n, it holds that $f(n) = g(n)+h(n)$. However, since $f(n)$ is unknown in the procedure, we replace the actual length $f(n)$, $g(n)$ and $h(n)$ with the estimated length $f^*(n)$, $g^*(n)$ and $h^*(n)$ satisfying that $f^*(n) = g^*(n) + h^*(n)$.

Since we can find $g(n)$ in the procedure, we set $g^*(n)$ to $g(n)$. On the other hand, since we cannot find $h(n)$ in the procedure, we set $h^*(n)$, called a *heuristic function*, to a function that is guaranteed to be always smaller than $h(n)$.

In particular, in order to design the A* algorithm for computing $\tau(T, S)$, in this paper, we formulate such a heuristic function h^* based on the following *lower bounding functions* on τ. Here, throughout of this paper, we assume that a cost function is a *unit cost function* μ [11] satisfying that $\mu(l_1, l_2) = \mu(l_1, \varepsilon) = \mu(\varepsilon, l_2) = 1$ for $l_1, l_2 \in \Sigma$ and $l_1 \neq l_2$. We can extend μ to an arbitrary cost function easily.

Definition 4 (Lower bounding function). Let T be a tree. We call the histogram consisting of the degree of a node and its frequency in T the *degree histogram* of T. Also we call the histogram consisting of the label of a node and its frequency in T the *label histogram* of T.

Let T and S be trees. Then, we define $n(T, S)$ as $\big||T| - |S|\big|$, that is, the difference between the number of nodes in T and S. Also we define the *degree histogram* L_1*-distance* $d_1(T, S)$ and the *degree histogram* L_∞*-distance* $d_\infty(T, S)$ (*resp.*, the *label histogram* L_1*-distance* $l_1(T, S)$ and the *label histogram* L_∞*-distance* $l_\infty(T, S)$) as the L_1- and L_∞-distances between the degree (*resp.*, label) histograms of T and S.

Lemma 1 (Horesh *et al.* **[4], Kailing** *et al.* **[5]).** *For trees T and S, the following statements hold.*

1. $\tau(T, S) \geq n(T, S)$ [4].
2. $\tau(T, S) \geq d_1(T, S)/3$ [4,5] *and* $\tau(T, S) \geq d_\infty(T, S)$ [4].
3. $\tau(T, S) \geq l_1(T, S)/2$ [5] *and* $\tau(T, S) \geq l_\infty(T, S)$.

Proof. It is sufficient to show that $l_\infty(T,S) \leq \tau(T,S)$, that is, how many values of $l_\infty(T,S)$ change when an edit operation is applied. We denote the frequency of the label a in a tree T by $f(a,T)$.

When transforming from T to S by a substitution of b for a, it holds that $f(a,T)-1 = f(a,S)$, $f(b,T)+1 = f(b,S)$ and $f(c,T) = f(c,S)$ for every label c except a and b. Then, it holds that $l_\infty(T,S) \leq 1$. On the other hand, when transforming from T to S by a deletion of a, it holds that $f(a,T)-1 = f(a,S)$ and $f(c,T) = f(c,S)$ for every label c except a. Then, it holds that $l_\infty(T,S) \leq 1$.

Hence, it holds that $l_\infty(T,S) \leq \tau(T,S)$. □

By Lemma 1, for nodes $t \in T$ and $s \in S$, we use a heuristic function $h^*(t,s)$ in the A* algorithm as the maximum value of the above 5 lower bounding functions with constant factors of $T(t)$ and $S(s)$, that is:

$$h^*(t,s) = \max \left\{ \begin{array}{l} n(T(t),S(s)), \\ d_1(T(t),S(s))/3,\ d_\infty(T(t),S(s)), \\ l_1(T(t),S(s))/2,\ l_\infty(T(t),S(s)) \end{array} \right\}. \qquad (1)$$

Example 1. Consider the trees T and S in Figure 2. Then, the degree histograms and the label histograms of T and S are described as follows.

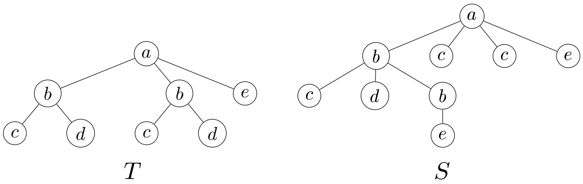


Fig. 2. Trees T and S in Example 1

degree	T	S
0	5	6
1	0	1
2	2	0
3	1	1
4	0	1

label	T	S
a	1	1
b	2	2
c	2	3
d	2	1
e	1	2

The values of the lower bounding functions for $r_1 = r(T)$ and $r_2 = r(S)$ are described as follows.

$$n(r_1,r_2) = 1, d_1(r_1,r_2) = 5, d_\infty(r_1,r_2) = 2, l_1(r_1,r_2) = 3, l_\infty(r_1,r_2) = 1.$$

Hence, it holds that $h^*(r_1,r_2) = \max\{1, 5/3, 2, 3/2, 1\} = 2$.

4 A* Algorithm for Computing the Tree Edit Distance for Unordered Trees

In order to transform the problem of finding the shortest path to one of computing $\tau(T,S)$ in the A* algorithm, we introduce the *edit distance search tree* $ET(T,S)$ between T and S. Suppose that every node $t \in T$ and $s \in S$ is numbered by its breadth-first search index $bf_T(t)$ and $bf_S(s)$ starting from 0.

Definition 5 (Edit distance search tree). For trees T and S, an *edit distance search tree* $ET(T,S)$ of T and S is a tree such that the depth is $|T|-1$, the label of the root is 0 and every non-leaf node has $|S|$ children labeled by $\varepsilon, 1, \dots, |S|-1$.

Furthermore, we say that a node v in $ET(T,S)$ is *valid* if the following set M_v of pairs of nodes in T and S forms a mapping between T and S.

$$M_v = \left\{ (t,s) \in T \times S \,\middle|\, \begin{array}{l} w \in UP_{ET(T,S)}(v) - \{\varepsilon\}, \\ d(w) = bf_T(t), l(w) = bf_S(s) \end{array} \right\}.$$

In this paper, we identify the edit distance search tree $ET(T,S)$ with one consisting of just valid nodes, and we also call the latter the edit distance search tree. Hence, the depth of $ET(T,S)$ is $|T|-1$ and the degree of $ET(T,S)$ is at most $|S|$. Every node $v \in ET(T,S)$ denotes the pair $(t,s) \in T \times S$, which is a component of a mapping, such that $d(v) = bf_T(t)$ and $l(v) = bf_S(s) (\neq \varepsilon)$.

Example 2. Consider the trees T and S in Figure 3 (left). Here, the number attached to nodes in T and the number of nodes in S denote the breadth-first search index of T and S, respectively.

Figure 3 (right) illustrates the edit distance search tree $ET(T,S)$ of T and S. For example, the path $\langle 0, 1, \varepsilon, 2 \rangle$ in $ET(T,S)$ consisting of the underlined number represents the mapping $\{(0,0),(1,1),(3,2)\}$ between T and S. In this path, the node labeled with 1 at depth 1 has just one child ε, because the mapping $\{(0,0),(1,1)\}$ cannot contain the pair $(2,2)$ satisfying the ancestor condition.

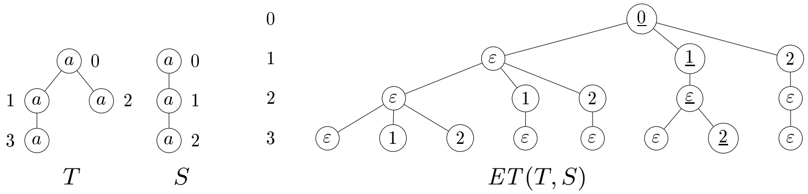


Fig. 3. Trees T and S (left) and the edit distance search tree $ET(T,S)$ (right) in Example 2

We explain how to compute the values of $g^*(t,s)$ and $h^*(t,s)$ in the A* algorithm. Let v be a node in $ET(T,S)$ such that $d(v) = bf_T(t)$ and $l(v) = bf_S(s)$ and v_p the parent of v in $ET(T,S)$ such that $d(v_p) = bf_T(t_p)$ and $l(v_p) = bf_S(s_p)$. Also we define the set $N_v(s) \subseteq S$ as follows.

$$N_v(s) = \left\{ s' \in S \,\middle|\, \begin{array}{l} 0 \leq bf_S(s') \leq bf_S(s), \text{ and} \\ M_v \cup \{(t', s')\} \text{ is not a mapping for every } t' \in T \end{array} \right\}.$$

Then, we compute $g^*(t, s)$ as follow.

$$g^*(t, s) = g^*(t_p, s_p) + |N_v(s)|. \tag{2}$$

On the other hand, consider $h^*(t, s)$. If $s \neq \varepsilon$, then we can compute $h^*(t, s)$ according to the equation (1). Otherwise, that is, in the case that $s = \varepsilon$, first search for the nearest ancestor $t' \in UP_T(t)$ to t in T such that $d(v') = bf_T(t')$ and $l(v') = bf_S(s') \neq \varepsilon$ for some $v' \in ET(T, S)$. For this t', let v be a node in $ET(T, S)$ such that $d(v) = t'$. Also suppose that $l(v) = s', d(v_p) = t'_p$ and $l(v_p) = s'_p$. Then, we define $c(t')$ as follows.

$$c(t') = \begin{cases} g^*(t'_p, s'_p) - g^*(t', s'), & \text{if } l(v) \neq \varepsilon, \\ g^*(t'_p, s'_p) - g^*(t', s') + 1, & \text{if } l(v) = \varepsilon. \end{cases}$$

Then, we compute $h^*(t, s)$ with the equation.

$$h^*(t, s) = \begin{cases} \text{the right hand side of the equation (1),} & \text{if } s \neq \varepsilon, \\ h^*(t', s') - \displaystyle\sum_{t'' \in T(t')} c(t''), & \text{if } s = \varepsilon. \end{cases} \tag{3}$$

Hence, the A* algorithm computes $\tau(T, S)$ by finding the path from the root to the leaves with the minimum estimated value $f^*(t, s) = g^*(t, s) + h^*(t, s)$ in $ET(T, S)$ according to the equations (1), (2) and (3). Since the full construction of $ET(T, S)$ in the A* algorithm is too redundant, we design the A* algorithm as the best-first search for $ET(T, S)$ with constructing just necessary branches of $ET(T, S)$.

Finally, we summarize the A* algorithm for computing $\tau(T, S)$ as follows, where L is a list of triples. Assume here that every mapping contains the pair of the roots in T and S.

1. Add $((0, 0), g^*(0, 0), h^*(0, 0))$ to L and draw the node labeled by 0 in $ET(T, S)$.
2. Select $((t, s), g^*(t, s), h^*(t, s))$ in L such that $g^*(t, s) + h^*(t, s)$ is minimum.
3. If $bf_T(t) = |T| - 1$, then output $g^*(t, s) + h^*(t, s)$ and halt. Otherwise:
    (a) Select $v \in ET(T, S)$ such that $d(v) = t$ and $l(v) = s$.
    (b) Draw the node u such that $l(u) = \varepsilon$ in $ET(T, S)$ as the child of v, and add $((t + 1, \varepsilon), g^*(t + 1, \varepsilon), h^*(t + 1, \varepsilon))$ to L.
    (c) For $t' \in T$ such that $bf_T(t') = d(v) + 1$ and for every $s' \in S$, if s' satisfies that $M_v \cup \{(t', s')\}$ forms a mapping between T and S, then draw the node u in $ET(T, S)$ such that $bf_S(s') = l(u)$ as the child of v and add $((t', s'), g^*(t', s'), h^*(t', s'))$ by the equations of (1), (2) and (3) to L.
    (d) Go to the statement 2.

Example 3. We apply the A* algorithm to compute the edit distance between T and S in Figure 4. Here, the number attached to nodes in T and S denotes the breadth-first search index of T and S, respectively.

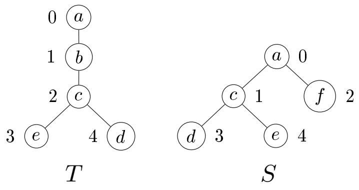


Fig. 4. Trees T and S in Example 3

Then, Figure 5 illustrates the running example of the A* algorithm for the trees T and S in Figure 4, with constructing the edit distance search tree $ET(T,S)$, from the root to leaves. Here, the number attached to a node n in $ET(T,S)$ denotes the value of $g^*(t,s)+h^*(t,s)$. Now we explain the run of the A* algorithm with Figure 5.

At the step (0), the A* algorithm adds the pair $(0,0)$ to a mapping and constructs the children of 0 at depth 1 in $ET(T,S)$.

At the step (1), the A* algorithm selects the path $\langle 0,\varepsilon\rangle$ in $ET(T,S)$, because the value $1+1$ of the node ε is minimum in the values at the depth 1 in $ET(T,S)$. Then, the A* algorithm adds no pair to a mapping and constructs the children of ε at depth 2 in $ET(T,S)$.

At the step (2), the A* algorithm selects the path $\langle 0,\varepsilon,1\rangle$ in $ET(T,S)$, because the value $2+0$ of the node 1 is minimum in the values at the depth 2 in $ET(T,S)$. Then, the A* algorithm adds the pair $(2,1)$ to a mapping and constructs the children of 1 at depth 3 in $ET(T,S)$.

At the step (3), the A* algorithm selects the path $\langle 0,\varepsilon,1,4\rangle$ in $ET(T,S)$, because the value $2+0$ of the node 4 is minimum in the values at the depth 3 in $ET(T,S)$. Then, the A* algorithm adds the pair $(3,4)$ to a mapping and constructs the children of 4 at depth 4 in $ET(T,S)$.

At the step (4), the A* algorithm selects the path $\langle 0,\varepsilon,1,4,3\rangle$ in $ET(T,S)$, because the value $2+0$ of the node 3 is minimum in the values at the depth 4 in $ET(T,S)$. Then, the A* algorithm adds the pair $(4,3)$ to a mapping.

Hence, the A* algorithm returns $2+0=2$ as $\tau(T,S)$, and its mapping between T and S is $\{(0,0),(2,1),(3,4),(4,3)\}$.

5 Experimental Results

We give experimental results by comparing the A* algorithm with the exhaustive search algorithm designed by Shasha *et al.* [6] and the clique-based algorithm designed by Fukagawa *et al.* [2]. Here, our computer environment is that OS is Microsoft Windows 7, CPU is Core i7 920 2.67GHz and RAM is 3GB.

We use the same dataset as [2], consisting of 352 glycan data including 137 leukemia and 14 erythrocyte. While we implement both the A* algorithm and

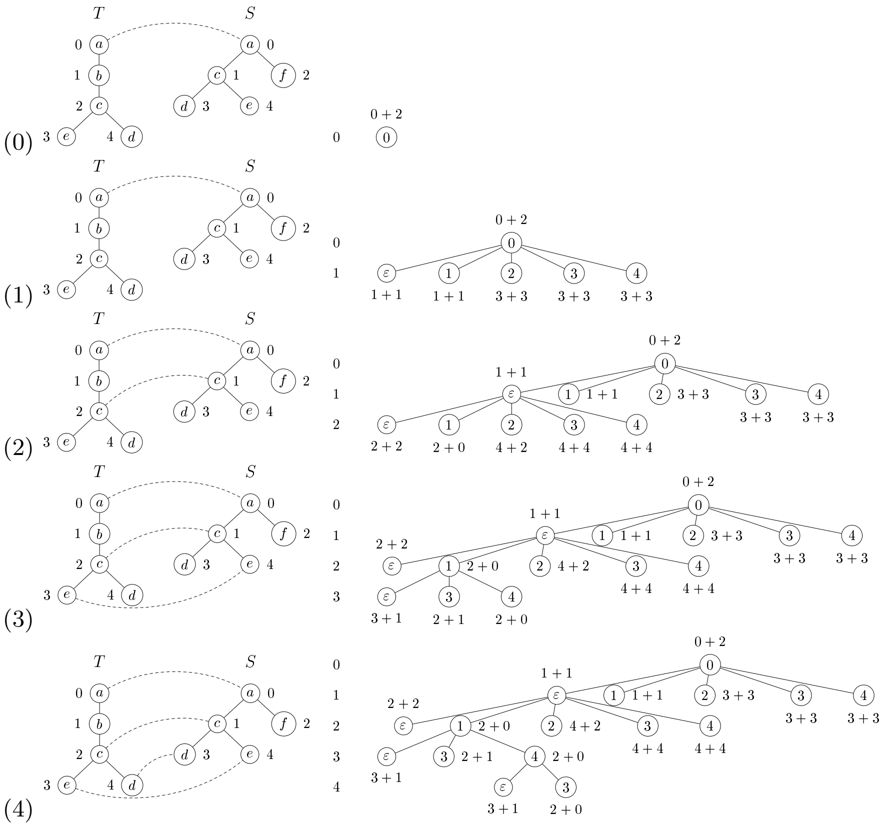


Fig. 5. The running example of the A* algorithm for the trees T and S in Figure 4, with constructing the edit distance search tree $ET(T, S)$

the exhaustive search algorithm, we cite the result of the clique-based algorithm to the paper [2][1]. Then, we obtain the result described in Table 1.

Hence, the A* algorithm is much efficient than the exhaustive search algorithm [6] and so efficient as the clique-based algorithm [2] to compute τ. It is advantage for the A* algorithm to adopt the unit cost function (and then extend an arbitrary cost function), while the clique-based algorithm [2] depends on a special cost function such that $\gamma(l_1, l_1) = 2$ and $\gamma(l_1, l_2) = 1$ for $l_1 \neq l_2$.

Furthermore, we evaluate the effect of lower bounding functions in the A* algorithm. Table 2 shows the running time of the A* algorithm when excluding at most one lower bounding function.

Hence, for glycan data, $n(T, S)$ is the most effective lower bounding function. Also $d_\infty(T, S)$ and $l_\infty(T, S)$ are more effective than $d_1(T, S)/3$ and $l_1(T, S)/2$, respectively.

---

[1] The computer environment of the clique-based algorithm [2] is that OS is Microsoft Windows XP, CPU is Intel Core 2 Duo 2.8GHz and RAM is 3.48GB.

Table 1. The running time (sec.) to compute unordered tree edit distance

		all data	leukemia and erythrocyte
exhaustive search algorithm	[6]	2133.03	751.28
A^* algorithm		161.56	21.36
clique-based algorithm	[2]	–	48.33

Table 2. The effect of lower bounding functions in the A^* algorithm (sec.)

excluded lower bounding function	all data	leukemia and erythrocyte
none	161.56	21.36
$n(T, S)$	338.61	41.60
$d_1(T, S)/3$	169.12	21.28
$d_\infty(T, S)$	207.06	32.64
$l_1(T, S)/2$	176.92	21.92
$l_\infty(T, S)$	187.44	24.32

6 Conclusion

In this paper, we have designed and implemented the A^* algorithm for computing an edit distance between rooted labeled unordered trees. Then, we have applied the A^* algorithm to glycan data and obtained the result that the A^* algorithm is much efficient than the exhaustive search algorithm [6] and so efficient as the clique-based algorithm [2]. Furthermore, we have evaluated that the difference of the nodes of subtrees is the most effective lower bounding function.

It is a future work to apply the A^* algorithm to other data such as XML data and evaluate the effect of lower bounding functions, that is, analyze the relationship between the structures of trees and the effect of lower bounding functions. It is also a future work to improve the A^* algorithm by introducing other lower bounding functions or to design other algorithm to compute an unordered edit distance more efficient than the A^* algorithm.

Acknowledgment. The authors would like to thank Prof. Tetsuji Kuboyama at Gakushuin University for his fruitful suggestion to improve the experimental results. Also they would like to thank the anonymous reviewers of ALSIP'11 for their valuable comments to revise this paper.

References

1. Bille, P.: A survey on tree edit distance and related problems. Theoret. Comput. Sci. 337, 217–239 (2005)
2. Fukagawa, D., Tamura, T., Takasu, A., Tomita, E., Akutsu, T.: A clique-based method for the edit distance between unordered trees and its application to analysis of glycan structures. BMC Bioinformatics 12 (2011)
3. Hirata, K., Yamamoto, Y., Kuboyama, T.: Improved MAX SNP-Hard Results for Finding an Edit Distance between Unordered Trees. In: Giancarlo, R., Manzini, G. (eds.) CPM 2011. LNCS, vol. 6661, pp. 402–415. Springer, Heidelberg (2011)
4. Horesh, Y., Mehr, R., Unger, R.: Designing an A^* algorithm for calculating edit distance between rooted-unordered trees. J. Comput. Bio. 13, 1165–1176 (2006)
5. Kailing, K., Kriegel, H.-P., Schönauer, S., Seidl, T.: Efficient Similarity Search for Hierarchical Data in Large Databases. In: Bertino, E., Christodoulakis, S., Plexousakis, D., Christophides, V., Koubarakis, M., Böhm, K. (eds.) EDBT 2004. LNCS, vol. 2992, pp. 676–693. Springer, Heidelberg (2004)
6. Shasha, D., Wang, J.T.-L., Zhang, K., Shih, F.Y.: Exact and approximate algorithms for unordered tree matching. IEEE Trans. Sys. Man and Cybernet. 24, 668–678 (1994)
7. Tai, K.-C.: The tree-to-tree correction problem. J. ACM 26, 422–433 (1979)
8. Zhang, K., Shasha, D.: Tree pattern matching. In: Apostolico, A., Galil, Z. (eds.) Pattern Matching Algorithms, pp. 341–371 (1997)
9. Zhang, K., Jiang, T.: Some MAX SNP-hard results concerning unordered labeled trees. Inform. Process. Lett. 49, 249–254 (1994)
10. Zhang, K., Statman, R., Shasha, D.: On the editing distance between unordered labeled trees. Inform. Process. Lett. 42, 133–139 (1992)
11. Zhang, K., Shasha, D.: Simple fast algorithms for the editing distance between trees and related problems. SIAM J. Comput. 18, 1245–1262 (1989)

Mining Closed Weighted Itemsets for Numerical Transaction Databases

Yuichi Kameda and Akihiro Yamamoto

Graduate School of Informatics Kyoto University
`y_kameda@iip.ist.i.kyoto-u.ac.jp`, `akihiro@i.kyoto-u.ac.jp`

Abstract. In this article we extend the notion of closed itemsets of binary transaction databases to numerical transaction databases, and give an algorithm to mine them. We compare the computation time of our method and the case using scaling technique. We consider the case that information of closed itemsets of binarized database is given, and investigate how changes if algorithm utilize the information for mining by some experiments.

1 Introduction

In data mining, frequent patterns are widely noticed as a fundamental of many useful applications, for example, association rules [1]. For binary transaction databases, various algorithms for mining frequent itemsets are proposed, such as Apriori [1] or FP-Growth [2]. To mine patterns from numerical transaction databases, one can convert numerical transaction databases to binary ones and can apply methods for binary databases to them, with using scaling technique [3]. But scaling usually increases the number of items much, so it makes mining more difficult. Therefore, direct methods of mining patterns from numerical databases are still useful and needed, and many researches performed until today, like closed interval pattern mining [4] or high utility mining [5].

In this article we focus on mining "closed itemsets" from numerical transaction databases. Frequent closed patterns [6] are the subset of frequent patterns that have enough information of all frequent patterns. Thus, by using closed patterns, one can reduce the number of patterns to be considered. Numerical databases are expected to have much more frequent patterns than binary ones, so the notion of closed patterns must be useful. Hence we propose the idea of closed weighted itemsets, as a natural extension of closed itemsets of binary databases, and consider a method for mining them.

2 Preliminaries

In this section we briefly review definitions for binary transaction databases, according to [7]. Throughout this article, $\mathcal{I} = \{p_1, p_2, \ldots, p_N\}$ denotes a set of *items*. We denote the set $\{1, 2, \ldots, N\}$ by $[1, N]$.

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 197–210, 2012.

A subset of $\mathcal{I}$ is called an *itemset*. A *binary transaction database* $\mathcal{D}$ over $\mathcal{I}$ is a finite sequence of itemsets $X_1, X_2, \ldots, X_M$. Elements of $\mathcal{D}$ are called *transactions* of $\mathcal{D}$. In the followings, N and M always represent the numbers of items and transactions of $\mathcal{D}$ respectively.

If an itemset $I \subseteq \mathcal{I}$ is included in a transaction X, then we say that I *occurs* in X. For an itemset $I \subseteq \mathcal{I}$, the *frequency* of I in $\mathcal{D}$ is defined by the number of transactions of $\mathcal{D}$ that I occurs in: that is, $\sharp\{i \in [1, M] \mid I \subseteq X_i\}$, where $\sharp S$ means the number of elements of the set S. We denote the frequency of I by $freq(I)$. Clearly, if $I \subseteq J$ then $freq(I) \geq freq(J)$.

An itemset I is called *closed* if $freq(I) > freq(J)$ for all $J \supsetneq I$. Closed itemsets can also be defined by using Galois connection as follows: let

$$\begin{aligned} \tau :\ & 2^{\mathcal{I}} \to 2^{[1,M]}, \quad \tau(I) = \{i \in [1, M] \mid I \subseteq X_i\}, \\ \iota :\ & 2^{[1,M]} \to 2^{\mathcal{I}}, \quad \iota(A) = \{p \in \mathcal{I} \mid p \in X_a \text{ for all } a \in A\}, \end{aligned}$$

where 2^S denotes the power set of the set S. It is known that the composition $\varphi = \iota \circ \tau$ satisfies the condition of closure operator: for each $I, J \subseteq \mathcal{I}$,
(CO1) $I \subseteq \varphi(I)$,
(CO2) $I \subseteq J \Rightarrow \varphi(I) \subseteq \varphi(J)$, and
(CO3) $\varphi(\varphi(I)) = \varphi(I)$.
Then the following holds:

Proposition 2.1. *Let I be an itemset. I is closed if and only if $I = \varphi(I)$.*

Example 2.2. A binary transaction database can be expressed as a table form illustrated on Table 1:

Table 1. An example of binary transaction database

ID	A	B	C	D	E
1	×	×	×		
2	×	×		×	
3	×			×	×
4		×	×		×
5	×	×	×		×
6	×		×		
7		×	×		×

For example, the first row of the table above corresponds an itemset $I = \{A, B, C\}$.

3 Numerical Transaction Database

In this article we define numerical transaction databases in the following way.

Let $\mathbb{N}$ be the set of all positive integers. A *weighted itemset* is a pair (I, ω) of a subset I of $\mathcal{I}$ and a mapping $\omega : I \to \mathbb{N}$. We say ω a *weight* on I and $\omega(p)$ a *weight* (or *value*) of the item p. A *numerical transaction database* $\mathcal{D}$ is a finite sequence of weighted itemsets $(I_1, \omega_1), (I_2, \omega_2), \ldots, (I_M, \omega_M)$.

Example 3.1. A numerical transaction database also can be expressed as a table form illustrated on Table 2:

Table 2. An example of numerical transaction database

ID	A	B	C	D	E
1	1	2	4		
2	2	3		2	
3	2			3	2
4		3	5		1
5	2	1	3		1
6	1		5		
7		3	3		3

For example, the first row of the table above corresponds a weighted itemset (I, ω) where $I = \{A, B, C\},\ \omega : I \to \mathbb{N},\ \omega(A) = 1,\ \omega(B) = 2$ and $\omega(C) = 4$.

An elements of $\mathcal{D}$ is called a *transaction* of $\mathcal{D}$, similar to the binary case. Next we define the occurrence and the frequency of a weighted itemset.

Definition 3.2. 1. Let (I, ω) and (J, ω') be weighted itemsets. If they satisfy the conditions (A) $I \subseteq J$ and (B) $\omega(p) \leq \omega'(p)$ for each $p \in I$, then we say that (I, ω) *occurs* in (J, ω'), and denote it by $(I, \omega) \preceq (J, \omega')$.
2. Let $\mathcal{D} = (I_1, \omega_1), (I_2, \omega_2), \ldots, (I, \omega_M)$ be a numerical transaction database and (I, ω) be a weighted itemset. Then the *frequency* of (I, ω), denoted by $freq(I, \omega)$, is defined by $\sharp\{k \in [1, M] \mid (I, \omega) \preceq (I_k, \omega_k)\}$.

Remark 3.3. Clearly the relation $\preceq$ becomes a partial order on the set of all weighted itemsets.

Intuitively, $(I, \omega) \preceq (J, \omega')$ means that "(I, ω) is contained in (J, ω') with taking values into account". In the rest of this article we fix a numerical transaction database $\mathcal{D} = (I_1, \omega_1), \ldots, (I_M, \omega_M)$. Using the relation $\preceq$ we define closed weighted itemsets as follows:

Definition 3.4. A weighted itemset (I, ω) is said to be *closed* if $freq(I, \omega) > freq(J, \omega')$ for every weighted itemset (J, ω') that satisfies $(J, \omega') \succ (I, \omega)$.

Likewise for the binary case, we can define closed weighted itemsets in an alternative way using Galois connection.

For two weighted itemsets (I, ω) and (J, ω'), we define

$$\begin{aligned}
&(I,\omega) \wedge (J,\omega') = (I \cap J, \omega \wedge \omega') \\
&\qquad\qquad \text{where } \ \omega \wedge \omega' : I \cap J \to \mathbb{N},\ p \mapsto \min(\omega(p), \omega'(p)), \\
&(I,\omega) \vee (J,\omega') = (I \cup J, \omega \vee \omega') \\
&\qquad\qquad \text{where } \ \omega \vee \omega' : I \cup J \to \mathbb{N},\ p \mapsto \begin{cases} \max(\omega(p), \omega'(p)) & \text{if } \ p \in I \cap J \\ \omega(p) & \text{if } \ p \in I \setminus J \\ \omega'(p) & \text{if } \ p \in J \setminus I, \end{cases}
\end{aligned}$$

and set two mappings τ and ι by

$$\begin{aligned}
\tau :&\ \mathcal{WI} \to 2^{[1,M]}, \quad \tau(I,\omega) = \{k \in [1,M] \mid (I,\omega) \preceq (I_k, \omega_k)\}, \\
\iota :&\ 2^{[1,M]} \to \mathcal{WI}, \quad \iota(A) = \bigwedge \{(I_k, \omega_k) \mid k \in A\},
\end{aligned}$$

where $\mathcal{WI}$ denotes the set of all weighted itemsets.

Remark 3.5. Note that the operations $\wedge$ and $\vee$ are the meet and join defined from the partial order $\preceq$ respectively.

Example 3.6. Here we give a simple example of $\wedge$ and $\vee$. Let (I, ω) and (J, ω') be weighted itemsets on the Table 3(a). Then the meet and join of (I, ω) and (J, ω') are illustrated on (b).

Table 3. Example of $\wedge$ and $\vee$

	A	B	C	D	E
(I,ω)	1		3	3	
(J,ω')		2	2	4	

(a)

	A	B	C	D	E
$(I,\omega) \wedge (J,\omega')$			2	3	
$(I,\omega) \vee (J,\omega')$	1	2	3	4	

(b)

Let $\varphi = \iota \circ \tau$. Then, similar to the binary case, it holds that

Proposition 3.7. *$\varphi : \mathcal{WI} \to \mathcal{WI}$ satisfies the following three conditions: for each (I, ω) and (J, ω'),*
(CO1)' *$(I,\omega) \preceq \varphi(I,\omega)$,*
(CO2)' *$(I,\omega) \preceq (J,\omega') \Rightarrow \varphi(I,\omega) \preceq \varphi(J,\omega')$, and*
(CO3)' *$\varphi(\varphi(I,\omega)) = \varphi(I,\omega)$.*

Proof. Let $A = \tau(I, \omega)$.
(CO1)' By definition we obtain

$$\varphi(I,\omega) = \bigwedge \{(I_k, \omega_k) \mid k \in A\} = (\cap_{k \in A} I_k, \wedge_{k \in A} \omega_k).$$

If $k \in A$, by the definition of τ, $I \subseteq I_k$ and $\omega(p) \leq \omega'(p)$ for all $p \in I$. Hence $I \subseteq \cap_{k \in A} I_k$ and $\omega(p) \leq \min_{k \in A}(\omega_k(p)) = (\wedge_{k \in A}\omega_k)(p)$. Thus $(I, \omega) \preceq \varphi(I, \omega)$.
(CO2)' Let $A' = \tau(J, \omega')$. Obviously $A \supseteq A'$. Therefore, we have

$$\begin{aligned} &\cap_{k \in A} I_k \subseteq \cap_{k \in A'} I_k, \text{ and} \\ &(\wedge_{k \in A}\omega_k)(p) = \min_{k \in A} \omega_k(p) \leq \min_{k \in A'} \omega_k(p) = (\wedge_{k \in A'}\omega_k)(p) \text{ for all } p \in \cap_{k \in A} I_k. \end{aligned}$$

These imply $\varphi(I, \omega) \preceq \varphi(J, \omega')$.
(CO3)' We claim that

Lemma 3.8. *$\tau(\varphi(I, \omega)) = \tau(I, \omega)$.*

By this lemma we have

$$\varphi(\varphi(I, \omega)) = \bigwedge_{k \in \tau(\varphi(I,\omega))} (I_k, \omega_k) = \bigwedge_{k \in \tau(I,\omega)} (I_k, \omega_k) = \varphi(I, \omega).$$

Proof. of Lemma 3.8. $\tau(\varphi(I, \omega)) \subseteq \tau(I, \omega)$ immediately follows from $(I_t, \omega_t) \succeq \varphi(I, \omega) \succeq (I, \omega)$.
Let $A = \tau(I, \omega)$. Assume $t \in A$. By the assumption we have $I_t \supseteq \cap_{k \in A} I_k$ and it holds that $\omega_t(p) \geq \min_{k \in A} \omega_k(p)$ for each $p \in \cap_{k \in A} I_k$. This implies $(I_t, \omega_t) \succeq \varphi(I, \omega)$, therefore $t \in \tau(\varphi(I, \omega))$. □

Proposition 3.9. *A weighted itemset (I, ω) is closed if and only if it satisfies that $\varphi(I, \omega) = (I, \omega)$.*

Proof. $\Rightarrow$) Assume that $(I, \omega) \neq \varphi(I, \omega)$. By (CO1)', this means $(I, \omega) \prec \varphi(I, \omega)$. Lemma 3.8 implies that $freq(\varphi(I, \omega)) = freq(I, \omega)$. Therefore, (I, ω) is not closed, and it is contradiction.
$\Leftarrow$) Assume that (I, ω) is not closed, that is, there exists $(J, \omega') \succ (I, \omega)$ such that $freq(I, \omega) = freq(J, \omega')$. This assumption implies that $\tau(I, \omega) = \tau(J, \omega')$, so we have $\varphi(I, \omega) = \varphi(J, \omega')$. By (CO1)', $\varphi(J, \omega') \succeq (J, \omega')$. Therefore, $(I, \omega) = \varphi(I, \omega) = \varphi(J, \omega') \succeq (J, \omega') \succ (I, \omega)$, and this is contradiction. □

For a weighted itemset (I, ω), we say the closed weighted itemset $\varphi(I, \omega)$ the *closure* of (I, ω).

Example 3.10. The closed weighted itemsets of Example 3.1 are listed in Table 4, in the table we use the notation such that the element $A1B1$ means the weighted itemset $(\{A, B\}, (A \mapsto 1, B \mapsto 1))$.
Note that, for a fixed set of items I, there can be more than one mapping ω that makes (I, ω) closed.

For a numerical transaction database $\mathcal{D} = (I_1, \omega_1), \ldots, (I_M, \omega_M)$, the sequence of itemsets $I_1, I_2, \ldots, I_M$ can be regarded as a binary transaction database. We denote it by $S(\mathcal{D})$ and call it the *support* of $\mathcal{D}$.

Table 4. Closed weighted itemsets of Example 3.1

freq	closed weighted itemsets
5	$A1, B1, C3$
4	$B2, B1C3, E1$
3	$A2, A1B1, B3, A1C3, B2C3, C4, B1C3E1$
2	$A2B1, A1B2, A1B1C3, A1C4, B2C4, C5, A2D2, A2E1, B3C3E1, E2$
1	$A1B2C4, A1C5, A2B3D2, A2B1C3E1, B3C5E1, A2D3E2, B3C3E3$

Remark 3.11. We use the terminology "support" nevertheless it is used to indicate the number $freq(I)/M$ for a given itemset I, because this is an analogy to the "support" of a function in mathematics. (Let $f : X \to \mathbb{R}$: the set of all real numbers. The support of f is the subset $\{x \in X \mid f(X) \neq 0\}$.)

Example 3.12. A binary transaction database of Example 2.2 is the support of the numerical transaction database of Example 3.1.

Then the following holds:

Proposition 3.13. *1. Let (I, ω) be a closed weighted itemset of $\mathcal{D}$. Then I is a closed itemset of $S(\mathcal{D})$.*
2. Conversely, let I be a closed itemset of $S(\mathcal{D})$. Then there exists a mapping $I \to \mathbb{N}$ that makes (I, ω) a closed weighted itemset of $\mathcal{D}$.

Proof. 1. If I is not closed, then there is an item $q \notin I$ such that $I_k \supseteq I \Rightarrow I_k \ni q$. This implies that, if $(I_k, \omega_k) \succeq (I, \omega)$, then $I_k \ni q$. Let

$$\omega' : I \cup \{q\} \to \mathbb{N}, \quad \omega'(p) = \omega(p)\ (p \in I), \quad \omega'(q) = 1.$$

Then $(I \cup \{q\}, \omega') \succ (I, \omega)$ and, all (I_k, ω_k)'s with $(I_k, \omega_k) \succeq (I, \omega)$ satisfy that $(I_k, \omega_k) \succeq (I \cup \{q\}, \omega')$. This means $freq(I, \omega) = freq(I \cup \{q\}, \omega')$, so (I, ω) is not closed.
2. Let $(I_{k_1}, \omega_{k_1}), \ldots, (I_{k_f}, \omega_{k_f})$ be all transactions of $\mathcal{D}$ such that $I_{k_j} \supseteq I$. Then $\cap I_{k_j} = I$ holds: otherwise I can not be closed. Now let $\hat{\omega}$ be the mapping defined by $\hat{\omega}(p) = 1$ for all $i \in I$. Clearly $(I_{k_j}, \omega_{k_j}) \succeq (I, \hat{\omega})$ for all j's. Set $(J, \omega') = \iota \circ \tau(I, \hat{\omega}) = \wedge\{(I_{k_j}, \omega_{k_j})\}$. Proposition 3.9 ensures that (J, ω') is closed, and by the definition of $\wedge$, $J = I$. □

Remark 3.14. Proposition 3.13 also holds for frequent weighted itemsets instead of closed weighted itemsets.

According to the argument so far, we can consider the following method to mine all closed weighted itemsets:

1. Mine all closed itemsets of $S(\mathcal{D})$;
2. For each closed itemset I, find all mappings $\omega : I \to \mathbb{N}$ that make (I, ω) closed.

In the next section we discuss this method more precisely.

At the end of this section, we mention scaling. *Scaling* [3] is a method to binarize a numerical transaction database. Scaling transforms a combination of an item and its value into an item under a certain rule called *scale*. There are various scales have been designed to suit each purpose. Here we adopt *ordinal scales*.

Let $\mathcal{D} = (I_1, \omega_1), \ldots, (I_M, \omega_M)$ be a given numerical transaction database in the set of items $\mathcal{I}$. The ordinal scale we consider here scales $\mathcal{D}$ in the following way. Let $p \in \mathcal{I}$ and let

$$W_p := \{a \mid \exists (I, \omega) : \text{ transaction of } \mathcal{D} \text{ such that } p \in I, \omega(p) = a\}$$

be the set of weights of p. Then we regard $(p \geq a)$ as an item of scaled database, where $a \in W_p$. More explicitly, define

$$\mathcal{I}' := \{(p \geq a) \mid p \in \mathcal{I}, a \in W_p\}$$

and, for a weighted itemset (I, ω), put

$$(I, \omega)' := \{(p \geq a) \mid p \in I, a \in W_p, \omega(p) \geq a\} \subseteq \mathcal{I}'.$$

The scaled database $sc(\mathcal{D})$ is the sequence $(I_1, \omega_1)', \ldots, (I_M, \omega_M)'$. Then the following holds true. This is why we adopt the ordinal scale stated above.

Lemma 3.15. *1.* $(I, \omega) \preceq (J, \omega') \Leftrightarrow (I, \omega)' \subseteq (J, \omega')'$.
2. (I, ω) *is a closed weighted itemset of* $\mathcal{D} \Leftrightarrow (I, \omega)'$ *is a closed itemset of* $sc(\mathcal{D})$.

Proof. (1) Straightforward. (2) Immediately follows from (1). □

We illustrate an example of scaling using an ordinal scale on Table 5. In this example, a closed weighted itemset $(\{A, B\}, (A \mapsto 1, B \mapsto 1))$ of original database corresponds to a closed itemset $\{A \geq 1, B \geq 1\}$ of scaled database, for instance.

Table 5. An example of scaling

ID	A	B
1	1	2
2	3	1
3	2	

⟶

ID	$(A \geq 1)$	$(A \geq 2)$	$(A \geq 3)$	$(B \geq 1)$	$(B \geq 2)$
1	×			×	×
2	×	×	×	×	
3	×	×			

4 Algorithm

In this section we give an algorithm for mining closed weighted itemsets. Let $\mathcal{I} = \{p_1, \ldots, p_N\}$ and $\mathcal{D} = (I_1, \omega_1), \ldots, (I_M, \omega_M)$.

Our algorithm consists of three parts: **CWM_main**, **CWM_recursive** and **GenOD**. In the following algorithms we define that, for a weighted itemset (I, ω), $\omega(p) = 0$ if and only if $p \notin I$, for convenience.

Here we use occurrence delivers. For a given weighted itemset (I, ω), put $T_{I,\omega} := \{j \mid (I_j, \omega_j) \succeq (I, \omega)\}$. Now define

$$OD_{I,\omega,k}(i) := \begin{cases} \{(j, w) \mid j \in T_{I,\omega}, w = \omega_j(p_i) - \omega(p_i) > 0\} & (\text{if } i \leq k) \\ \{j \mid j \in T_{I,\omega}, \omega_j(p_i) - \omega(p_i) > 0\} & (\text{otherwise}), \end{cases}$$

for a given $k \in I$.

Remark 4.1. If $OD_{I,\omega,k}(i) \neq \emptyset$ then p_i can be added to (I, ω) to create a new weighted itemset (I', ω') that occurs in some transactions of $\mathcal{D}$. More explicitly, suppose $OD_{I,\omega,k}(i) \neq \emptyset$. Let $(OD_{I',\omega',i}, T)$ be the output of **GenOD**$(OD_{I,\omega,k}, i)$ that stated later. Then $\tau(I', \omega') = OD_{I,\omega,k}(i)$.

The *occurrence deliver* $OD_{I,\omega,k}$ associated with the weighted itemset (I, ω) and the current index k is the pair $(Head, Tail)$, where $Head$ is the list of $OD_{I,\omega,k}(i)$'s $(i \leq k)$ and $Tail$ is the set

$$\{OD_{I,\omega,k}(i) \mid i > k, \nexists i' \neq i, i' > k \text{ such that } OD_{I,\omega,k}(i) \subset OD_{I,\omega,k}(i')\}.$$

We distinguish $Tail$ from $Head$ because $Tail$ is not used except for pruning. We present an example of occurrence deliver on Table 6. For the sake of simplifying initialize, we define

$$OD_0(i) := \{(j, w) \mid i \in I_j, \omega_j(i) = w\},$$
$$OD_{\emptyset,\omega_0,N+1} := (\{OD_0(i) \mid i \in \mathcal{I}\}, \emptyset),$$

where ω_0 is defined by $\omega_0(p) = 0$ for all $p \in \mathcal{I}$.

Table 6. An example of occurrence deliver

$\mathcal{D}$

ID	p_1	p_2	p_3	p_4	p_5
1	1	2	4		
2	2	3		2	
3	2			3	2
4		3	5		1
5	2	1	3		1
6	1		5		
7		3	3		3

$OD_{I,\omega,3}$
where $I = \{3\}, \omega(p_3) = 3$

p_1	p_2	p_3	p_5
$(1,1)$	$(1,2)$	$(1,1)$	4
$(5,2)$	$(4,3)$	$(4,2)$	5
$(6,1)$	$(5,1)$	$(6,2)$	7
	$(7,3)$		

The algorithm starts with **CWM_main**.

```
Procedure: CWM_main
Input: Numerical transaction database $\mathcal{D}$, and the set $\mathcal{S}$ of
        all closed itemsets of $S(\mathcal{D})$;
Output: The set $C$ of closed weighted itemsets of $\mathcal{D}$;
begin
    $C \leftarrow \emptyset$;
    calculate $OD_{\emptyset,\omega_0,N+1}$ from $\mathcal{D}$;
    for $k' = 1, \ldots, N$ do
        CWM_recursive($OD_{\emptyset,\omega_0,N+1}, k', C, \mathcal{S}$);
    output $C$;
end.
```

CWM_recursive recursively increments the value of an item of the weighted itemset (I, ω) and create (I', ω'). We call (I', ω') the *current weighted itemset* at this step. Then this module decides whether (I', ω') is closed weighted itemset that has to be listed at this step or not, with using occurrence deliver.

```
Procedure: CWM_recursive
Input: Occurrence deliver $OD_{I,\omega,k}$ with weighted itemset $(I, \omega)$
        and index $k$, index $k' \leq k$, $C$ and $\mathcal{S}$;
begin
    $(OD_{I',\omega',k'}, T) \leftarrow$ GenOD($OD_{I,\omega,k}, k'$);
    if $\exists i > k'$ such that $OD_{I',\omega',k'}(i) = T$ then return;
    if $\nexists i < k'$ such that $\{j \mid (j, w) \in OD_{I',\omega',k'}(i)\} = T$ then
        add $(I', \omega')$ to $C$;
    for $i = 1, \ldots, k' - 1$ do
        if there is $\tilde{I} \in \mathcal{S}$ such that $\tilde{I} \supseteq I' \cup \{p_i\}$ and
        $\max\{i_1 \mid p_{i_1} \in \tilde{I}\} = \max\{i_1 \mid p_{i_1} \in I\}$ then
            CWM_recursive($OD_{I',\omega',k'}, i, C, \mathcal{S}$);
    if $OD_{I',\omega',k'}(k') \neq \emptyset$ then CWM_recursive($OD_{I',\omega',k'}, k', C, \mathcal{S}$);
end.
```

GenOD generates occurrence delivers.

```
Procedure: GenOD
Input: Occurrence deliver $OD_{I,\omega,k}$
        and the index $k' \leq k$ of item that will be added to $(I,\omega)$;
Output: Occurrence deliver $OD_{I',\omega',k'}$ and a set of transaction IDs $T$;
begin
    $T(i) \leftarrow \begin{cases} \{j \mid (j,w) \in OD_{I,\omega,k}(i)\} \ (i \leq k) \\ OD_{I,\omega,k}(i) \ \text{(otherwise)}; \end{cases}$
    $m \leftarrow \min\{w \mid (j,w) \in OD_{I,\omega,k}(k')\}$;
    $I' \leftarrow I \cup \{p_{k'}\}$;
    define $\omega'$ by $\omega'(p) = \omega(p)\ (p \neq p_{k'}),\ \omega'(p_{k'}) = \omega(p_{k'}) + m$;
    for $i = k'+1, \ldots, N$ do $OD_{I',\omega',k'}(i) \leftarrow T(i) \cap T(k')$;
    $OD_{I',\omega',k'}(k') \leftarrow \{(j, w-m) \mid (j,w) \in OD_{I,\omega,k}(k'), w-m > 0\}$;
    for $i = 1, \ldots, k'$ do $OD_{I',\omega',k'}(i) \leftarrow \{(j,w) \in OD_{I,\omega,k}(i) \mid j \in T(k')\}$;
    $Head' \leftarrow \{OD_{I',\omega',k'}(i) \mid i \leq k'\}, Tail' \leftarrow \{OD_{I',\omega',k'}(i) \mid i > k'\}$;
    for $i = k'+1, \ldots, N$ do
        if $\exists i' \neq i, i' > k'$ such that $OD_{I',\omega',k'}(i) \subset OD_{I',\omega',k'}(i')$ then
            remove $OD_{I',\omega',k'}(i)$ from $Tail'$;
    output $OD_{I',\omega',k'} = (Head', Tail')$ and $T(k')$;
end.
```

If the condition judgment at the step 6 of **CWM_recursive** is neglected, then the algorithm can mine closed weighted itemsets without using information of closed itemsets of $S(\mathcal{D})$. We say the algorithm with using $S(\mathcal{D})$'s closed itemsets **CWM**, and the algorithm without using them **CWM_woc**. In the next section we test these algorithms and the case using scaling, and compare their results.

It is easy to show that **CWM** stops within finite steps. Now we verify that **CWM** is sound.

Theorem 4.2. *Let C be the output of* **CWM** *and $\mathcal{CWI}$ be the set of all closed weighted itemsets of $\mathcal{D}$. Then $C = \mathcal{CWI}$.*

Proof. Let $(OD_{I',\omega',k'}, T)$ be an output of **GenOD**(I,ω,k). T is the set $\{j \mid (I_j, \omega_j) \succeq (I',\omega')\}$, that is, $\tau(I',\omega')$. If (I',ω') is not closed, then there is $(I'',\omega'') \succ (I',\omega')$ such that $\tau(I'',\omega'') = \tau(I',\omega')$. Let p_i be an item that $\omega''(p_i) > \omega'(p_i)$. If $i > k'$, then $OD_{I',\omega',k'}(i) = T$ (see Remark 4.1). Similarly, if $i < k'$, then $\{j \mid (j,w) \in OD_{I',\omega',k'}(i)\} = T$. Now i can not equal to k', since $\{j \mid (j,w) \in OD_{I',\omega',k'}(k')\} \subset T$. Thus (I',ω') is not outputted. Thus, $C \subseteq \mathcal{CWI}$. Conversely, it is easy to show that if the current weighted itemset (I',ω') is closed, then (I',ω') must be outputted. Therefore it suffices to show that, for every $(J,\varpi) \in \mathcal{CWI}$, (J,ϖ) becomes the current weighted itemset at a certain step.

Let $J = \{p_{i_1}, \ldots, p_{i_r}\}$, $i_1 < \ldots < i_r$. We denote procedure **CWM_recursive** with argument $k' = i$ by $CR(i)$. Let $w(i_r) = \{\omega_j(i_r) \mid i_r \in I_j\}$ and let $w_1 < w_2 < \ldots$ be the elements of $w(i_r)$. There is q such that $w_q = \omega(i_r)$, otherwise (I, ω) can not be closed. Hence, if we call $CR(i_r)$ in **CWM_main** and then call $CR(i_r)$ $q-1$ times recursively, then we have the current weighted itemset $(I^{(r)}, \omega^{(r)}) = (\{p_{i_r}\}, (p_{i_r} \mapsto \omega(i_r)))$. (Note that $OD_{I,\omega,i_r}(i_r) \neq \emptyset$ in each step.) Now, $OD_{I^{(r)},\omega^{(r)},i_r}(i_{r-1})$ is not empty. So we call $CR(i_{r-1})$ a certain times and have the current weighted itemset $(I^{(r-1)}, \omega^{(r-1)}) = (\{p_{i_{r-1}}, p_{i_r}\}, (p_{i_{r-1}} \mapsto \omega(i_{r-1}), p_{i_r} \mapsto \omega(i_r)))$. Then continue this process and finally we have the current weighted itemset $(I^{(1)}, \omega^{(1)}) = (J, \varpi)$. $\square$

Remark 4.3. Let $(OD_{I',\omega',k'}, T)$ be an output of **GenOD**(I, ω, k). Then $\sharp T$ is the frequency of (I', ω'). By utilize this fact, one can modify slightly **CWM** in order to mine closed weighted itemsets of their frequency are more than a given threshold.

5 Experiments

We have implemented both of **CWM** and **CWM_woc** by C++ and have practiced some experiments. In **CWM**, the algorithm first reads the data file of closed itemsets of $S(\mathcal{D})$, and then it holds the data of closed itemsets in the form of prefix tree. **CWM** searches the tree of closed itemsets in depth-first order, and sets weights on the closed itemsets incrementally. Information of closed itemsets of $S(\mathcal{D})$ is also used to simplify the process of generating occurrence delivers in **CWM**.

Test Environment:

- Core 2 Quad Q9400 @2.66GHz
- 2GB Memory
- Windows XP sp3

Experiment 1
First we test our algorithm for sparse datasets. Here we choose Bag of Words - KOS blog entries dataset in UCI Machine Learning Repository[1] and restrict it to 50 – 700 transactions so that we can observe changes. Table 7 shows the statistics of datasets. Here the number of nonzero elements means $\sum\{\sharp I \mid (I, \omega) :$ transaction $\}$, that is intuitively the number of cells filled by some (nonzero) numbers.

For each case we mine all closed weighted itemsets by three methods: (1) scale given databases and execute **LCM** [7], (2) **CWM_woc** and (3) **CWM**. In order to provide the list of closed itemsets of the support database we use **LCM**. In the case (3) we suppose that the list is given, so we exclude the computation time of **LCM** for the support (we record it in the row (**LCM** for support)). Similarly, we exclude the time for scaling from (1) and the time for making the support database from (3) (these times are much smaller than the mining times).

The results are illustrated in Table 8 and Figure 1.

---

[1] `http://archive.ics.uci.edu/ml/`

Table 7. Statistics of databases used in Experiment 1

Number of	**50**	**100**	**200**	**300**	**400**	**500**	**600**	**700**
Original database:								
items	2,136	3,114	4,382	5,135	5,579	5,924	6,154	6,310
transactions	50	100	200	300	400	500	600	700
nonzero elements	4,853	9,586	18,937	28,906	38,478	48,687	58,961	69,458
clo. itemsets	3,909	22,180	112,072	324,362	609,634	1,214,875	1,984,404	3,107,003
clo. w. itemsets	4,593	27,712	157,639	462,816	862,628	1,781,811	2,996,046	4,789,732
Scaled database:								
items	3,226	4,808	7,253	9,104	10,389	11,575	12,421	13,172
nonzero elements	6,535	12,665	24,898	38,096	50,577	64,253	77,612	91,437

Table 8. Results of Experiment 1

Computation time[sec]

	50	**100**	**200**	**300**	**400**	**500**	**600**	**700**
LCM+Scaling	0.689	4.330	27.971	155.504	404.015	1,025.628	h/o	h/o
LCM for support	0.406	1.838	9.160	48.427	112.670	319.983	655.222	h/o
CWM	0.293	1.476	7.711	22.421	42.934	89.922	148.934	*237.574
CWM_woc	0.348	1.758	8.963	25.981	52.718	106.367	184.009	271.182

h/o: heap overflow *: use CWM_woc for mining closed itemsets of the support database

(a)

Number of times of generating occurrence deliver

	50	**100**	**200**	**300**	**400**	**500**	**600**	**700**
CWM	15,181	90,430	456,550	1,419,549	2,624,743	5,438,729	9,295,676	15,046,947
CWM_woc	149,044	700,580	2,983,753	8,978,451	16,575,045	30,848,991	49,601,432	77,181,597

(b)

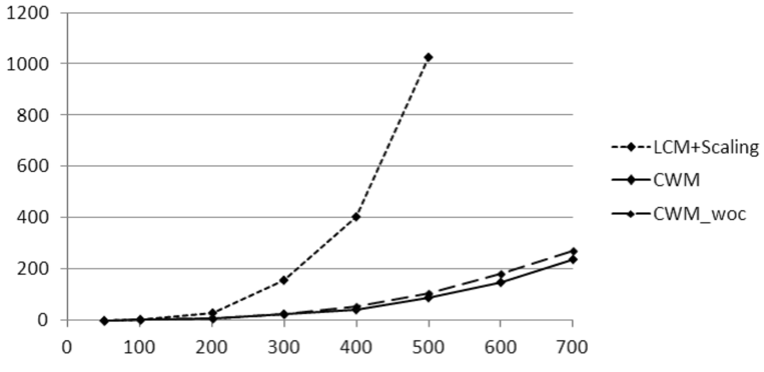


Fig. 1. Graph of the computation times of Experiment 1

As a result, **CWM** and **CWM_woc** are much faster than **LCM**+Scaling. In addition, **CWM** is around fifteen percents faster than **CWM_woc**. Table 8(b) indicates the numbers of times of generating occurrence deliver. By using the tree of closed itemsets of $S(\mathcal{D})$, **CWM** reduces unnecessary attempts considerably.

Experiment 2

Next we test for dense datasets. For dense datasets, **CWM** seems to be not faster than **CWM_woc** for dense datasets, because the set of closed itemsets of $S(\mathcal{D})$ of dense dataset has less information. Here we use Bolts (**BL**), Airport (**AP**) and Low Birth Weight (**LW**) datasets from the Bilkent Repository[2]. We present the statistics of datasets on Table 9. The results are in Table 10 and Figure 2.

Table 9. Statistics of databases used in Experiment 2

Original databases

Number of	**BL**	**AP**	**LW**
items	8	5	10
transactions	40	135	189
nonzero elements	300	672	1,048
clo. itemsets	5	5	38
clo. w. itemsets	5,659	18,243	48,829

Scaled databases

Number of	**BL**	**AP**	**LW**
items	134	672	247
nonzero elements	2,808	45,496	21,652

Table 10. Results of Experiment 2

Computation time[sec]

	BL	**AP**	**LW**
LCM+Scaling	0.069	1.272	1.141
LCM for support	≈0	≈0	≈0
CWM	0.172	0.495	0.894
CWM_woc	0.166	0.477	0.890

(a)

Number of times of generating occurrence deliver

	BL	**AP**	**LW**
CWM	17,889	42,402	59,689
CWM_woc	17,898	42,405	59,697

(b)

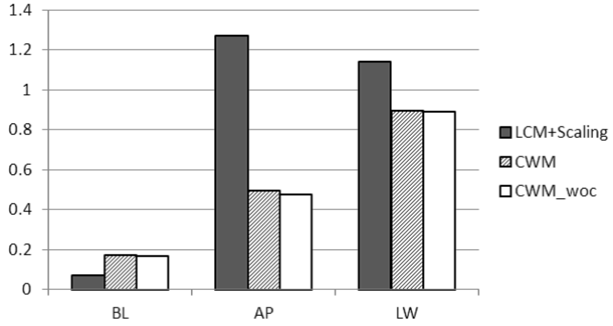


Fig. 2. Graph of the computation times of Experiment 2

---

[2] `http://funapp.cs.bilkent.edu.tr/`

This time **CWM** takes more computation time than **CWM_woc** as we expected. As for comparison to **LCM**+Scaling, it depends on dataset which method is faster.

6 Conclusion

We have introduced the notion of closed weighted itemsets and attempted to mine them. In mining we compare our method to the way using scaling. Furthermore, we investigate an influence of availability of closed itemsets of the support of a given numerical transaction database, by applying two mining algorithms, **CWM** uses the closed itemsets and **CWM_woc** does not use them. We experimented our algorithms for sparse and dense datasets. As a result, for a given sparse dataset, using closed itemsets of the support of the given database can reduce the computation time for mining closed weighted itemsets, and each way is much faster than the method using scaling. Of course, it takes time for mining closed itemsets of support database to apply **CWM**, so the total computation time may exceed the computation time of **CWM_woc**. However, there still remains an advantage that **CWM** can divide the problem of mining closed weighted itemsets into two parts. Thus we think that using closed itemsets of support database has a certain usefulness for mining closed weighted itemsets.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207–216 (1993)
2. Han, J., Pei, J., Yin, Y., Mao, R.: Mining frequent patterns without candidate generation: A frequent-pattern tree approach. Data Mining and Knowledge Discovery 8, 53–87 (2004)
3. Ganter, B., Wille, R.: Applications of combinatorics and graph theory to the biological and social sciences. The IMA Volumes in Mathematics and Its Applications, vol. 17, pp. 139–167 (1989)
4. Kaytoue, M., Kuznetsov, S.O., Napoli, A.: Revisiting numerical patterns with formal concept analysis. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence, pp. 1342–1347 (2011)
5. Chan, R., Yang, Q., Shen, Y.-D.: Mining high utility itemsets. In: Proceedings of the Third International Conference on Data Mining, pp. 19–26 (2003)
6. Pasquier, N., Bastide, Y., Taouil, R., Lakhal, L.: Discovering Frequent Closed Itemsets for Association Rules. In: Beeri, C., Bruneman, P. (eds.) ICDT 1999. LNCS (LNAI), vol. 1540, pp. 398–416. Springer, Heidelberg (1998)
7. Uno, T., Asai, T., Uchida, Y., Arimura, H.: An Efficient Algorithm for Enumerating Closed Patterns in Transaction Databases. In: Suzuki, E., Arikawa, S. (eds.) DS 2004. LNCS (LNAI), vol. 3245, pp. 16–31. Springer, Heidelberg (2004)

On Computing Tractable Variations of Unordered Tree Edit Distance with Network Algorithms*

Yoshiyuki Yamamoto[1], Kouichi Hirata[2], and Tetsuji Kuboyama[3]

[1] Graduate School of Computer Science and Systems Engineering
yamamoto@dumbo.ai.kyutech.ac.jp
[2] Department of Artificial Intelligence
Kyushu Institute of Technology
Kawazu 680-4, Iizuka 820-8502, Japan
hirata@dumbo.ai.kyutech.ac.jp
[3] Computer Center, Gakushuin University
Mejiro 1-5-1, Toshima, Tokyo 171-8588, Japan
ori-alsip2011@tk.cc.gakushuin.ac.jp

Abstract. The problem of computing the standard edit distance between unordered trees is known to be intractable. To circumvent this hardness result, several tractable variations have been proposed. The algorithms of these variations include the submodule of a network algorithm, either the minimum cost maximum flow algorithm or the maximum weighted bipartite matching algorithm. In this paper, we point out that these network algorithms are replaceable, and give the experimental results of computing these variations with both network algorithms.

1 Introduction

Comparing tree-structured data such as HTML and XML data in web mining or DNA and glycan data for bioinformatics is one of the important tasks for data mining. In this paper, we formulate such data as *rooted labeled unordered trees* (*trees*, for short) and then focus on distance measures between trees.

The most famous distance measure between trees is the *edit distance* [10]. The edit distance in the standard definition is formulated as the minimum cost to transform from a tree to another tree by applying *edit operations* of a *substitution*, a *deletion* and an *insertion* to trees. It is known that the edit distance is closely related to the notion of a Tai *mapping* [10], which is a one-to-one node correspondence between trees preserving ancestor relations. The minimum cost of possible mappings coincides with the edit distance [10].

The problem of computing the edit distance for *unordered trees* is intractable, that is, NP-hard [16] and MAX SNP-hard [4,15]. In order to overcome such

---

* This work is partially supported by Grand-in-Aid for Scientific Research 20500126, 20240014, 21500145, 22240010 and 23300061 from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 211–223, 2012.

computational inefficiency with approximating the edit distance well, many variations of the standard edit distance such as the *top-down* (or the *degree*-1) distance [2,9,12], the *accordant* distance [5,6], the *degree*-2 (or the *Lu's*, the *LCA-preserving*) distance [7,17] and the *constrained* (or the *isolated-subtree*) distance [11,13,14] have been developed by restricting mappings such as *top-down*, *accordant*, *degree*-2 and *constrained mappings*, respectively. Then, we can compute both top-down and constrained distances in $O(n^2 D \log_2 D)$ time [12,14], while the degree-2 distance in $O(n^2 \sqrt{d} \log_2 d)$ time [17], where D is the maximum degree and d is the minimum degree of given two trees. The time complexity of computing the accordant distance for unordered trees is unknown so far.

The difference between the above time complexity follows from the adoption of different network algorithms (*cf.*, [1,3]) in order to compute distances between forests. While the algorithms for computing both top-down and constrained distances [12,14] adopt the *minimum cost maximum flow algorithm*, the algorithm for computing the degree-2 distance [17] adopts the *maximum weighted bipartite matching algorithm*.

In this paper, we point out that both network algorithms can be applicable to computing all the variations, including an accordant distance. Hence, the time complexity of computing all the variations is $O(n^2 \sqrt{d} \log_2 d)$. Furthermore, we give experimental results for all the variations and both two network algorithms.

2 Unordered Tree Edit Distance

A *tree* is a connected graph without cycles. For a tree $T = (V, E)$, we denote V and E by $V(T)$ and $E(T)$, respectively. We sometimes denote $v \in V(T)$ by $v \in T$. A *rooted tree* is a tree with one node r chosen as its *root*.

For each node v in a rooted tree with the root r, let $UP_r(v)$ be the unique path from v to r. The *parent* of $v (\neq r)$, denoted by $p(v)$, is its adjacent node on $UP_r(v)$ and the *ancestors* of $v (\neq r)$ are the nodes on $UP_r(v) - \{v\}$. We denote $u \leq v$ if v is an ancestor of u, which we denote by $u < v$, or $u = v$. The parent and the ancestors of the root r are undefined. We say that u is a *child* of v if v is the parent of u, and u is a *descendant* of v if v is an ancestor of u. Two nodes with the same parent are called *siblings*. Also we say that w is the *least common ancestor* of u and v, denoted by $u \sqcup v$, if $u \leq w$, $v \leq w$ and there exists no node w' such that $u \leq w'$ and $v \leq w'$. A *leaf* is a node having no children. The *degree* $d(v)$ of a node v is the number of the children of v, and the *degree* $d(T)$ of T is the maximum degree for every node in T. For a tree T and a node $v \in T$, we define the *complete subtree* $T[v]$ of T rooted by v as a tree such that $V(T[v]) = \{u \in V(T) \mid u \leq v\}$ and $E(T[v]) = \{(u, w) \in E(T) \mid u, w \in V(T[v])\}$.

We say that a rooted tree is *ordered* if a left-to-right order among siblings is given; *Unordered* otherwise. Also we say that a tree is *labeled* over Σ if each node is assigned a symbol from a fixed finite alphabet Σ, where we denote the label of a node v by $l(v)$. We sometimes identify v with $l(v)$. In this paper, we call a rooted labeled unordered tree over Σ a *tree*, simply. We call an ordered sequence of trees a *forest*. Also we denote the forest obtained by deleting v from $T[v]$ by $F[v]$. We denote an empty tree or forest by $\emptyset$.

Definition 1 (Edit operations). Let T be a tree. Then, we call the following three operations *edit operations*. Also see Figure 1.

1. *Substitution*: Change the label of the node v in T (from l_1 to l_2).
2. *Deletion*: Delete a non-root node v in T (labeled by l_1) with a parent v' (labeled by l'), making the children of v become the children of v'. The children are inserted in the place of v as a subset of the children of v'.
3. *Insertion*: The complement of deletion. Insert a node v (labeled by l_2) as a child of v' (labeled by l') in T making v the parent of a subset of the children of v'.

For a special *blank* symbol $\varepsilon \notin \Sigma$, let $\Sigma_\varepsilon = \Sigma \cup \{\varepsilon\}$. Then, we represent each edit operation by $l_1 \mapsto l_2$, where $(l_1, l_2) \in (\Sigma_\varepsilon \times \Sigma_\varepsilon - \{(\varepsilon, \varepsilon)\})$. The operation is a substitution if $l_1 \neq \varepsilon$ and $l_2 \neq \varepsilon$, a deletion if $l_2 = \varepsilon$, and an insertion if $l_1 = \varepsilon$.

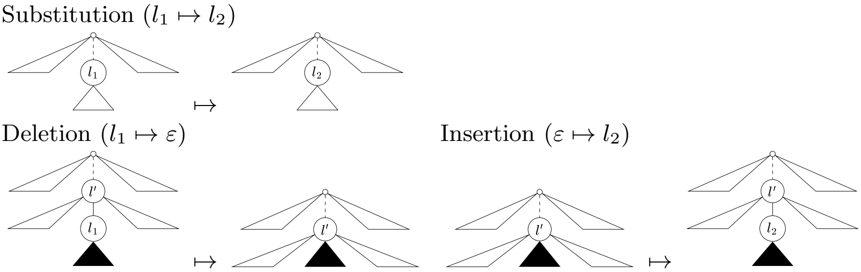


Fig. 1. Edit operations for trees

We define a *cost function* $\gamma : (\Sigma_\varepsilon \times \Sigma_\varepsilon - \{(\varepsilon, \varepsilon)\}) \mapsto \mathbf{R}$ on pairs of labels. For a cost function γ, we define the *cost* of an edit operation by setting $\gamma(l_1 \mapsto l_2) = \gamma(l_1, l_2)$. The *cost* of a sequence $S = s_1, \ldots, s_k$ of edit operations is given by $\gamma(S) = \sum_{i=1}^{k} \gamma(s_i)$.

Definition 2 (Edit distance). Let T_1 and T_2 be trees and γ a cost function. Then, an *edit distance* $\tau(T_1, T_2)$ between T_1 and T_2 under γ is defined as follows:

$$\tau(T_1, T_2) = \min \left\{ \gamma(S) \;\middle|\; \begin{array}{l} S \text{ is a sequence of edit operations} \\ \text{transforming from } T_1 \text{ to } T_2 \end{array} \right\}.$$

The edit distance is closely related to the following mapping [10].

Definition 3 (Mapping). Let T_1 and T_2 be trees. Then, we say that the triple (M, T_1, T_2) is a *Tai mapping* (or a *mapping* simply) between T_1 and T_2 if $M \subseteq V(T_1) \times V(T_2)$ and every pair (v_1, w_1) and (v_2, w_2) in M satisfies the following conditions.

1. $v_1 = v_2$ iff $w_1 = w_2$ (one-to-one condition).
2. $v_1 \leq v_2$ iff $w_1 \leq w_2$ (ancestor condition).

We will use M instead of (M, T_1, T_2) when there is no confusion.

Let M be a mapping between T_1 and T_2. Also let I_1 (*resp.*, I_2) be the set of nodes in T_1 (*resp.*, T_2) but not in M. Then, the *cost* $\gamma(M)$ of M is given as $\sum_{(v,w)\in M} \gamma(l(v), l(w)) + \sum_{v\in I_1} \gamma(l(v), \varepsilon) + \sum_{w\in I_2} \gamma(\varepsilon, l(w))$.

Theorem 1 (Tai [10]). *For trees T_1 and T_2, the following statement holds.*

$$\tau(T_1, T_2) = \min\{\gamma(M) \mid M \text{ is a mapping between } T_1 \text{ and } T_2\}.$$

3 Tractable Variations of Unordered Tree Edit Distance

In this section, we introduce the tractable variations of τ, which are formulated as the minimum cost of the restricted mappings as the same form of Theorem 1.

Definition 4 (Restricted mappings). Let $T_i = (V_i, E_i)$ be a tree with the root r_i $(i = 1, 2)$ and $M \subseteq V_1 \times V_2$ a mapping between T_1 and T_2.

1. M is a *constrained mapping* [13,14] if M satisfies that:
   $\forall(v_1, w_1), (v_2, w_2), (v_3, w_3) \in M$ $[v_3 < v_1 \sqcup v_2 \iff w_3 < w_1 \sqcup w_2]$.
2. M is an *accordant mapping* [5,6] if M satisfies that:
   $\forall(v_1, w_1), (v_2, w_2), (v_3, w_3) \in M$
   $[v_1 \sqcup v_2 = v_1 \sqcup v_3 \iff w_1 \sqcup w_2 = w_1 \sqcup w_3]$.
3. M is a *degree-2 mapping* [7,17] if M satisfies that:
   $\forall(v_1, w_1), (v_2, w_2) \in M$ $[(v_1 \sqcup v_2, w_1 \sqcup w_2) \in M]$.
4. M is a *top-down mapping* [2,9] if M satisfies that:
   $\forall(v, w) \in M \backslash \{(r_1, r_2)\}$ $[(p(v), p(w)) \in M]$.

The above restricted mappings provide the following implications [5,11], but none of the inverse implications hold in general.

M: top-down mapping $\Rightarrow$ M: degree-2 mapping
$\Rightarrow$ M: accordant mapping $\Rightarrow$ M: constrained mapping
$\Rightarrow$ M: Tai mapping

Definition 5 (Variations of τ). For trees T_1 and T_2, a *constrained distance* $\tau_{\downarrow}(T_1, T_2)$, an *accordant distance* $\tau_{\leftrightarrow}(T_1, T_2)$, a *degree-2 distance* $\tau_{\sqcup}(T_1, T_2)$ and a *top-down distance* $\tau_{\top}(T_1, T_2)$ are defined as the minimum cost of a constrained, an accordant, a degree-2 and a top-down mappings between T_1 and T_2.

In the remainder of this paper, we sometimes denote the variations of $\tau_{\downarrow}$, $\tau_{\leftrightarrow}$, $\tau_{\sqcup}$ and $\tau_{\top}$ by using the form $\tau_{\circ}$ $(\circ \in \{\downarrow, \leftrightarrow, \sqcup, \top\})$.

By the above implications of restricted mappings, the distances of τ and $\tau_{\circ}$ $(\circ \in \{\downarrow, \leftrightarrow, \sqcup, \top\})$ between T_1 and T_2 under a fixed (and an arbitrary) cost function provide the following sequence.

$$\tau(T_1,T_2) \leq \tau_{\downarrow}(T_1,T_2) \leq \tau_{\leftrightarrow}(T_1,T_2) \leq \tau_{\sqcup}(T_1,T_2) \leq \tau_{\top}(T_1,T_2).$$

In order for every distance of τ and $\tau_\circ$ ($\circ \in \{\downarrow,\leftrightarrow,\sqcup,\top\}$) to be a *metric*, it is necessary for a cost function γ to be a *metric*, that is, $\gamma(l_1,l_2) \geq 0$, $\gamma(l_1,l_1)=0$, $\gamma(l_1,l_2)=\gamma(l_2,l_1)$ and $\gamma(l_1,l_3) \leq \gamma(l_1,l_2)+\gamma(l_2,l_3)$. Also, under a cost function that is a metric, it holds that $\tau_{\sqcup}(T_1,T_2) = \tau_{\leftrightarrow}(T_1,T_2)$ [5,6]. Hence, the distances of τ and $\tau_\circ$ ($\circ \in \{\downarrow,\leftrightarrow,\sqcup,\top\}$) between T_1 and T_2 under a fixed cost function that is a metric provide the following sequence.

$$\tau(T_1,T_2) \leq \tau_{\downarrow}(T_1,T_2) \leq \tau_{\leftrightarrow}(T_1,T_2) = \tau_{\sqcup}(T_1,T_2) \leq \tau_{\top}(T_1,T_2).$$

Example 1. Consider the mappings M_i ($1 \leq i \leq 5$) concerned with the trees T_j ($1 \leq j \leq 6$) described in Figure 2, where the dashed line denotes the element of the mapping. Also consider a *unit cost function* μ such that $\mu(l_1 \mapsto l_2) = \mu(l_1 \mapsto \varepsilon) = \mu(\varepsilon \mapsto l_2) = 1$ and an *indel cost function* ι such that $\iota(l_1 \mapsto l_2) = 3$ and $\iota(l_1 \mapsto \varepsilon) = \iota(\varepsilon \mapsto l_2) = 1$ for $l_1, l_2 \in \Sigma$ and $l_1 \neq l_2$. Note that, while μ is a metric, ι is not.

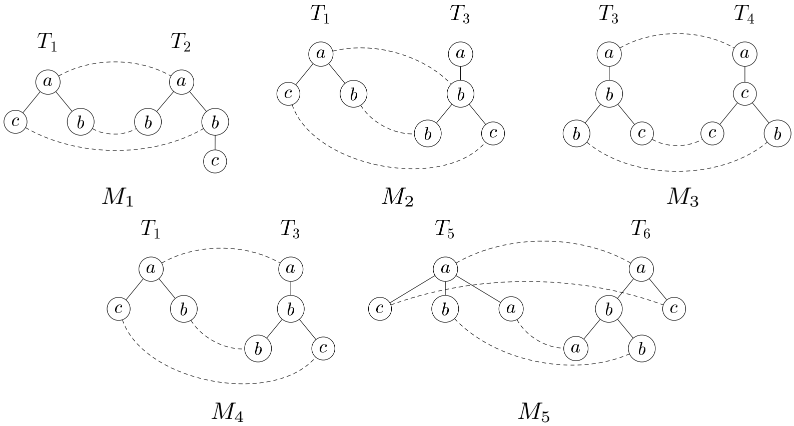


Fig. 2. Mappings M_i ($1 \leq i \leq 5$)

The mapping M_1 is a top-down mapping. Hence, it is also a degree-2, an accordant, a constrained and a Tai mapping. For T_1 and T_2 under μ, it holds that $\tau_{\top}(T_1,T_2) = 2$ and $\tau_{\sqcup}(T_1,T_2) = \tau_{\leftrightarrow}(T_1,T_2) = \tau_{\downarrow}(T_1,T_2) = \tau(T_1,T_2) = 1$ (because of $(M_1 - \{(c,b)\}) \cup \{(c,c)\}$). In this case, M_1 is the minimum cost mapping for $\tau_{\top}$.

The mapping M_2 is not a top-down mapping but a degree-2 mapping. Hence, it is also an accordant, a constrained and a Tai mapping. For T_1 and T_3 under μ, it holds that $\tau_{\top}(T_1,T_3) = 3$, $\tau_{\sqcup}(T_1,T_3) = \tau_{\leftrightarrow}(T_1,T_3) = 2$ and $\tau_{\downarrow}(T_1,T_3) = \tau(T_1,T_3) = 1$ (because of M_4). In this case, M_2 is the minimum cost mapping for $\tau_{\sqcup}$ and $\tau_{\leftrightarrow}$.

The mapping M_3 is not a degree-2 mapping but an accordant mapping. Note that the definition of an accordant mapping does not require that a mapping M always contains $v_1 \sqcup v_2$ and $w_1 \sqcup w_2$. Hence, it is also a constrained and a Tai mapping. For T_3 and T_4 under ι, it holds that $\tau_{\top}(T_3, T_4) = \tau_{\sqcup}(T_3, T_4) = 3$ (because of $M_3 \cup \{(b, c)\}$) and $\tau_{\leftrightarrow}(T_3, T_4) = \tau_{\downarrow}(T_3, T_4) = \tau(T_3, T_4) = 2$. In this case, M_3 is the minimum cost mapping for $\tau_{\leftrightarrow}$, $\tau_{\downarrow}$ and τ. On the other hand, under μ, it holds that $\tau(T_3, T_4) = \tau_{\circ}(T_3, T_4) = 1$ by using a mapping $M_3 \cup \{(b, c)\}$.

The mapping M_4 is not an accordant mapping, because $c \sqcup a = a = c \sqcup b$ in T_1 but $c \sqcup a = a \neq b = c \sqcup b$ in T_3. On the other hand, it is a constrained and a Tai mapping. For T_1 and T_3 under μ, it holds that $\tau_{\top}(T_1, T_3) = 3$, $\tau_{\sqcup}(T_1, T_3) = \tau_{\leftrightarrow}(T_1, T_3) = 2$ (because of M_2) and $\tau_{\downarrow}(T_1, T_3) = \tau(T_1, T_3) = 1$. In this case, M_4 is the minimum cost mapping for $\tau_{\downarrow}$ and τ.

The mapping M_5 is not a constrained mapping but a Tai mapping. For T_5 and T_6 under μ, it holds that $\tau_{\circ}(T_5, T_6) = 3$ and $\tau(T_5, T_6) = 1$. In this case, M_5 is the minimum cost mapping for τ.

4 Algorithms for Computing the Variations

Figure 3 illustrates the algorithms for computing the variations of an unordered tree edit distance τ, that is, a constrained distance $\tau_{\downarrow}$, an accordant distance $\tau_{\leftrightarrow}$, a degree-2 distance $\tau_{\sqcup}$ and a top-down distance $\tau_{\top}$, according to [12,14,17].

The algorithm for computing a top-down distance $\tau_{\top}$ [12] is based on the algorithm for computing a constrained distance $\tau_{\downarrow}$ [14]. On the other hand, according to the algorithm for computing an *ordered* accordant distance [5,6], we newly design the algorithm for computing an (*unordered*) accordant distance $\tau_{\leftrightarrow}$ based on the algorithm for computing a degree-2 distance $\tau_{\sqcup}$ [17].

The algorithms in Figure 3 adopt two different network algorithms, that is, the *minimum cost maximum flow algorithm* [1] for computing $\tau_{\downarrow}$ and $\tau_{\top}$ and *the maximum weighted bipartite matching algorithm* [3] for computing $\tau_{\sqcup}$ and $\tau_{\leftrightarrow}$.

When computing `MinCostMaxFlow`(v, w) as the formula (A) in Figure 3, it is necessary to compute the *minimum cost maximum flow* (*MCMF*, for short) [12,14] in a network N_F with the source s and the sink t interleaving a complete bipartite graph $G_F = ((V \cup \{\lambda_1\}) \cup (W \cup \{\lambda_2\}), E_F)$. Here, V is the set of children of v in T_1, W is the set of children of w in T_2 and λ_i $(i = 1, 2)$ is a dummy node representing an empty tree $\emptyset$. For $v' \in V$ and $w' \in W$, the cost of (v', w') is $\tau_{\circ}(T_1[v'], T_2[w'])$, the costs of (v', λ_2) and (λ_1, w') are respectively $|T_1[v']|$ and $|T_2[w']|$, and the costs of all other edges are 0. All the edges have capacity 1 except (s, λ_1), (λ_1, λ_2) and (λ_2, t) whose capacities are $d(w)$, $|d(v) - d(w)|$ and $d(v)$. Then, the value of (A) is equal to the MCMF of N_F.

Example 2. Consider the trees S and T in Figure 4 (left), and suppose that $\tau_{\circ}(S_i, T_j)$, $\tau_{\circ}(\lambda_1, T_j)$ and $\tau_{\circ}(S_i, \lambda_2)$ $(i = 1, 2, 3; j = 1, 2)$ is given as Figure 4 (right). Then, we explain how to compute the value of the formula (A) as the distance between forests (S_1, S_2, S_3) and (T_1, T_2), by using the MCMF algorithm (for $\tau_{\downarrow}$ and $\tau_{\top}$).

$\tau_{\downarrow}$ (constrained distance)

$$\begin{cases} \tau_{\downarrow}(T_1[v], T_2[w]) \\ \qquad = \min \left\{ \begin{array}{l} \tau_{\downarrow}(\emptyset, T_2[w]) + \min\limits_{1 \le j \le d(w)} \{\tau_{\downarrow}(T_1[v], T_2[w_j]) - \tau_{\downarrow}(\emptyset, T_2[w_j])\}, \\ \tau_{\downarrow}(T_1[v], \emptyset) + \min\limits_{1 \le i \le d(v)} \{\tau_{\downarrow}(T_1[v_i], T_2[w]) - \tau_{\downarrow}(T_2[v_i], \emptyset)\}, \\ \tau_{\downarrow}(F_1[v], F_2[w]) + \gamma(v \mapsto w). \end{array} \right\}. \\ \tau_{\downarrow}(F_1[v], F_2[w]) \\ \qquad = \min \left\{ \begin{array}{l} \tau_{\downarrow}(\emptyset, F_2[w]) + \min\limits_{1 \le j \le d(w)} \{\tau_{\downarrow}(F_1[v], F_2[w_j]) - \tau_{\downarrow}(\emptyset, F_2[w_j])\}, \\ \tau_{\downarrow}(F_1[v], \emptyset) + \min\limits_{1 \le i \le d(v)} \{\tau_{\downarrow}(F_1[v_i], F_2[w]) - \tau_{\downarrow}(F_2[v_i], \emptyset)\}, \\ \texttt{MinCostMaxFlow}(v, w) \dotfill \text{(A)} \end{array} \right\}. \end{cases}$$

$\tau_{\sqcup}$ (degree-2 distance)

$$\begin{cases} \tau_{\sqcup}(T_1[v], T_2[w]) \cdots \text{same as } \tau_{\downarrow}(T_1[v], T_2[w]) \\ \tau_{\sqcup}(F_1[v], F_2[w]) \\ \qquad = \sum\limits_{1 \le i \le d(v)} \tau_{\sqcup}(T_1[v_i], \emptyset) + \sum\limits_{1 \le j \le d(w)} \tau_{\sqcup}(\emptyset, T_2[w_j]) - \sum\limits_{(v,w) \in \texttt{BM}} \omega((v, w)). \ \ldots \text{(B)} \end{cases}$$

$\tau_{\leftrightarrow}$ (accordant distance)

$$\begin{cases} \tau_{\leftrightarrow}(T_1[v], T_2[w]) \\ \qquad = \min \left\{ \begin{array}{l} \tau_{\leftrightarrow}(\emptyset, T_2[w]) + \min\limits_{1 \le j \le d(w)} \{\tau_{\leftrightarrow}(T_1[v], T_2[w_j]) - \tau_{\leftrightarrow}(\emptyset, T_2[w_j])\}, \\ \tau_{\leftrightarrow}(T_1[v], \emptyset) + \min\limits_{1 \le i \le d(v)} \{\tau_{\leftrightarrow}(T_1[v_i], T_2[w]) - \tau_{\leftrightarrow}(T_2[v_i], \emptyset)\}, \\ \tau_{\leftrightarrow}(F_1[v], F_2[w]) + \gamma(v \mapsto w), \\ \tau_{\leftrightarrow}(F_1[v], F_2[w]) + \gamma(v \mapsto \varepsilon) + \gamma(\varepsilon \mapsto w). \end{array} \right\}. \\ \tau_{\leftrightarrow}(F_1[v], F_2[w]) \cdots \text{same as } \tau_{\sqcup}(F_1[v], F_2[w]) \dotfill \text{(B)} \end{cases}$$

$\tau_{\top}$ (top-down distance)

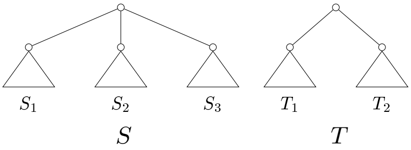
$$\begin{cases} \tau_{\top}(T_1[v], T_2[w]) \\ \qquad = \tau_{\top}(\emptyset, T_2[w]) + \min\limits_{1 \le j \le d(w)} \{\tau_{\top}(T_1[v], T_2[w_j]) - \tau_{\top}(\emptyset, T_2[w_j])\}. \\ \tau_{\top}(F_1[v], F_2[w]) = \texttt{MinCostMaxFlow}(v, w) \dotfill \text{(A)} \end{cases}$$

Fig. 3. The algorithms for computing the variations $\tau_{\downarrow}$ [14], $\tau_{\sqcup}$ [17], $\tau_{\leftrightarrow}$ and $\tau_{\top}$ [12] of an unordered tree edit distance τ

In using the MCMF algorithm for the formula (A), we construct the network N_F described as Figure 5 (left). Here, the cost of an edge (S_i, T_j) is $\tau_{\circ}(S_i, T_i)$. For example, the costs of (S_1, T_1), (S_2, T_1), (S_3, T_1) and (λ_1, T_1) are 5, 7, 10 and 5, respectively. Also the capacities of (s, λ_1), (λ_1, λ_2) and (λ_2, t) are 2, 1 and 3, respectively.

By using the MCMF algorithm, we obtain the minimum cost maximum flow of N_F described as Figure 5 (right). Then, we obtain the value of the formula (A) as the cost of the minimum cost maximum flow of N_F that $3 \times 1 + 7 \times 1 + 5 \times 1 + 0 \times 2 = 15$.

The time complexity of the MCMF algorithm is $O(fm \log_2 n)$ [1], where f is the maximum flow. In our setting, since $n = d(v) + d(w) + 4$, $m = d(v)d(w) + 2d(v) + 2d(w) + 3$ and $f = d(v) + d(w)$, we can rewrite the time complexity as $O(d(v)d(w)(d(v) + d(w)) \log_2(d(v) + d(w)))$. Hence, Zhang [14] has shown that the time complexity of computing $\tau_{\downarrow}(T_1, T_2)$ is:



$\tau_\circ(S_i,T_j)$	λ_1	S_1	S_2	S_3
λ_2	0	3	9	11
T_1	5	5	7	10
T_2	6	3	6	5

Fig. 4. Trees S and T (left) and $\tau_\circ(S_i,T_j)$ (right)

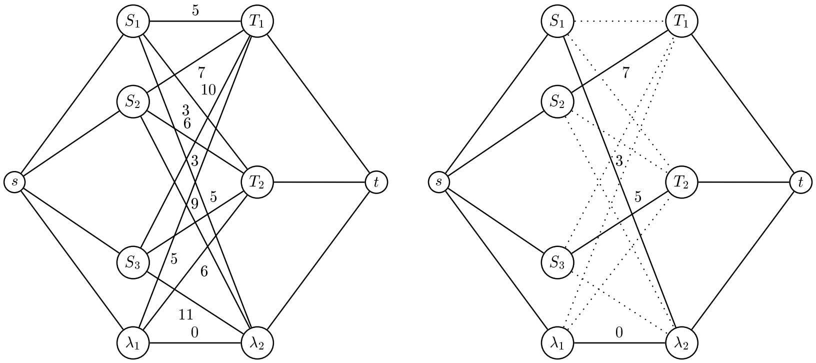


Fig. 5. The network N_F for trees S and T in Figure 4 (left) and its minimum cost maximum flow (right)

$$O(|T_1||T_2|(d(T_1)+d(T_2))\log_2(d(T_1)+d(T_2))),$$

that is, $O(n^2 D\log_2 D)$, where $n=\max\{|T_1|,|T_2|\}$ and $D=\max\{d(T_1),d(T_2)\}$. The time complexity of computing $\tau_\top(T_1,T_2)$ [12] is same as above.

On the other hand, when computing $\sum_{(v,w)\in\mathtt{BM}}\omega((v,w))$ in the formula (B) in Figure 3, it is enough to compute the *maximum weighted bipartite matching* (*MWBM*, for short) `BM` (with weight ω) in a complete bipartite graph $G_M=(V\cup W,E_M)$ [17]. Here, V is the set of children of v in T_1, W is the set of children of w in T_2 and the weight of an edge $e=(v',w')\in E_M$ is set to $\tau_\sqcup(T_1[v'],\emptyset)+\tau_\sqcup(\emptyset,T_2[w'])-\tau_\sqcup(T_1[v'],T_2[w'])$.

Example 3. For the same trees S and T in Example 2, we explain how to compute the value of the formula (B) as the distance between forests (S_1,S_2,S_3) and (T_1,T_2), by using the MWBM algorithm (for $\tau_\sqcup$ and $\tau_\leftrightarrow$).

Consider the formula (B) to use the MWBM algorithm. Then, we construct the complete bipartite graph G_M described as Figure 6 (left). Here, the weight of an edge (S_i,T_j) is given as $\tau_\circ(S_i,\emptyset)+\tau_\circ(\emptyset,T_i)-\tau_\circ(S_i,T_i)$. For example, the weights of (S_1,T_1), (S_2,T_1) and (S_3,T_1) are $3+5-5=3$, $9+5-7=7$ and $11+5-10=6$, respectively.

By using the MWBM algorithm, we obtain the maximum weighted bipartite matching of G_M illustrated in Figure 6 (right). Then, we obtain the value of the

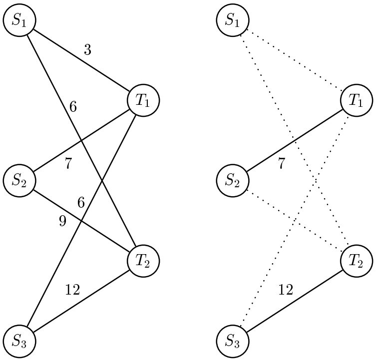


Fig. 6. The complete bipartite graph G_M for trees S and T in Figure 4 (left) and its maximum weighted bipartite matching (right)

formula (B) as the sum of the distance between a forest and an empty forest minus the distance of the maximum weighted bipartite matching of G_M that $(3+9+11)+(5+6)-(7+12)=15$.

The time complexity of the MWBM algorithm is $O(\sqrt{n}m\log_2(nW))$ [3], where W is the maximum weight. When $d(v) < d(w)$, the MWBM of (V, E) is equal to one of (V', E'), where E' consists of $d(v)$ edges in E whose weight is larger than any edge in $E - E'$ and V' is a set of vertices adjacent to E'. In our setting, since $n = d(v) + d(v)^2$, $m = d(v)^2$ and W is bounded by cn for a constant $c > 0$, we can rewrite the time complexity as $O(\sqrt{d(v)+d(v)^2}d(v)^2\log_2(d(v)+d(v)^2)) = O(d(v)d(w)\sqrt{d(v)}\log_2 d(v))$, where $d(v)^2 \leq d(w)$ and $d(v)^2 \leq d(v)d(w)$. Hence, Zhang *et al.* [17] have shown that the time complexity of computing $\tau_{\sqcup}(T_1, T_2)$ is:

$$O(|T_1||T_2|\sqrt{\min\{d(T_1), d(T_2)\}}\log_2(\min\{d(T_1), d(T_2)\})),$$

that is, $O(n^2\sqrt{d}\log_2 d)$, where $d = \min\{d(T_1), d(T_2)\}$. The time complexity of computing $\tau_{\leftrightarrow}(T_1, T_2)$ is the same as above.

The following theorem claims that the formula (A) and the formula (B) in the algorithms in Figure 3 are replaceable, as presented in Example 2 and 3.

Theorem 2. *In the algorithms in Figure 3, the value of the formula* (A) *is same as one of* (B)*.*

Proof. Consider the case in computing $\tau_{\circ}(F_1[v], F_2[w])$. Let v_i $(1 \leq i \leq d(v))$ and w_j $(1 \leq j \leq d(w))$ be the child of v and w, respectively. Also let $V = \{v_1, \ldots, v_{d(v)}\}$ and $W = \{w_1, \ldots, w_{d(w)}\}$. Then, the formula (A) adopts a network N_F including a complete bipartite graph $G_F = ((V \cup \{\lambda_1\}) \cup (W \cup \{\lambda_2\}), E_F)$ and the formula (B) adopts a complete bipartite graph $G_M = (V \cup W, E_M)$. The cost of the edge $(v_i, w_j) \in G_F$ is $\tau_{\circ}(T_1[v_i], T_2[w_j])$ and the weight of the edge $(v_i, w_j) \in G_M$ is $\tau_{\circ}(T_1[v_i], \emptyset) + \tau_{\circ}(\emptyset, T_2[w_j]) - \tau_{\circ}(T_1[v_i], T_2[w_j])$. Furthermore, let vl(A) and vl(B) the value of the formula (A) and (B), respectively.

Suppose that $F \subseteq (V \cup \{\lambda_1\}) \times (W \cup \{\lambda_2\})$ is the set of edges in E_F selected by some maximum flow of N_F. Furthermore, we denote the restriction of F in $V \times W$ by $F|_{V \times W}$. Then, we can evaluate $vl(\mathrm{A})$ as follows, where the equality holds when F is the minimum cost maximum flow of N_F.

$$\begin{aligned}
vl(\mathrm{A}) \leq \tau_{\circ}(\emptyset, \emptyset) &+ \sum_{(v,w) \in F|_{V \times W}} \tau_{\circ}(T_1[v], T_2[w]) \\
&+ \sum_{(v,\lambda_2) \in F - F|_{V \times W}} \tau_{\circ}(T_1[v], \emptyset) + \sum_{(\lambda_1,w) \in F - F|_{V \times W}} \tau_{\circ}(\emptyset, T_2[w]).
\end{aligned}$$

On the other hand, suppose that $M \subseteq V \times W$ is the set of edges in E_M selected by some weighted bipartite matching of G_M. Furthermore, we denote $\{v \in V \mid (v,w) \in M\}$ (*resp.*, $\{w \in W \mid (v,w) \in M\}$) by V_M^+ (*resp.*, W_M^+) and $V - V_M^+$ (*resp.*, $W - W_M^+$) by V_M^- (*resp*, W_M^-). Then, by substituting the weight of the edge $(v,w) \in E_M$, we can evaluate $vl(\mathrm{B})$ as follows, where the equality holds when M is the maximum weighted bipartite matching of G_M.

$$\begin{aligned}
vl(\mathrm{B}) &\leq \sum_{v \in V} \tau_{\circ}(T_1[v], \emptyset]) + \sum_{w \in W} \tau_{\circ}(\emptyset, T_1[w]) - \sum_{(v,w) \in M} \omega((v,w)) \\
&= \sum_{v \in V} \tau_{\circ}(T_1[v], \emptyset]) + \sum_{w \in W} \tau_{\circ}(\emptyset, T_1[w]) \\
&\quad - \sum_{(v,w) \in M} \{\tau_{\circ}(T_1[v], \emptyset) + \tau_{\circ}(\emptyset, T_2[w]) - \tau_{\circ}(T_1[v], T_2[w])\} \\
&= \sum_{v \in V} \tau_{\circ}(T_1[v], \emptyset]) + \sum_{w \in W} \tau_{\circ}(\emptyset, T_1[w]) \\
&\quad - \sum_{v \in V_M^+} \tau_{\circ}(T_1[v], \emptyset) - \sum_{w \in W_M^+} \tau_{\circ}(\emptyset, T_2[w]) + \sum_{(v,w) \in M} \tau_{\circ}(T_1[v], T_2[w]) \\
&= \sum_{(v,w) \in M} \tau_{\circ}(T_1[v], T_2[w]) + \sum_{v \in V_M^-} \tau_{\circ}(T_1[v], \emptyset) + \sum_{w \in W_M^-} \tau_{\circ}(\emptyset, T_1[w]).
\end{aligned}$$

If F is the minimum cost maximum flow of N_F, then the following sequence holds by setting a bipartite matching M' of G_M to $F|_{V \times W}$, the set $V_{M'}^-$ of vertices to $\{v \in V \mid (v, \lambda_2) \in F - F|_{V \times W}\}$ and the set $W_{M'}^-$ of vertices to $\{w \in W \mid (\lambda_1, w) \in F - F|_{V \times W}\}$, where $\tau_{\circ}(\emptyset, \emptyset) = 0$.

$$\begin{aligned}
vl(\mathrm{A}) &= \tau_{\circ}(\emptyset, \emptyset) + \sum_{(v,w) \in F|_{V \times W}} \tau_{\circ}(T_1[v], T_2[w]) \\
&\quad + \sum_{(v,\lambda_2) \in F - F|_{V \times W}} \tau_{\circ}(T_1[v], \emptyset) + \sum_{(\lambda_1,w) \in F - F|_{V \times W}} \tau_{\circ}(\emptyset, T_2[w]) \\
&= \sum_{(v,w) \in M'} \tau_{\circ}(T_1[v], T_2[w]) + \sum_{v \in V_{M'}^-} \tau_{\circ}(T_1[v], \emptyset) + \sum_{w \in W_{M'}^-} \tau_{\circ}(\emptyset, T_1[w]). \\
&\geq vl(\mathrm{B}).
\end{aligned}$$

On the other hand, if M is the maximum weighted bipartite matching of G_M, then the following sequence holds by setting the maximum flow F' of N_F to $M \cup \{(v, \lambda_2) \mid v \in V_M^-\} \cup \{(\lambda_1, w) \mid w \in W_M^-\}$, where $\tau_\circ(\emptyset, \emptyset) = 0$.

$$\begin{aligned} vl(\mathrm{B}) &= \sum_{(v,w) \in M} \tau_\circ(T_1[v], T_2[w]) + \sum_{v \in V_M^-} \tau_\circ(T_1[v], \emptyset) + \sum_{w \in W_M^-} \tau_\circ(\emptyset, T_1[w]) \\ &= \tau_\circ(\emptyset, \emptyset) + \sum_{(v,w) \in F'|_{V \times W}} \tau_\circ(T_1[v], T_2[w]) \\ &\quad + \sum_{(v,\lambda_2) \in F' - F'|_{V \times W}} \tau_\circ(T_1[v], \emptyset) + \sum_{(\lambda_1,w) \in F' - F'|_{V \times W}} \tau_\circ(\emptyset, T_2[w]) \\ &\geq vl(\mathrm{A}). \end{aligned}$$

Hence, it holds that $vl(\mathrm{A}) = vl(\mathrm{B})$. □

Corollary 1 (Zhang *et al.* [**17**]**).** *We can compute* $\tau_\circ(T_1, T_2)$ $(\circ \in \{\downarrow, \leftrightarrow, \sqcup, \top\})$ *in* $O(n^2\sqrt{d}\log_2 d)$ *time, where* $n = \max\{|T_1|, |T_2|\}$ *and* $d = \min\{d(T_1), d(T_2)\}$.

5 Experimental Results

In this section, we give an experimental results for the algorithms for computing $\tau_\circ(T_1, T_2)$ $(\circ \in \{\downarrow, \leftrightarrow, \sqcup, \top\})$ under the unit cost function (*cf.*, Example 1). Here, we implement the algorithms based on a Hungarian method [1], which runs in $O(V^3)$ time, so V is equal to $d(T_1) + d(T_2)$ for the MCMF algorithm and $\max\{d(T_1), d(T_2)\}$ for the MWBM algorithm.

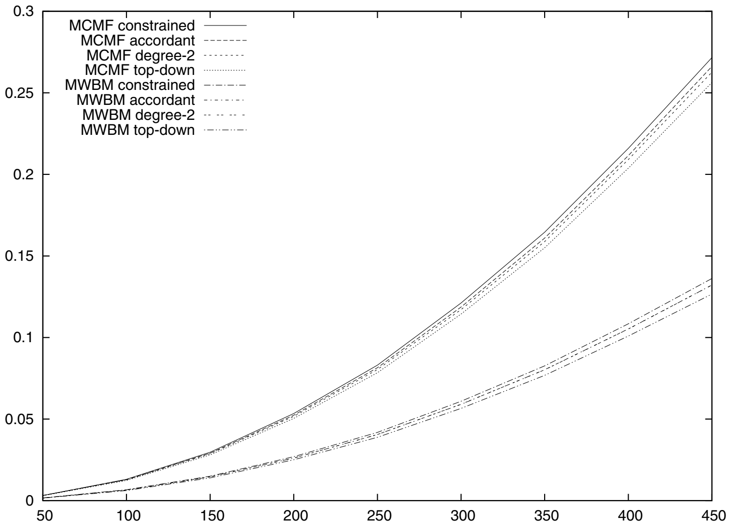


Fig. 7. The running time of computing $\tau_\circ$ $(\circ \in \{\downarrow, \leftrightarrow, \sqcup, \top\})$

Figure 7 shows the running time of $\tau_\circ$ ($\circ \in \{\downarrow, \leftrightarrow, \sqcup, \top\}$) for 4950 pairs of 100 randomly generated trees by the algorithm PTC [8], where the number of nodes varies from 50 to 450, the maximum degree is 20 and the number of labels is 26. Then, the algorithms based on the MWBM algorithm are faster than ones based on the MCMF algorithm and, under the same network algorithm, the algorithms for computing $\tau_\top$, $\tau_\sqcup$, $\tau_\leftrightarrow$ and $\tau_\downarrow$ are faster in this order.

6 Conclusion

In this paper, we have analyzed the adoption of two network algorithms, the minimum cost maximum flow and the maximum weighted bipartite matching algorithms, in computing the tractable variations of an unordered tree edit distance. Then, we have shown that both network algorithms are replaceable to computing all the variations, with designing an algorithm of computing an *accordant distance*. Hence, we have reduced the time complexity of computing all the variations to $O(n^2\sqrt{d}\log_2 d)$ time. Furthermore, we have given experimental results for all the variations and both two network algorithms.

It is a future work to apply the variations to other practical data in order to evaluate the effect of the variations. Furthermore, it is also a future work to introduce another distance that approximates an unordered tree edit distance or its variations and is more efficient.

Acknowledgment. The authors would like to thank the anonymous reviewers of ALSIP'11 for their valuable comments to revise this paper.

References

1. Ahuja, R.K., Magnanti, T.L., Orlin, J.B.: Network flows. Prentice Hall (1993)
2. Chawathe, S.S.: Comparing hierarchical data in external memory. In: Proc. VLDB 1999, pp. 90–101 (1999)
3. Gabow, H.N., Tarjan, R.E.: Faster scaling algorithms for network problems. SIAM J. Comput. 18, 1013–1036 (1989)
4. Hirata, K., Yamamoto, Y., Kuboyama, T.: Improved MAX SNP-Hard Results for Finding an Edit Distance between Unordered Trees. In: Giancarlo, R., Manzini, G. (eds.) CPM 2011. LNCS, vol. 6661, pp. 402–415. Springer, Heidelberg (2011)
5. Kuboyama, T.: Matching and learning in trees, Ph.D thesis, University of Tokyo (2007), `http://tk.cc.gakushuin.ac.jp/doc/kuboyama2007phd.pdf`
6. Kuboyama, T., Shin, K., Kashima, H.: Flexible tree kernels based on counting the number of tree mappings. In: Proc. MLG 2006, pp. 61–72 (2006)
7. Lu, S.-Y.: A tree-to-tree distance and its application to cluster analysis. IEEE Trans. Pattern Anal. Mach. Intell. 1, 219–224 (1979)
8. Luke, S., Panait, L.: A survey and comparison of tree generation algorithms. In: Proc. GECCO 2001, pp. 81–88 (2001)
9. Selkow, S.M.: The tree-to-tree editing problem. Inform. Process. Lett. 6, 184–186 (1977)
10. Tai, K.-C.: The tree-to-tree correction problem. J. ACM 26, 422–433 (1979)

11. Wang, J.T.L., Zhang, K.: Finding similar consensus between trees: An algorithm and a distance hierarchy. Pattern Recog. 34, 127–137 (2001)
12. Wang, Y., DeWitt, D.J., Cai, J.-Y.: X-Diff: An effective change detection algorithm for XML documents. In: Proc. ICDE 2003, pp. 519–530 (2003)
13. Zhang, K.: Algorithms for the constrained editing distance between ordered labeled trees and related problems. Pattern Recog. 28, 463–474 (1995)
14. Zhang, K.: A constrained edit distance between unordered labeled trees. Algorithmica 15, 205–222 (1996)
15. Zhang, K., Jiang, T.: Some MAX SNP-hard results concerning unordered labeled trees. Inform. Process. Let. 49, 249–254 (1994)
16. Zhang, K., Statman, R., Shasha, D.: On the editing distance between unordered labeled trees. Inform. Process. Let. 42, 133–139 (1992)
17. Zhang, K., Wang, J., Shasha, D.: On the editing distance between undirected acyclic graphs. Int. J. Found. Comput. Sci. 7, 43–58 (1995)

Multimodality in Multispace Interaction (MiMI)

Mayumi Bono and Nobuhiro Furuyama

National Institute of Informatics, Japan

Preface

We held International Workshop on Multimodality in Multispace Interaction (MiMI), at Sunport Hall Takamatsu, Takamatsu City, Kagawa prefecture in Japan on December 1-2, 2011. The workshop was part of JSAI International Symposia on Artificial Intelligence (JSAI-isAI 2011) sponsored by the Japanese Society for Artificial Intelligence. All he papers collected here were presented at the workshop either as invited talks or as accepted papers . Incorporating discussions, comments, and questions, workshop presenters revised their papers and submit them to this proceedings . The submitted papers were peer-reviewed once again and three out of the eight papers were accepted in the end. Our special gratitude goes to the anonymous reviewers of the papers for their dedicated efforts to make very constructive and useful comments for the authors to make their papers more convincing and intriguing. Before we proceed to the papers themselves, we would like to introduce the readers to the aims and scope of MiMI 2011, by showing a piece of memo that we made in preparation of a proposal of the workshop to JSAI. Here it is:

Multiplicity in talk and its social surrounding involving tools, artifacts and technologies, material objects and their distributions have been a canonical concern in research on human communication. In the last couple of decades, in particular, there has been a growing interest in complex ways in which multimodality including vocal, visual, aural, and gestural apparatus of speaking is played out in seemingly simple activities such as an everyday conversation between two people. While research on multimodality has revealed fundamental aspects of human communication, how habituated ways of conversing can be transformed and thereby coordinated in technologically enhanced environments is not yet well understood. Drawing on approaches from conversation analysis, ethnographic studies, Human-Computer interactions, workplace studies, and data mining in real space (e.g., meeting mining), we discuss how multispace is managed in socially, temporally, and sequentially complex environments. Our foci on multispace include, but are not limited to computer-mediated interactions (e.g., video conference), media-mediated interaction (e.g., remote human-agent interaction supported by computer technologies), longitudinally established interactional spaces, and multiple cognitive images within a speaker's viewpoints for producing language and gesture (e.g., sign language and gestures in narrative discourses), In such multispace interactions, conventional orders of talking may be negotiated and re-organized in order to achieve coherent courses of interaction. We particularly pay attention to how multimodal resources are

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 224–225, 2012.

sequentially organized in order to make sense of the multispace interaction. Traditionally, research on sequential organization of talk has relied on turn-taking systems revealed in telephone conversations. The application of theoretical and analytical concepts derived from such audio-focused studies needs to be reconsidered because emerging multispace interactions employ not only audio, but also visual, gestural and spatial information. Recent research on multimodality has addressed the issues on multiplicity by emphasizing the importance of "peripheral" modes of communication such as gesture, gaze, and posture. However, such accumulative approach only reproduces the simplistic view of speech as a "dominant" modality in interaction. Instead of inclusively describing multiple modes of communication, our workshop focuses on processes in which speakers select and coordinate multimodal resources to achieve their interactional goals. We believe that the readers will find the collection of papers on "MiMI" all stimulating and thought provocative, and hopefully come up with their own ideas to approach this fascinating topic.

MiMI 2011 Organizers,
Mayumi Bono
Nobuhiro Furuyama

PC Members,
Katsuya Takanashi, Kyoto University, Japan
Chiho Sunakawa, University of Texas, Austin, USA
Hideaki Kuzuoka, University of Tsukuba, Japan
Yoshinori Kuno, Saitama University, Japan
Yasuyuki Sumi, Future University Hakodate, Japan
Yasuharu Den, Chiba University, Japan
Yasuhiro Katagiri, Future University Hakodate, Japan
Kazuhiro Otsuka, NTT Communication Science Loboratories, Japan
Junji Yamato, NTT Communication Science Loboratories, Japan
Toshihiro Takada, NTT Communication Science Loboratories, Japan
Kouhei Kikuchi, National Institute of Informatics, Japan
Hideyuki Nakanishi, Osaka University, Japan
Tomohiro Kuroda, Kyoto University, Japan
Fernando Ferri, IRPPS-CNR, Italy
Patrizia Grifoni, IRPPS-CNR, Italy
Frederic Andres, National Institute of Informatics, Japan

An Engineering Approach to Conversational Informatics*

Toyoaki Nishida

Graduate School of Informatics, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan
`nishida@i.kyoto-u.ac.jp`

Abstract. Conversational informatics is a field of research that focuses on investigating human behaviors and designing artifacts that can interact with people in a conversational fashion. The field draws on a foundation provided by artificial intelligence, natural language processing, speech and image processing, cognitive science, and conversation analysis. It aims to shed light on meaning creation and interpretation during conversations, in search of better methods of computer-mediated communication, human-computer interaction, and support for knowledge creation. In this article, I will highlight recent developments in an engineering approach to conversational informatics. I also address empathic agents as future challenges.

Keywords: Conversational Informatics, Intelligent Virtual Agents, Empathic Agents.

1 Conversational Informatics as an Interdisciplinary Study on Conversation

Conversation is everywhere around us. Conversation is the most natural and popular means for people to communicate with each other. People are not only proficient in expressing their thoughts and emotions in conversation but also quite skillful to catch the flow of the discourse and make sense of the verbal and nonverbal signals other participants produce in conversation. Conversation has been studied by numerous authors. Among others, Clark studied conversation, or *language use*, from the viewpoint of joint actions [1]. Conversation is used to do joint activities and conversation itself is an emergent joint action. Beneath the surface underlies a fairly sophisticated interaction mechanism consisting of levels of hierarchical signals, layers of discourse, and tracks of channels for social exchange to take place.

Conversational informatics is an interdisciplinary area of research [2]. It sheds light on computational aspects of language use through a sophisticated mechanism of the verbal/nonverbal interactions during conversations (Fig. 1). It involves broad issues

---

* This paper is an extended version of the paper presented under the same title at International Workshop on Multimodality in Multispace Interaction (MiMI), Sunport Hall Takamatsu, Takamatsu, December 1-2, 2011.

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 226–239, 2012.

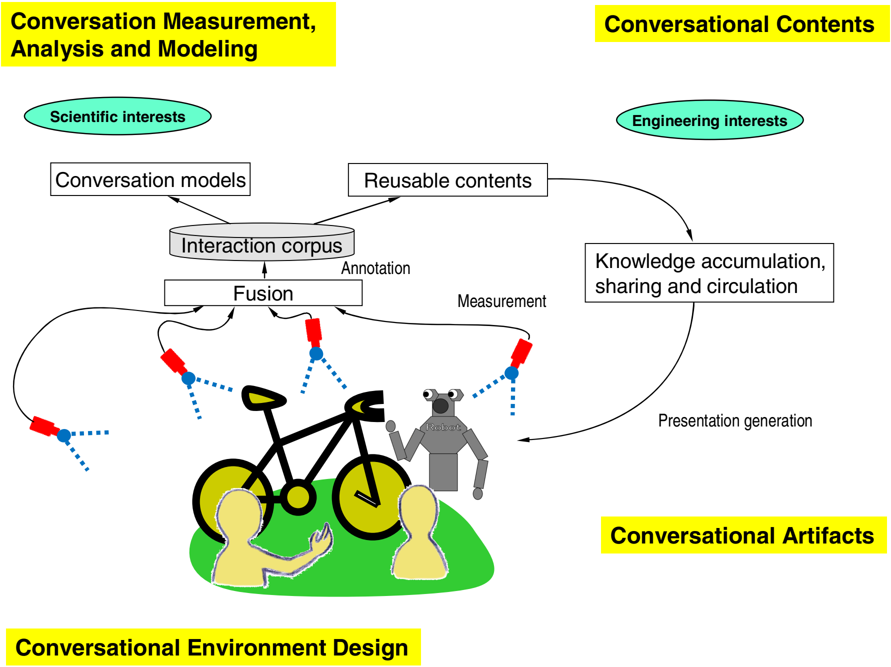


Fig. 1. Conceptual framework of conversational Informatics

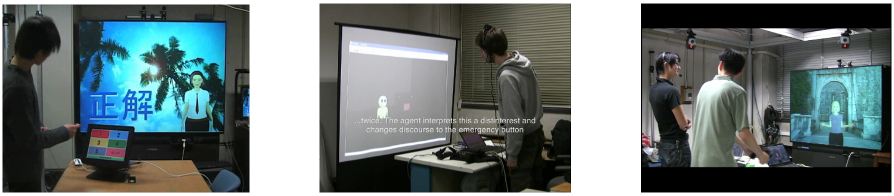
related to artificial intelligence, natural language processing, speech and image processing, cognitive science, and conversation analysis.

Conversational informatics consists of four subfields. The first subfield is about conversational artifacts that can participate in conversations with people. Our scope includes both interactive synthetic characters that show up on a computer screen and talk with the user and intelligent robots that can help the user in the physical environment by making conversation using not only natural language but also eye contact, facial expressions, gestures, or other nonverbal means of communication.

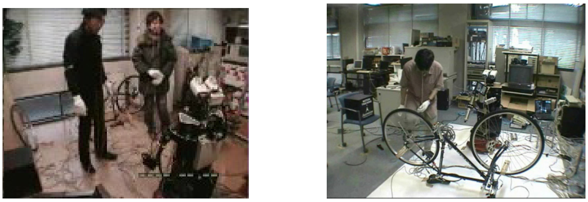
The second subfield is about conversational contents that can encapsulate information and knowledge arising in a conversational situation for reusing it in a new conversational situation. We aim at developing methods of capturing, accumulating, transforming, and applying conversational contents.

The third subfield is about conversation environment design whose goal is to build a complete space that can provide participants with proper resources in conversation to enable smooth and effective interaction. We address a method for sensing social signals in conversation to effectively help participants find necessary resources in conversation and retain the essences obtained in conversation for later use so that key information can be restored with supporting information.

The last subfield is about conversation measurement, analysis, and modeling. Motivated by scientific interest, we take a data-driven quantitative approach to



(a) Quiz agent [3] (b) VTSE: Virtual Technical Support Engineer [4] (c) Guide agent [5]



(d) The listener robot [6] (e) The presenter robot [6]

Fig. 2. Conversational artifacts we have developed so far

understanding conversational behaviors by measuring conversational behaviors using advanced sensing technologies and thereby we aim at building detailed quantitative models of conversation.

Synergy of the four subfields is critical, for it is hard to think about success in design without analysis, modeling without application, virtual humans without contents, contents without virtual humans, etc.

2 Conversational Artifacts

Conversational artifacts are autonomous systems that can communicate with people in a conversational fashion to pursue a given mission. Typical conversational artifacts are embodied conversational agents that exploit interactive synthetic characters to implement fluent communication between humans and computers and conversational robots that employ physical embodiment of robots, as shown in Fig. 2. A quiz agent entertains the user by giving quizzes. Our quiz agent [3] is attentive in the sense that it can monitor and control the conversation process by employing an attentive utterance policy that allows for whether, when, and to whom to make utterance, depending on the users' reactions. The virtual technical support engineer (VTSE) agent [4] provides a human user with technical information about a complex industrial device on demand. A virtual tour guide agent takes visitors on a virtual city tour to introduce its cultural heritage, history, and monuments, on demand [5]. The listener robot listens to the conversation among the users to acquire the task knowledge in a conversational fashion which will in turn be used as the knowledge source of the presenter robot that will provide the user with critical information for pursuing a task by estimating the user's information demand [6]. In each case, conversational communication with verbal and nonverbal cues is key to make the user interaction both easy and effective.

Development of a conversational artifact is not easy, for appropriate conversational behaviors are needed to be produced in real time by integrating multimodal input. The history of the development of conversational artifacts goes back to 1960's when natural language dialogue systems such as Eliza [7] or SHRDLU [8] were developed. It was followed by speech dialogue systems such as HEARSAY-II [9] and multimodal dialogue systems such as Put-that-there [10]. Use of synthetic characters to bestow the dialogue engine with embodiment was proposed by the Knowledge Navigator video [11], which encouraged the researchers to intensively employ nonverbal cues in human-agent interaction, resulting in the notion of embodied conversational agents [12]. Initially, script or mark-up languages were developed to program the behaviors of embodied conversational agents [13].

The contemporary methodology for the development of conversational artifacts might be called "behavior from observation", for it implies that a quantitative dialogue model is generated by observing exactly how people behave in face-to-face conversation, as reported in [14]. This methodology exploits recent advances in sensing technology that allows for capturing nonverbal communication behaviors that people exhibit consciously or unconsciously during conversation to build a quantitative model of exchanging propositions encompassing uncertainty in various situations or controlling the discourse of conversation.

Unfortunately, the behavior-from-observation approach does not work well when the physical features of the conversational artifact, shape and size in particular, significantly differ from those of humans, for such a difference often results in different communication patterns. In fact, we have found that people tend to use clear, emphasized, complete, and redundant expressions in human-robot interaction, as opposed to vague, subtle, incomplete, and parsimonious ones in human-human interaction [15]. A probable reason is that robotic agents are deemed not as competent as humans in communication and the common ground is not well-established.

In order to overcome the difficulties, two methods need to be employed. The first is the Wizard of Oz (WOZ) method to realizing the "human-in-the-artifacts" setting. The second is to develop a method of effectively producing the behaviors of the conversational artifact from the data collected in the WOZ experiments.

Our immersive WOZ environment (Fig. 3) provides the human operator with a feeling as if s/he stayed "inside" a conversational artifact to receive incoming visual and auditory signals and to create conversational behaviors in a natural fashion [16]. At the human-robot interaction site, a 360-degree camera is placed near the robot's head, which is used to acquire the image of all directions around it. The WOZ operator's cabin is surrounded by the cylindrical display, which is a set of large-sized displays that are circularly aligned. The current display system uses eight 64-inch display panels arranged in a circle with about 2.5 meters diameter. Eight surround speakers are used to reproduce the acoustic environment. The WOZ operator stands in the cylindrical display and controls the robot from there. The image around the robot is projected on an immersive cylindrical display around the WOZ operator. This setting gives the operator exactly the same view as the robot sees. When a scene is displayed on the full screen, it will provide a sense of immersion.

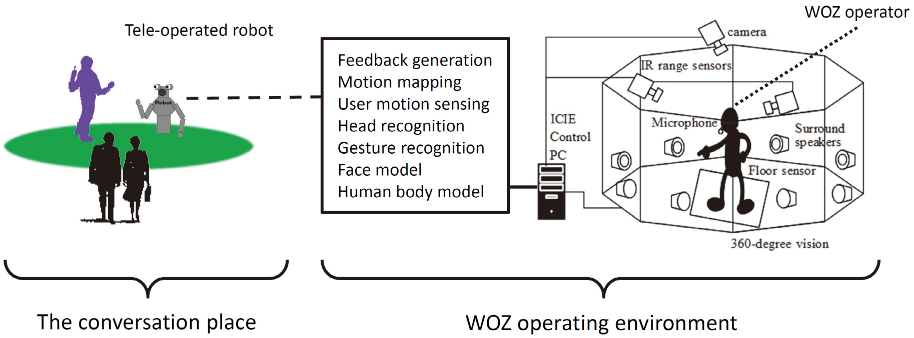


Fig. 3. Immersive WOZ environment

The WOZ operator's behavior, in turn, is captured by one or more range sensor to reproduce a mirrored behavior of the robot, which will take the same poses as the operator does by calculating the angles of the operator's joints at every frame. We can control the robot's head, shoulders, elbows, wrists, fingers, hip joints, knees, and ankles, and we think they are enough to represent basic actions in communication. The sound on each side of the WOZ operator is gathered by microphones and communicated via network so that everyone can hear the sound of the other side. Semi-autonomous control features are introduced in order to cope with differences in embodiment between humans and robots [17]. For example, as soon as the user's behavior is recognized as a pointing gesture, the optimal pointing behavior for the robot can be produced rather than simply mapping the trajectory of the user's behavior to the robot.

Learning by mimicking is a computational framework for producing the interactive behaviors of conversational artifacts from a collection of data obtained from the WOZ experiment. As shown in Fig. 4, the learning robot "watches" how people interact with each other, estimates how the target actor reacts according to the communicative behavior of the communication partner, and applies the acquired knowledge as estimated patterns of actions to the actual situations it encounters in conversation. Currently, the communicative behaviors the robot "observes" are approximated as a collection of continuous time series. We have developed a suite of unsupervised learning algorithms for this framework [18,19].

The learning algorithm consists of four stages:

1) the discovery stage on which the robot discovers the action and command space;
2) the association stage on which the robot associates discovered actions and commands generating a probabilistic model that can be used either for behavior understanding or generation;
3) the controller generation stage on which the behavioral model is converted into an actual controller to allow the robot to act in similar situations; and
4) the accumulation stage on which the robot combines the gestures and actions it learned from multiple interactions.

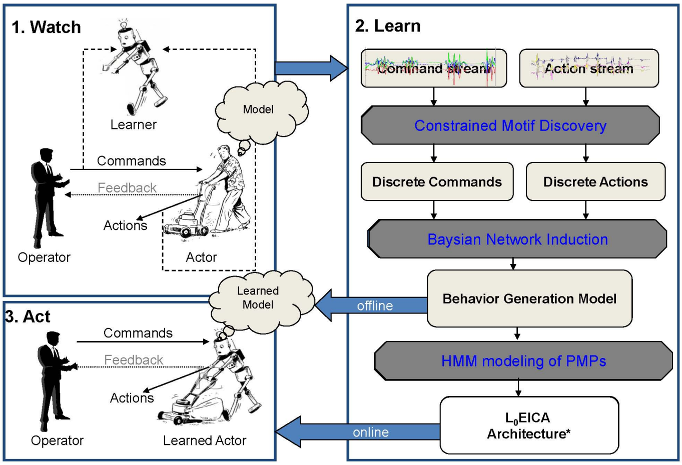


Fig. 4. A framework for learning by imitation [18]

So far, the learning by imitation framework has been applied only to nonverbal interactions. In the future, we plan to extend it to the entire communicative behaviors in which verbal and nonverbal communication are integrated with each other. Another future challenge involves realizing a fluid communication engine by integrating verbal and nonverbal communication skills and building a service agent by integrating fluid communication and high task performance engines.

3 Conversational Contents

Technologies for conversational contents aim at helping people create and manage a large amount of contents collected from conversations. Circulating conversational contents is critical to increase the quality. Public Opinion Channel (POC) is an early social-information processing system that continuously collects messages from people in a community and feeds edited messages back to them [20]. Central to circulating conversational contents is conversation quantization which is a computational framework of circulating conversation quanta that encapsulate discourse units into annotated audio-visual video segments. Conversation quantization is based on the idea of approximating a continuous flow of conversation by a series of minimally coherent segments of discourse called conversation quanta [21] (Fig. 5). A conversation quantum contains a segment of verbal and nonverbal interactions together with contextual information. Conversation quanta can be captured in a conversation for use as an “essence of conversation” in later conversation.

Vickey [22] is an augmented conversational environment for a driving simulator. It can ground the conversation on the events observed through the simulated window of the vehicle, by analyzing pointing gestures of the participants. A simple sensing technique is used to capture the conversational situations.

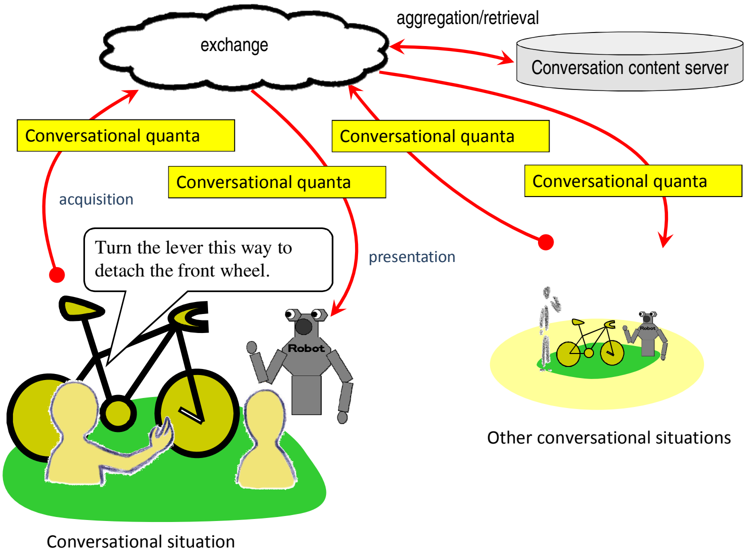


Fig. 5. Conversational framework of conversation quantization

Merckel and Nishida developed a system that can associate conversation quanta on varying places in the environment. An augmented reality system allows for retrieving the spatial coordinates in the real world from a corresponding two-dimensional image point [23]. Special emphasis is placed to minimize the overhead for building the spatial model of the environment and operations, for if it involved extra work, it would not be sustainable in a practical way.

The future challenge includes fully automated conversation quanta capture and application that allows for preserving and restoring essence of conversation together with proper contextual information.

4 Conversational Environment Design

The goal of conversational environment design is to realize a smart environment that permits people to pursue effective knowledge creation through conversations. Approaches vary depending on how many participants are involved, whether the environment is distributed or not, how much auxiliary devices can be introduced, how much quality is required, how much cooperation is expected from the participants, how much cost can be spent on the environment, etc.

ICIE (Immersive Collaborative Interaction Environment) [24], as shown in Fig. 6, is a generalization of an immersive WOZ environment described in Section 2. The ICIE can serve as a "bubble" that can move around in a larger virtual space artificially created or result from virtualization of the real world. The user can move around the virtual space to meet other users or autonomous agents that are designed to provide various services.

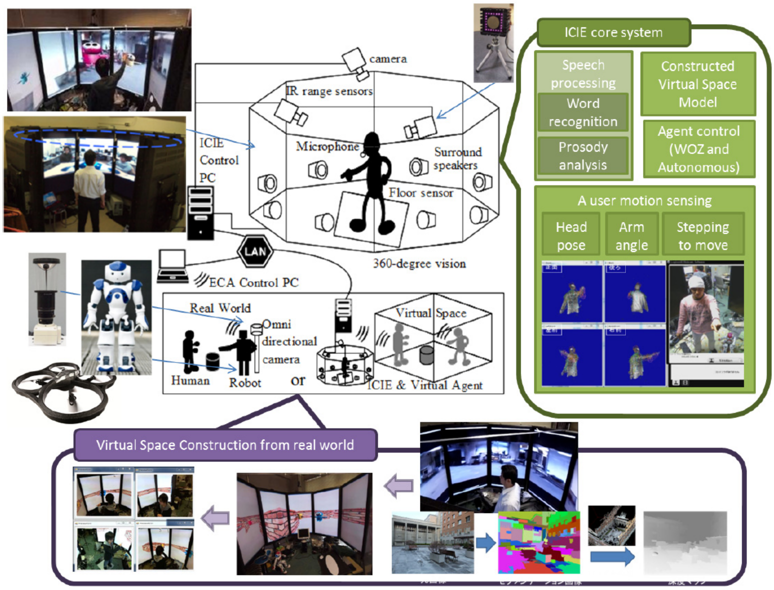


Fig. 6. ICIE (Immersive Collaborative Interaction Environment) [24]

The real world can be projected into the ICIE screen in either online or offline. In the online mode, scenes captured by one or more CCD camera or even omnidirectional camera possibly attached to a mobile vehicle can be projected on the screen on the fly. It will give a real-time immersive image that permits the user to feel as if the user in ICIE were moving around the projected space.

In the offline mode, interaction with the user is made in the virtual world that has been built beforehand. We have developed a method for reconstructing a virtual space consisting of panoramic images for a given area of the real world, by combining multiple computer-vision techniques such as structure from motion, multi view stereo, and depth map [25]. As the user virtually walks around the given space, a panoramic image for the location is computed almost in real time. The current version can automatically re-construct from approximately 5000 digital photos a 3D scene for a 50m x 50m space in 1 days (Fig. 7).

The ICIE allows for building an interactive collaborative system for a virtual space. We are currently building a system called simulated crowd [26], which is based on the idea of synthetic culture, i.e., “role profiles for enacting dimensions of national culture” [27]. Simulated crowd allows the user to virtually walk in a crowd of a given culture to experience culture-specific non-verbal signals by which people cooperate or negotiate with each other to avoid collision and achieve the respective goals [28] (Fig. 8). We have found that the different settings of values for generic parameters may influence not only the physical characteristics of the place (such as average travel time or wait time in the space) [29] but also cultural interpretations by the user [30].

The future challenge involves a full scale integration of conceptual information space with our physical/virtual living space that will allow for creating and leveraging common ground for conversation.

Fig. 7. The user walks around a virtual space corresponding to a place of 50*m* x 50*m* in our university campus [25]

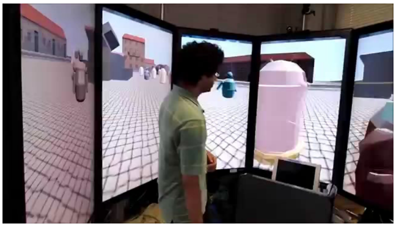


Fig. 8. The user interacting with a simulated crowd

5 Conversation Measurement, Analysis and Modeling

The goal of conversation measurement, analysis and modeling is to uncover principles of verbal and nonverbal interactions that people engage everyday as a part of intellectual activities. The insights and data obtained from conversational analysis permit engineers to incorporate the insights to various applications.

In order to study conversations by measurement and corpus building, we developed an environment called IMADE (the real world Interaction Measurement, Analysis and Design Environment. In addition to multi-modal sensing devices, such as the wearable motion capture devices or eye mark recorders, we introduced biological and brain measurement devices so that we can observe the internal activities and their interdependencies of each participant in a given conversational situation (Fig. 9).

Masaharu Yano has prototyped a portable 3D multi-party conversation recording environment that integrates static recording of environment and 3D recording using multiple range sensors to re-construct a 3D movie for multi-party conversation [31]. Fig. 10 illustrates a snapshot of tracking a first-person view moving with a given actor (Fig. 10), allowing the user to understand how other people may observe the world from a given perspective.

We are developing a method for extracting sequential pattern of nonverbal behaviors in the context of multiparty conversation for extracting machine-readable interaction protocols that we unconsciously use in our daily conversations. We have built interaction corpora containing multi-viewpoint videos, voice sound, motion data, and eye tracking data of each participant to the conversation as well as interaction primitives articulated from the above data, i.e., turn taking of speech, back-channel feedback, head nodding, viewpoint movement, pointing gesture, etc. In our interaction mining method, the sequential patterns of these nonverbal behaviors are represented in *N*-gram. In the *N*-gram representation, we could find characteristic patterns depending on individual conversational situations such as turn-taking mechanisms among the participants of poster presentation.

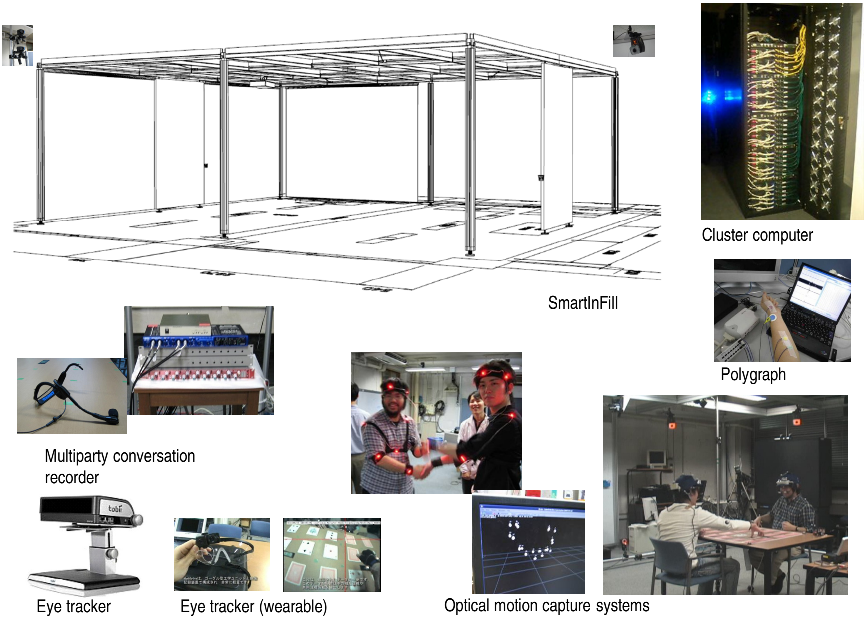


Fig. 9. Devices and sensors to sense the interaction behavior

The future challenge involves a high-resolution, full-scale integrated 3D conversation capture that will allow the user to investigate conversations from any viewpoint together with brain physiological measurement.

6 Toward Empathic Agents

Empathy has been defined as “the ability to understand others’ emotions and/or perspectives and, often, to resonate with others’ emotional states,” or as “an affective response that is identical, or very similar, to what the other person is feeling or might be expected to feel given the context: a response stemming from an understanding of another’s emotional state or condition” [32]. Empathy can also be considered to be equivalent to conviviality that allows individuals to identify with each other thereby experiencing each other’s feelings, thoughts and attitudes, and hence is deemed a central concept to design a community [33]. According to [34], mirror neurons help us understand the mental state of other people by making some form of inner imitation to pretend to be “in other people’s shoes” (the mirror neuron hypothesis of empathy).

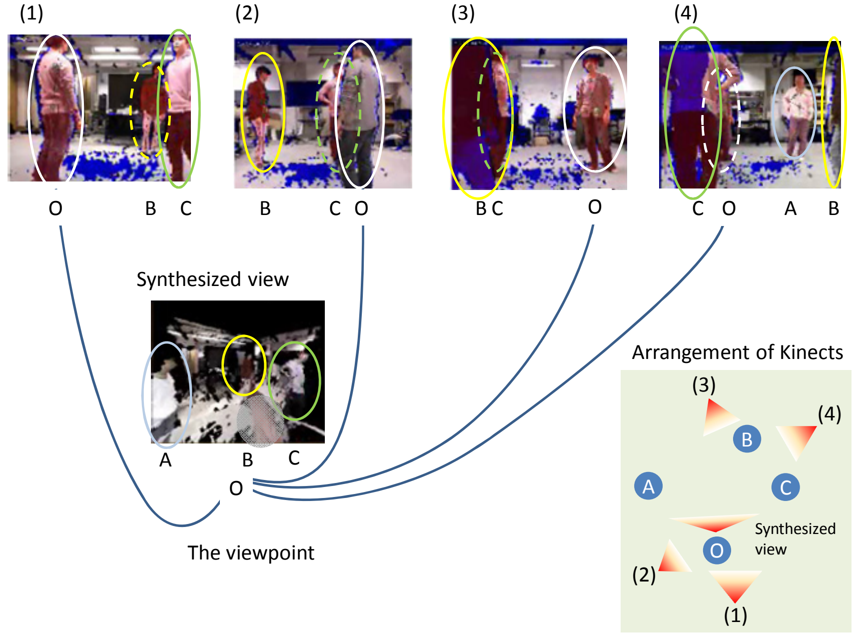


Fig. 10. A portable 3D multi-party conversation recording environment [32]. In this example, four Kinects are used to re-produce the conversation of four participants {"O", "A", "B", "C"}. (*i*) shows a scene from the *i*th Kinect, arranged as shown in the lower-right corner. The synthesized view shows an estimated scene captured by a virtual camera placed in front of the participant "O", an approximation of what "O" sees.

Building empathic agents, which can create and maintain empathy with people to efficiently and securely understand our intentions to provide with maximal service, is an ultimate goal of an engineering approach to conversational informatics. The computational model for empathy might be the one shown in Fig. 11. The architecture of an individual person's cognition may include an "internal theater." In addition to a routine work of human information processing that takes information from receptors and produces motor commands, the input is sent to the internal theater where what happening is reproduced for reflection or planning. The communication partner's behavior is reproduced with the assistance of mirror units in the internal theater and interpreted by using the actor's own mechanism for generating emotional appraisals.

Building an empathic agent may be counted as one of the most challenging problems in AI, for it involves reproduction of a substantial portion of a mental process of humans. The key issues appear to consist in simulation and metaphor. We may use simulations to reproduce the mental process at some levels of representation abstracted from the neuro-physiological levels. On the other hand, we need to depend

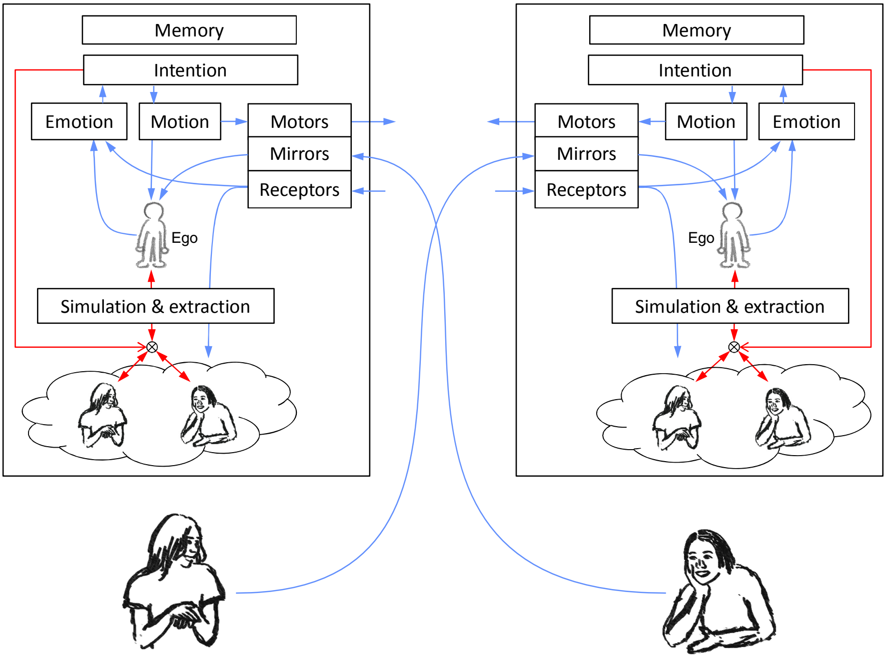


Fig. 11. Hypothesized mechanism of empathy [35]

on metaphor to project the embodiment and emotion of the human to the empathic agent and vice versa. Although the empathy between human and agent might be weaker than the empathy among people, the sharing hypothesis suggests that it would be much better than nothing.

It appears that empathy has different degrees of intensity, depending on how much is shared and how much the participants are aware of it. According to the *sharing* hypothesis, the more is shared, the more empathy is gained [35]. ICIE allows people to share the first-person view. It helps people look at the world from other angles to understand how other people might conceive the shared world. According to the sharing hypothesis, increase in shared world perception will induce increased empathy.

References

1. Clark, H.: Using Language. Cambridge University Press (1996)
2. Nishida, T. (ed.): Conversational Informatics: an Engineering Approach. John Wiley & Sons Ltd., London (2007)
3. Huang, H.H., Furukawa, T., Ohashi, H., Nishida, T., Cerekovic, A., Pandzic, I.S., Nakano, Y.I.: How multiple concurrent users react to a quiz agent attentive to the dynamics of their game participation. In: Proc. AAMAS 2010, pp. 1281–1288 (2010)

4. Kiselev, A., Hacker, B.A., Wankerl, T., Abdikeev, N., Nishida, T.: Toward incorporating emotions with rationality into a communicative virtual agent. AI Soc. 26(3), 275–289 (2011)
5. Cerekovic, A., Huang, H.-H., Pandzic, I.S., Nakano, Y., Nishida, T.: Towards a Multicultural ECA Tour Guide System. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 364–365. Springer, Heidelberg (2007)
6. Nishida, T., Terada, K., Tajima, T., Hatakeyama, M., Ogasawara, Y., Sumi, Y., Xu, Y., Mohammad, Y.F.O., Tarasenko, K., Ohya, T., Hiramatsu, T.: Toward robots as embodied knowledge media. IEICE Transactions 89-D(6), 1768–1780 (2006)
7. Weizenbaum, J.: ELIZA — A computer program for the study of natural language communication between man and machine. CACM 9(1), 36–45 (1966)
8. Winograd, T.: Understanding Natural Language. Academic Press (1972)
9. Erman, L.D., Hayes-Roth, F., Lesser, V.R., Reddy, R.: The Hearsay-II speech-understanding system: integrating knowledge to resolve uncertainty. ACM Computing Surveys 12(2), 213–253 (1980)
10. Bolt, R.A.: Put-that-there: Voice and gesture at the graphics interface. In: Proc. the 7th Annual Conference on Computer Graphics and Interactive Techniques, vol. 14(3), pp. 262–270 (1980)
11. Byrne, J.A., Sculley, J.: Odyssey: Pepsi to Apple: A Journey of Adventure, Ideas, and the Future. Harpercollins (1987)
12. Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.): Embodied Conversational Agents. The MIT Press (2000)
13. Prendinger, H., Ishizuka, M. (eds.): Life-like Characters—tools, affective functions and applications. Springer (2004)
14. Rehm, M., Nakano, Y., André, E., Nishida, T., Bee, N., Endrass, B., Huang, H.H., Lipi, A.A., Wissner, M.: From observation to simulation: generating culture-specific behavior for interactive systems. AI & Society 24(3), 267–280 (2009)
15. Nishida, T.: Talking with Uncertainty. In: Tang, Y., Huynh, V.-N., Lawry, J. (eds.) IUKM 2011. LNCS, vol. 7027, pp. 1–6. Springer, Heidelberg (2011)
16. Ohashi, H., Okada, S., Ohmoto, Y., Nishida, T.: A proposal of a novel WOZ environment for realizing essence of communication in social robots. In: 9th International Workshop on Social Intelligence Design (SID 2010), London, United Kingdom, September 13-16 (2010)
17. Saiga, H.: Construction of A Semi-automatic Telepresence System Correcting Motion by Integrating A Controller's Behavior with A Remote Sensing Data. Master thesis, Department of Intelligence Science and Technology, Kyoto University (2012) (in Japanese)
18. Mohammad, Y.F.O., Nishida, T., Okada, S.: Unsupervised simultaneous learning of gestures, actions and their associations for Human-Robot Interaction. In: Proc. IROS 2009, pp. 2537–2544 (2009)
19. Mohammad, Y.F.O., Nishida, T.: Learning interaction protocols using Augmented Baysian Networks applied to guided navigation. In: Proc. IROS 2010, pp. 4119–4126 (2010)
20. Nishida, T., Fujihara, N., Azechi, S., Sumi, K., Yano, H.: Public Opinion Channel for communities in the information age. New Generation Computing 14(4), 417–427 (1999)
21. Nishida, T.: Conversation Quantization for Conversational Knowledge Process. In: Bhalla, S. (ed.) DNIS 2005. LNCS, vol. 3433, pp. 15–33. Springer, Heidelberg (2005)
22. Okamura, G., Kubota, H., Sumi, Y., Nishida, T., Tsukahara, H., Iwasaki, H.: Quantization and reuse of driving conversations. Transactions of Information Processing Society of Japan 48(12), 3893–3906 (2007) (in Japanese)

23. Merckel, L., Nishida, T.: Enabling situated knowledge management for complex instruments by real-time reconstruction of surface coordinate system on a mobile device. AI & Society 24(1), 85–95 (2009)
24. Ohmoto, Y., Ohashi, H., Lala, D., Mori, S., Sakamoto, K., Kinoshita, K., Nishida, T.: ICIE: Immersive environment for social interaction based on socio-spatial information, Conference. In: Proc. Conference on Technologies and Applications of Artificial Intelligence (TAAI 2011), Taoyuan, Taiwan, pp. 119–125 (2011)
25. Mori, S., Ohmoto, Y., Nishida, T.: Constructing immersive virtual space for HAI with photos. In: Proc. GrC 2011, pp. 479–484 (2011)
26. Thovuttikul, S., Lala, D., Ohashi, H., Okada, S., Ohmoto, Y., Nishida, T.: Simulated Crowd: Towards a synthetic culture for engaging a learner in culture-dependent nonverbal interaction. In: 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction at IUI 2011, California, USA (2011)
27. Hofstede, G.J., Pedersen, P., Hofstede, G.: Exploring Culture, Exercise, Stories and Synthetic Cultures. Intercultural Press, Yarmouth (2002)
28. Lala, D., Nishida, T.: VISIE: A spatially immersive interaction environment using real-time human measurement. In: Proc. GrC 2011, pp. 363–368 (2011)
29. Lala, D., Nishida, T.: Towards a virtual environment for capturing behavior in cultural crowds. In: Proc. ICDIM 2011, pp. 310–315 (2011)
30. Lala, D., Nishida, T.: The interpretation of emotion based on gestural parameters and culture. In: CMVC 2011, Reykjavik, Iceland (2011)
31. Yano, M.: Construction of 3-dimensional Recording Environments for Multi-party Conversation with RGB-depth Sensors. Master thesis, Department of Intelligence Science and Technology, Kyoto University (2012) (in Japanese)
32. Eisenberg, N., Eggum, N.D., Di Giunta, L.: Empathy-related responding: associations with prosocial behavior, Aggression, and Intergroup Relations. Social Issues and Policy Review 4(1), 143–180 (2010)
33. Caire, P.: Designing convivial digital cities: A social intelligence design approach. AI Soc. 24(1), 97–114 (2009)
34. Iacoboni, M.: Mirroring People – The New Science of How We Connect with Others. Farrar, Straus & Giroux, New York (2008)
35. Nishida, T.: Toward mutual dependency between empathy and technology. AI Soc. 25th Anniversary volume – a Faustian Exchange: what Is to be Human in the Era of Ubiquitous Technology? (Online First, 2012), doi: 10.1007/s00146-012-0403-5

Technology and Social Interaction in the Multimodal, Multispace Setting of Audiometric Testing

Maria Egbert

University of Southern Denmark, Alsion 2, 6400 Sonderborg, Denmark
meg@sitkom.sdu.dk

Abstract. A frequent motivation for integrating the technological and the social sciences lies in understanding the users of technologies in order to innovate [1-3]. This paper argues that a shift is necessary from ‘the user’ to ‘the interaction in the participation framework’ because it is here where the interactants display to each other their relevancies. Using Conversation Analysis [5-8], this point is exemplified by the examination of interaction in an audiological consultation where the interface of sociality and technology is relevant as a barrier. The analysis focuses on what aspects the participants in their talk and nonverbal conduct orient to as problematic given the task and the technology in this multimodal, multispace environment. The analytical results are discussed for innovation within the framework of User-Centered Design.

Keywords: Innovation in technology and interaction, User-Centered Design, Conversation Analysis, communication disability, audiological interaction, communication with hearing loss, audiometric testing.

1 Introduction

Assistive and assessment technologies for hearing loss are a prime example for the interconnectedness of technology and sociality because the tremendous technological advancement has not resulted in a wide acceptance of hearing aids, although for most persons, it is the only help to improve their hearing ability. Adult-onset hearing loss is estimated as “the second largest cause of Years Lost to Disease” [9], affecting adults at all age ranges, e.g. age 18-44 = 23% prevalence, age 45-64 = 29%, age 65 = 30% [10], with rising rates at all ages [11]. In most countries with a health care system accessible to all citizens, an average of 80% of persons with hearing loss does not use hearing aids [12], with a compliance rate ranging from less than 15% (e.g. USA, Europe, Japan) to around 40% (e.g. Australia, Denmark, UK). Limited access to health care correlates to an even lower usage of less than 5% (e.g. China, India, Lithuania). Hearing loss leads to stress in communication, withdrawal from interaction, increased sick leave, early retirement, and social isolation [13, 9].

Most of the barriers to hearing aid use are rooted in phenomena learned and experienced in social interaction. These include shame and stigmatization due to associations with old age and dumbness, denial of having hearing loss, problems in the interaction with hearing health specialists, unrealistic expectations that the hearing aid will bring back normal hearing, and technical problems during the use [14].

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 240–252, 2012.

Hearing health encounters are a barrier due to problems in information transfer and sociality [14, 15]. In a survey of 190 patients, 88% dropped out during the path through the health care system [17].

These problems have been identified in studies based on interviews, questionnaires, surveys and focus groups, yet in order to fully understand and address these barriers, research needs to investigate how interactants co-create these relevancies and handle them in the situations where they actually occur, namely in naturally-occurring encounters. Using Conversation Analysis [5-8] to study a video-taped hearing aid consultation, this study yields that in this encounter alone, the participants bring up all but one of the barriers listed above, and additional problems emerge.

A key result is that the client, her husband and the audiologist pursue different agendas. Whereas during history taking, the client and her husband bring up problems in the area of psychology, sociality and technology, the audiologist focuses exclusively on technological aspects. During audiometric testing, a tension between institutional procedure and interactional concerns is handled by departures from the testing procedure. In addition, the audiologist's and the client's descriptions of hearing differ.

The research is conducted within the international network “Hearing Aids Communication” [19], in which experts in Conversation Analysis and User-Centered Design collaborate with audiologists and hearing health professionals to better understand the social aspects of the problems with hearing loss and hearing aids. This novel approach aims at overcoming the current fragmentation of the research fields of social interaction and the medical/technological disciplines with the final goal of improving social participation with hearing loss.

2 Divergent Relevancies of Barriers to Hearing Aid Use

The data were collected in a German research center which offers consultations independently from commercial hearing aid dispensers. The video-taped encounter (1:18 hour) consists of three phases: history taking (36:25 min.), audiometric testing (14:17 min.) and the ensuing recommendation (27:15 min.). In this audiological consultation, the client’s husband and the audiologist pursue the goal of finding a hearing aid for her, while she goes along only with resistance and reservations. The three participants also differ in what barriers to the client’s use of hearing aids are relevant. The audiologist addresses only those barriers relevant to technology, i.e. the selection and fitting of the most appropriate hearing aid. In contrast, when the client or her husband mention psychological and social barriers, the audiologist does not address these issues and returns to his agenda.

Of prime importance to understanding this encounter is the conversation analytical concept of ‘preference’ in sequences of interaction [20, 7]. For example, question/answer sequences are a frequent structure underlying medical encounters [21]. In response to a question, the range of answers underlies the social rule for agreement, which is the ‘preferred’ way, whereas disagreements are ‘dispreferred’. Preferred responses are usually short and delivered promptly, while in contrast, dispreferred actions are delayed, often mitigated and more indirect, thus resulting in a longer turn where the disagreeing elements are pushed back towards the end of the turn. There are also dispreferred ‘first’ actions, such as signaling trouble in hearing and understanding ('other-initiated repair' [22]), complaints, self-deprecations and requests [7]. The preference for agreement is related to the preference for contiguity in that agreeing actions are placed contiguously and disagreeing actions are delayed. Thus, the concept of 'preference' describes a social structure and is not to be confused with a person's likings.

2.1 Hearing Loss as a Delicate Issue

Both the husband and the client treat talking about hearing loss as a sensitive matter, thus orienting to the stigma and taboo associated with hearing loss [23-25]. At the beginning of history taking, the audiologist asks the client about her problems and how he can help. She starts a response, but then self-interrupts and turns the floor to her husband. The husband then names as reason for the visit that according to his impression, her hearing has become worse. His formulations contain smile voice, laugh tokens, hesitations, and mitigations. This is followed by a non-serious attempt to align with her hearing loss by considering the possibility of him having 'marriage deafness', alluding to the stereotype that husbands do not listen to their wives. In using this term, he comes as close as possible to aligning with her hearing loss although he has normal hearing. The combination of these dispreferred features are displayed in segment #1 below.

#1: Hearing loss as a sensitive matter (5:57)

Hus: ja. also ich habe den eindruck als wenn das (.) ä: .h (0.3) sich verändert hat=
yes. well I have the impression as if it (.) e:h .h (0.3) has changed=

=als wenn (sie jetz) das hörvermögen schlechter geworden is.
=as if (she now) the hearing ability has become worse.

Aud: °ja:a,°
°ye:s,°

smile voice
Hus: .hhh [aber (0.3) *das kann ja auch he e (0.2) he e (.) ne ehe, taubheit sein.=
[*but (0.3)) it can also be he e (0.2) he e (.) a marriage, deafness.=*
[
Aud: [°okee°
[*°okay°*

Hus: =hätt ich fast gesacht, .h he he he .hhh he he
=I almost would have said, .h hehehe .hhh he he

The client herself also orients to this stigma and taboo. In response to the audiologist's question about her hearing ability in groups of people, the husband delays his answer with hesitations and a long pause. The client ends his silence by explicitly permitting him so 'say it', thus implying that she takes his hesitations as an orientation to this topic being a taboo matter.

#2: Hearing loss as a sensitive issue (6:50)

Hus: also ä:::m: (1.5)
well eh:::m: (1.5)

Cli: du darfst das ruhig sagen.
you may say it really

Hus: nee. [nein=nein. also [nein=nein. im im [gegenteil.
no. [no=no. well [no=no. on the on the [contrary
[[[
Cli: [he wei(h)ßt du [do(h)och [.hh he he
[bu(h)t you kno(h)w [that [.hh he he

In addition to these dispreferred elements, the underlying structure of this sequence is dispreferred in that it consists of a misunderstanding ['third position repair', 26, 27], where the husband rejects his wife's interpretation of his hesitation. In turn, the wife disagrees with his rejection of her interpretation.

2.2 Denial

A strong barrier to hearing aid use is denial of the disability or mitigation of its severity. This is one reason why it is frequently a close relative with normal hearing who notices that his or her family member is affected by hearing loss [28]. In the case at hand, the client and her husband disagree about the severity of her hearing loss. While the husband provides as one reason for the consultation that her hearing has become worse (excerpt #1 above), the client insists that there has been no change: 'das hatn sich nach meiner meinung nicht geändert.' ('*that has according to my opinion not changed.*') (8:30). She repeats this stance: '.hhh aber so lange ich::: (0.3) denken kann, .h hat sichs nich verändert' ('*.hhh but as far back as I::: (0.3) can think, .h it hasn't changed*') (8:43). By placing contrastive stress on the references to herself, she emphasizes the difference between her own assessment and that of her husband.

2.3 Client's Resistance to Hearing Aid Use

The client shows reluctance and reservations to acquiring hearing aids throughout the encounter, and she provides a number of accounts for her non-compliance. Her negative stance is observable already in the pre-beginning, when she and her husband sit down at the desk for consultation while the audiologist leaves the room for a moment. In reaction to the husband admiring the computer equipment, the client withholds a response to his positive assessment, then initiates repair and launches a negative assessment about the arrangement of the computer cables: 'und auch so'n kabelsalat' ('*and also such a mess with the cables*' (00:30). By placing emphasis on 'auch' ('also'), she implies that her negative assessment covers other aspects as well, thus displaying a more general negative stance. All these features are dispreferred. A little later during history taking, the client gives an unmitigated dispreferred response to the audiologist's question about what hearing aids she is currently using. When he poses a confirmation question, she shakes her head, followed by a nonsensical account. Again, her action is dense with dispreferred features.

#6: Unmitigated dispreferred response by client

Aud: welche hörgeräte tragen sie derzeit.=
what hearing aids are you wearing at present.=

Cli: =ga:r keine.
=none at a:ll.

(0.5)

Aud: ham sie gar keine geräte?
do you have no aids at all?

(0.4) / ((Client shakes head))

*smile voice
Cli: ich- die liegen im schrank seitdem *sie nich mehr da sind.
*I- they are lying in the closet since *they are no longer there*

In her account why she owns but does not use hearing aids, she refrains from any agency, and the hearing aids in the drawer seem to exist no longer. The interaction continues in that she leads into a report of the last failed attempt: 'ich weiß im moment nich mehr wie die: dinger hießen. die hab ich wieder zurück absolut katastrophe. ('*I don't know at the moment any more what the:se things were called. I returned them absolute catastrophe.*') (01:35). An 'extreme formulation' [29] like 'absolute catastrophe' can serve the purpose of normalization.

2.4 Audiologist's Exclusive Focus on Technology

Throughout the encounter, the audiologist concentrates on the barriers which lie in his area of expertise in that he pursues the institutional goal of finding the best hearing aid, while disregarding non-technical reasons for non-compliance, as the following data excerpt exemplifies. In describing and accounting for the client's resistance to the consultation and the use of hearing aids, her husband brings up psychological reasons.

#7: Audiologist does not take up psychological issue (6:10) (simplified)

Hus: du musst dich damit (h) ausenandersetzen
you must deal with (h) it

mit der ganzen geschichte dass es mal n hörgerät gibt.
with the whole story that there will be a hearing aid some time.

.hh das is ja auch ne psychologische angelegenheit
.hh that is also a psychological matter

Aud: also allgemein sprachverstehen so vom gefühl her, (0.5) schlechter geworden.
so in general speech understanding it feels, (0.5) has become worse.

The audiologist is selective in what elements of the husband's description he zeros in. The husband's subjective theory of her compliance ('also a psychological matter') receives no uptake. Rather, the audiologist returns to the issue of language understanding, a piece of information relevant to the choice of hearing aids. The same strategy is observable when the client herself mentions a psychological reason for the non-use of hearing aids.

2.5 Negotiating Unrealistic Expectations

Research on hearing aid barriers reports that many persons with hearing loss expect too much from the technology [22, 23, 44]. Unrealistically high expectations, e.g. that hearing aids brings back normal hearing, are probably due to lack of knowledge and to the hearing aid industry's advertisements [45]. Clients' expectations have been researched, but in the data analyzed, the audiologist's expectations also play a role in that he lowers his initial optimism of finding the right hearing aid for her:

#9: Audiologist's initial expectation of success (excerpt simplified)

Aud: tausend siebenhundert verschiedene hörgeräte, (0.2)
one thousand seven hundred different hearing aids, (0.2)

.h dass man da eins findet was passend is, glaub ich schon.
.h that one finds one which fits is, I do believe that.

.hh nu:r, (0.2) man muss natürlich verschiedene hersteller testen.
.hh only, (0.2) one has to test different producers of course.

Ensuing this positive forecast, the client's expectations are as follows: She rejects in-the-ear hearing aids because they give a sense of seclusion, she hears her own chewing sounds when eating, and the instruments feel too big due to her narrow ear canal. For in-door sports, where the client serves as a tennis instructor, she wants to be able to understand people who address her from a distance. This is difficult due to the echo effect in large rooms and the multi-person setting. Hearing aids need to be water repellent or water proof because during sports she sweats much and she likes to swim. When she goes on bike tours with her husband and friends, she needs to be able to hear their talk even when there are sounds of the wind and she is wearing a helmet. She has discontinued going to public lectures because she is not able to follow when the speaker doesn't use a microphone. In small theaters, when actors gaze away from the audience, she feels like being in a 'silent movie'. In normal everyday conversation she has trouble when others mumble or speak softly. She does not want hearing aids that make her ears stick out. Hearing aids should also not make her ears produce popping sounds. Watching TV with her husband is problematic because she needs to have the volume turned up so high that it disturbs him. Several times a year, the client has sinus infections which make using hearing aids uncomfortable.

The audiologist addresses all of these complex needs. For her degree of hearing loss, in-the-ear hearing aids are appropriate, and models are available where the portion inserted into the ear cannel only consists of a narrow pipe. He describes a preliminary package of technical choices and explains the purpose of the features (transcript not displayed here: language recognition and amplification, filtering out of disturbing non-language sounds like noise and the sounds of the wind, directional microphone, feedback suppression, suppression of echo effects, and automatic telephone recognition. For watching TV, a range of special head-sets are available. His explanations also cover price options.

As a further barrier, the client places the condition that the process of acquiring a hearing aid should not again be a 'rigmarole'. In German, she is using the derogatory term 'affentheater' (literally 'theater of monkeys'): 'ich hab k<u>ei</u>ne lust noch mal so'n so'n (.) affentheater mitzumachen dass es ein: (.)·hhh ä: von vorn herein zum scheitern verurteilt is.' (' *I have no desire to once more be part of such a such a (.) rigmarole that it is a (.) from the beginning it is sentenced to failure*') (13:50).

The audiologist takes up this condition by providing clear information about the process, the options within this process and criteria for pricing. He explains that the next step in selecting a hearing aid would be to test products by different companies in order to find out which sound quality she is most satisfied with. Once the brand is identified, the price can be balanced with the types of technological features to be selected. He summarizes that the client will have to expect to test at least two to three instruments, and that for each instrument two to three consultations are necessary, totally six to nine appointments.

Towards the end of history taking, there is a noteworthy downward shift is the audiologist's prior forecast of finding the right hearing aid: 'eh:m es gi:bt möglichkeiten=aber es gibt auch grenzen.' ('*uh:m there a:re possibilities=but there are also limitations*') (34:40).

In the ensuing stretch of interaction, the client brings up more conditions and questions, to which the audiologist provides clear information. Whereas the husband receives these with agreements (verbal, continuers, head nods), the client stays reserved (minimal responses, other-initiated repair) and places more conditions. She never agrees, and when her verbal resistances discontinue, the audiologist moves forward to the next point in the agenda, the pure tone audiogram.

3 Problems during Audiometric Testing

Pure tone audiometric testing is a standardized procedure to objectively assesss the client's degree of hearing loss based on the perception of 'pure' or 'sinus' tones. It is usually conducted in a sound proof room. By means of a special computer program the audiologist feeds a series of tones, each in continuously increasing volume. The client is instructed to signal when she first hears a tone fed to her through head-sets (phase 1) or the tone's loudness becomes uncomfortable (phase 2). In phase 3, the tones are played through an instrument the client holds behind her ear. In the case at hand, all three phases are completed within 11:45 minutes.

The audiometric test represents a 'situation', which, according to Goffman, is "an environment of mutual monitoring possibilities, anywhere within which an individual will find himself accessible to the naked senses of all others who are 'present', and similarly find them accessible to him" [32]. In this specific situation, mutual monitoring is complex due to the spatial set-up, the institutional goal, and possibly the client's hearing loss. It consists of two interconnected dimensions which can be described in terms of interaction and space. During sequences of direct interaction between the audiologist and the client, they gaze at each other. The participants use this constellation for instructions and for dealing with trouble. For the actual testing, the client is wearing head-sets (phases 1 and 2) and is thus not acoustically available to the audiologist. By turning her head downwards and sideways to him and by closing her eyes she is not visually available to him (Fig. 1).

Fig. 1. Focus on testing

Fig. 2. Focus on interaction

The desk, computer and chairs are arranged for testing, to which the participants orient by positioning their torso and lower bodies rectangular to each other. Through this ‘position’ [33] they mark the larger unit of interaction. When they temporarily side-track to the smaller unit ‘point’, they do so through head shifts only. In line with Scheflen’s findings [47], these physical orientations indicate that the main activity is testing hearing to which direct interactional sequences are subordinate.

3.1 Departures from Testing Procedure Induced by Audiologist

During pure tone testing, five departures occur, three induced by the audiologist and two by the client. In all cases they concern test instructions. The analysis briefly summarizes the first three departures and then examines the client-induced shifts in more detail. For an analysis of all five departures, cf. [18].

The first three shifts from testing to talk are induced by the audiologist. 53 seconds into the first testing phase, he solicits the client’s attention to let her know that the sound feeding is about to change from the left to the right ear. To mobilize her attention, he employs nonverbal and verbal resources. First he lifts his right hand. When she turns her gaze to him, he points to her right ear and swiftly draws a semi-circle until he is pointing to her left ear. During this movement he utters ‘now this side’. The client nods after his first word and then shifts her gaze back to the downward testing position. While the audiologist’s actions serve to orient the client to an upcoming shift, this action achieves that she moves her focus away from testing. Whether she understands his talk through lip reading or surmises its meaning through his pointing gestures is neither accessible to the audiologist nor to the researcher.

In the second departure, the audiologist also tries to preempt a potential problem. First, the client holds the signaling instrument in her right hand and presses the button several times. When she changes hands and shakes out her right hand, the audiologist mobilizes her attention by lifting up his head and eye brows. After she turns her gaze to him, he verbally offers that she can also nod with her head. In response, the client takes off the head-sets, thus indicating that she did not hear him but can now listen. The audiologist repeats his offer, which she rejects, then he reformulates the offer which she rejects again and accompanies with an account.

The third departure occurs in the second testing phase, where the client’s discomfort level of loudness is measured. She is instructed to signal this by uttering ‘stop’. Although the client displayed understanding during instruction giving, she does not use the verbal signal ‘stop’. Rather, she employs a facial expression of frowning and a head shake. This departure from the agreed upon signal is the source for the audiologist to initiate repair. He inserts a candidate understanding [22, 27] in the form of 'STOP?', delivered loudly and with question intonation. She confirms this with a head nod. The participants have thus collaborated in establishing a new signal.

In all three instances, the audiologist induces a momentary shift from testing to talking by departing from the procedure, thus handling problem solution immediately. The next two problematic sequences are induced by the client, who delays the timing to the next possible transition phase in the testing procedure.

3.2 Problems Signaled by the Client

The first problem brought up by the client consists in finding out exactly at what point she needs to signal that a tone fed through head-sets is too loud. When the audiologist explains this, she first signals understanding, but a turn later produces an understanding check.

#13: Negotiation the meaning of 'uncomfortably loud'

Aud: also, (0.5) <wenn es <u>un</u>angenehm laut wird> wenn sie sagen
so, (0.5) <when it gets too <u>un</u>comfortably loud> when you say

ohe, (0.2) <<u>viel</u> zu laut> sagen sie einfach nur s:topp.
oh no, (0.2) <<u>much</u> too lout> say simply just s:top.

*nods
Cli: m*h[m.
[
Aud: [auf jedem ohr mache ich vier töne.
[*on every ear I put four tones.*

Cli: solange wie ich das ertr<u>a</u>gen kann?
as long as I can end<u>u</u>re it?=

Aud: genau.
exactly.

(0.2)

*Cli nods
Aud: also jetz=jetz nich so <u>un</u>mäßig quä:ln aber wenns *wirklich- (0.5) wirklich
*well now=now not like to:rturing too much but when it *really- (0.5) really*

<u>zu</u> laut is *einfach nur stopp [sagen.
*is <u>too</u> loud *simply just say stop.*
[
Cli: [gut.
[*all right.*

Although she claims trouble resolution with 'gut' ('all right.'), she brings up this problem again. Directly upon the audiologist's pre-closing of the third testing phase, she takes off her head-sets and formulates a specification of when she pressed the button to signal uncomfortable loudness.

#14: Client double-checks understanding of ‘uncomfortable’ (9:55-10:01)

Aud: oke:e,?
oka:y,?

Cli: also man konnte es länger ertra:gen: so nich aber es war unangenehm
well one could endu:re it longer: but it was uncomfortable

Aud: okay j: das is in ordnung.
okay j: that is all right.

Again, her problem lies in the meaning of ‘uncomfortable’ and whether she could ‘endure’ it. This indicates continued insecurity with this part of the instructions. Her timing shows that she orients to the progression of the testing procedure as a priority over signaling her trouble.

The second problem brought up by the client occurs in the last testing phase, where the tones are no longer fed through head-sets but through an instrument the client holds behind her left ear and then her right ear. Prior to the sound feeds, the audiologist provides instructions and the client signals understanding. In the course of testing, she complies with the instructions, yet when the first transition phase emerges (the audiologist instructs her to change the instrument from the left to the right side), she uses the opportunity to ask for a clarification. The problem the client articulates pertains to the perception of tones and to the accuracy of signaling. To her, the detection of a tone is not a point but rather a period in time. During this period, the perception changes qualitatively. Reducing the phenomenology of hearing to pure tones is not in line with the client’s report of perception. In both client-induced departures, both participants minimize the disruption. She does so by positioning her specification as non-intrusively as possible, and he by his minimal reaction.

As these instances show, the testing procedure has no provision for signaling trouble. All departures from the test’s contiguity serve the purpose of dealing with instructions, and are thus, paradoxically, produced in service of following the protocol while simultaneously disturbing it. Given the test design’s lack of a provision to signal trouble once it is in progress, the participants resort to nonverbal actions in order to mobilize the other’s attention and to make them available for talk.

4 Relevance for Innovation

The micro-analysis has shown that all but one of the barriers reported in prior research are mentioned in this single case as relevant by the interactants. The exception is that this audiologist, as shown above, provides detailed information about the technical features of hearing aids in relation to the client’s needs and in relation to the selection and fitting process. A barrier not mentioned in prior research lies in the dispensing of hearing aids. In the consultation under investigation, the audiologist describes a procedure for selecting hearing aids used in the research center which is unique in Germany. While this hearing center can offer clients to try out the hearing aids from all European companies, the usual dispensers sell one or two brands. In business, it is unlikely that a hearing aid dispenser would send clients to competitors.

Clients would first have to be aware of this crucial information in order to then consult different dispensers and to test their hearing aids on a trial and error basis.

The meso level of a society's structure pertains to its organization of institutions and companies, such as, in the case of hearing loss, the facilities for hearing health care, the ways hearing aids are dispensed and the training of professionals. The micro-analysis reveals a domination of the institutional and technological aspects. If audiological consultations are not the place to address socio-psychological barriers, hearing health professionals should be able to recommend such care. However, this presupposes that such care is available. In the German health care system, treating the socio-psychological aspects of hearing loss is not an integral part. In contrast, the national Danish health care system provides hearing pedagogues and rehabilitation experts, which may be a reason why Denmark has one of highest compliance rates of hearing aid use worldwide.

A further problem at the meso level concerns the way hearing aids are dispensed. The industry-independent audiologist in the data of this study describes as the first step in selecting a hearing aid to find out which company's product fits best, and only in the second step, once the company has been identified, to narrow down the search within the product range of this company. In the German system, the first step provided here is not routine. Dispensers offer products of one or two companies and compete with other dispensers. For this reason, innovation at the meso level should take seriously the patients' needs for information and full range of choices. It would also preempt frustration if hearing aid companies refrained from raising unrealistic expectations.

At the macro level, governmental policy is a crucial point of consideration because at present, it fails to address most of the problems associated with hearing loss beyond medicine and technology. A systematic inclusion of socio-psychological aspects would eventually save costs because more hearing aids would be used, and secondary economical loss could be lessened. Untreated hearing disability creates economic loss to individuals, companies and national economies. The economical loss for each person dropping out of the workplace early due to hearing loss is estimated at 200,000 US Dollars to society [13]. A higher compliance rate would also support the hearing aid industry in raising the currently low market saturation of 20%. While new and improved hearing aids are developed constantly, innovation research is disregarding the potential of developing interactional approaches which address the social problems.

At the global level, the WHO and the United Nations promote a shift in conceptualizing disability [34, 35]. The United Nations 2006 draft on the "Convention on the Rights of Persons with Disabilities" replaces the concept of disability as a condition of an individual, who needs to be treated, by a holistic concept of disability as a participatory socio-cultural phenomenon, which a multi-cultural society needs to address by integrating all members as full participants. This shift widens the focus from an individual with hearing loss to participation in socio-cultural interaction.

An approach to innovation which embraces this goal is User-Centered Design. Initial projects exploring the integration of these dimensions have been undertaken in the network 'Hearing Aids Communication' [19] which focuses on hearing loss in interaction because it is the central place where negative and positive conditions rooted at the micro, meso, macro and global level are observable in how they are relevant to the interactants.

References

1. Suchman, L., Blomberg, J., Orr, J., Trigg, R.: Reconstructing technologies as social practice. Special Issue of the American Behavioral Scientist on Analysing Virtual Societies: New Directions in Methodology 43(3), 392–408 (1999)
2. Suchman, L.A., Trigg, R.: Understanding practice: Video as a medium for reflection and design. Design at Work: Cooperative Design of Computer Systems, 65–89 (1991)
3. Buur, J., Matthews, B.: Participatory innovation. International Journal of Innovation Management 12(3), 255–273 (2008)
4. Egbert, M.: Conversation analysis applied to user-centered design: a study of who 'the user' is. In: Antaki, C. (ed.) Applied Conversation Analysis, pp. 207–221. Palgrave Macmillan, Basingstoke (2011)
5. Sacks, H.: Lectures on conversation, vol. 2. Basil Blackwell, Oxford (1992); Edited by G. Jefferson with introductions by E.A. Schegloff
6. Pomerantz, A., Fehr, B.J.: Conversation analysis: an approach to the study of social action as sense making practices'. In: van Dijk, T.A. (ed.) Discourse as Social Interaction, pp. 64–91. Sage, London (1997)
7. Schegloff, E.A.: A primer for conversation analysis: Sequence organization. Cambridge University Press, Cambridge (2007)
8. Antaki, C. (ed.): Applied conversation analysis. Palgrave Macmillan (2011)
9. World Health Organization. Global health risks: Mortality and burden of disease attributable to selected major risks. WHO Press, Geneva (2009), `http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf`(accessed August 18, 2011)
10. Shield, B.: Evaluation of the social and economic costs of hearing impairment. A report for Hear-it (2006), `http://www.hear-it.org/multimedia/Hear_It_Report_October_2006.pdf` (accessed March 1, 2012)
11. Sorri, M., Jounio-Ervasti, K., Uimonen, S., Huttunen, K.: Will hearing health care be affordable in the next millennium? Scandinavian Audiology 30, 203–204 (2001)
12. Bisgaard, N.: Global hearing aid usage. Internal hearcom-Workshop (2009), `http://hearcom.eu/lenya/hearcom/authoring/about/DisseminationandExploitation/Workshop/5_Nikolai_Bisgaard_Industry-perspectives.pdf` (accessed August 18, 2011)
13. World Health Organisation. Deafness and hearing impairment survey. WHO Project ICP DPR 001, New Delhi (2001), `http://www.searo.who.int/LinkFiles/Publications_SEA-DEAF-4.pdf` (accessed August 18, 2011)
14. Kochkin, S.: MarkeTrak III: Why 20 million in US don't use hearing aids for their hearing loss. Hearing Journal 46(1) 20–27; 46(2), 26–31; 46(4), 36–37 (1993)
15. ProMatura. Exploring the consumer's journey: A report of an in-depth survey with hearing aid users to learn what impacts their sense of delight with their hearing aid (Hearing Industries Association) (February 2007a)
16. ProMatura. Report of four focus groups with hearing aid users to define the best practices for dispensing hearing instruments that will improve the success rate (August 21, 2007b)
17. Meis, M., Gabriel, B.: Barriers in the supply with hearing systems: The view of the customer. In: Proceedings of the 51st International Congress of Hearing Aid Acousticians, Frankfurt/Main, October 10-20 (2006)

18. Egbert, M.: Interaktion bei der Tonaudiometrie: Zur Paradoxie der Relevanzmarkierung von Störungen [Interaction during tone audiometry: About the paradoxy of relevancies in troubles]. In: Ayaß, R., Meyer, C. (eds.) Sozialität in Slow Motion: Theoretische und Empirische Perspektiven [Sociality in Slow Motion: Theoretical and Empirical Perspectives], pp. 463–479. Springer VS, Wiesbaden (2012)
19. Egbert, M., Deppermann, A. (eds.): Hearing aids communication. Integrating social interaction, audiology and user centered design to improve communication with hearing loss and hearing technology use. Verlag für Gesprächsforschung, Mannheim (2012)
20. Sacks, H.: On the preferences for agreement and contiguity in sequences in conversation. In: Button, G., Lee, J.R. (eds.) Talk and Social Organisation, pp. 54–69. Multilingual Matters, Clevedon (1973/1987)
21. Heritage, J., Clayman, S.: Talk in action. Interactions, identities, and institutions. Wiley (2010)
22. Schegloff, E.A., Jefferson, G., Sacks, H.: The preference for self-correction in the organisation of repair in conversation. Language 53, 361–382 (1977)
23. Hetu, R.: The stigma attached to hearing impairment. Scandinavian Audiology 25(43), 12–24 (1996)
24. McKenna, L.: Psychological aspects of acquired hearing loss. In: Graham, J., Martin, M. (eds.) Ballantyne's Deafness, pp. 258–271. Whurr, Philadelphia (2001)

Designing Smooth Communication Environment for Telemedicine for General Practitioner

Tomohiro Kuroda[1], Kenta Hori[2], Masahiko Ando[3], Takashi Kawamura[3],
Tadamasa Takemura[1], and Hiroyuki Yoshihara[1]

[1] Department of Medical Informatics, Kyoto University Hospital
Shogo-in Kawahara-cho 54, Sakyo-ku, Kyoto 606-8507, Japan
{tomo,takemura,lob}@kuhp.kyoto-u.ac.jp
[2] School of Radiological Technology, Gunma Prefectural College of Health Sciences
323-1 Kamioki-machi, Maebashi, Gunma 371-0052, Japan
hori@gchs.ac.jp
[3] Kyoto University Health Service
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan
{mando,kawax}@kuhp.kyoto-u.ac.jp

Abstract. Today's rapid advancement of the information and communication technologies enables us to perform telemedicine over the information networks. However, the realtime telemedicine is still not widely performed as a routine clinical activity. The authors developed a prototype multimedia communication system to support the realtime telemedicine and evaluated communication and technical barriers through protocol analysis. The results tell that ill synchronization and lack of nonverbal information may harm diagnostic process although the prototype enables general practitioners to diagnose a new patient from a remote site.

Keywords: Telemedicine, General practitioner, Realtime multimedia communication, Protocol analysis.

1 Introduction

Today's rapid advancement of the information and communication technologies enables us to perform diagnostic process, named telemedicine, over the information networks among distant sites[2]. From technical point of view, telemedicine can be classified into two groups; the store-and-forward type telemedicine and the realtime telemedicine.

The store-and-forward type telemedicine is the approach to store obtained data once into certain storage and share the stored data among distant sites. This type of telemedicine, including telepathology[12] and teleradiology[1], is already technically and legally available and widely performed. Electronic health record (EHR)[3], which is to share electronic patient record (EPR) over the information networks among multiple clinical organizations, is also regarded as a variation of the store-and-forward type telemedicine.

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 253–263, 2012.

The realtime telemedicine is the approach to transmit obtained data in realtime to enable a medical doctor to diagnose or to treat a patient over the information networks. Although innumerous trials are successfully performed, including the very early trial of telepsychiatry by Massachusetts General Hospital (MGH) back in 1968[6] and the transatlantic tele-robotic-surgery called "Operation Lindbergh" [4], the realtime telemedicine is still not widely performed as a routine clinical activity. In order to put the realtime telemedicine in practical use, we need to evaluate conventional realtime multimedia communication tools and to specify technical barriers of them under the most standard clinical process, which a general practitioner performs for a new patient.

The healthcare service section of schools and companies is one of the most common places to start clinical treatments. Recent mergers and relocation of schools and companies force the healthcare service section to provide their service, mainly pre-diagnosis for new patients, for remote sites.

The authors developed a realtime multimedia communication environment to realize telemedicine for newly opened campus of Kyoto University from its main campus. This paper discusses barriers and requirements of telemedicine under conventional communication tools through evaluation of the developed system.

2 Designing and Prototyping Telemedicine System

The common protocol to diagnose a new patient is to check the patient through interview, inspection, auscultation, and additional physiological / radiographic tests. Then, the doctor at the healthcare service division mainly provides consultation and guidance along the diagnosis, although the doctor sometimes introduces the patient in need for clinics and hospitals. Thus, the multimedia communication system to support telemedicine process of the healthcare service division needs to mediate conditions of the patient as well as messages from the doctors. Therefore, the communication support system should equips three channels, the standard TV-phone system for interview and inspection, another TV-phone system for detailed inspection and auscultation, and a shared memo pad for guidance. As a matter of course, existence of EPR to share physiological and radiographic test results helps diagnosis.

Fig.1,2 and Tab.1 shows configuration of the developed prototype[7]. The upper channel is a standard TV-phone system for interview and the lower channel is a shared memo pad. The middle channel dedicated to auscultation and detailed inspection equips an electric stethoscope, a standard inspection camera, a high definition still image camera, and a specially designed inspection camera for the oral cavity. The system utilizes DVTS [8] to transmit high-quality video and audio from the stethoscope with low latency, and FTP for the high definition still image. Fig.3 shows the state transition of the middle channel. The doctor switches modes of the middle channel along the progress of the diagnostic process.

The patient side terminal is installed to Katsura office of the Kyoto University Health Service where is ten kilometers away from the doctor side terminal at the

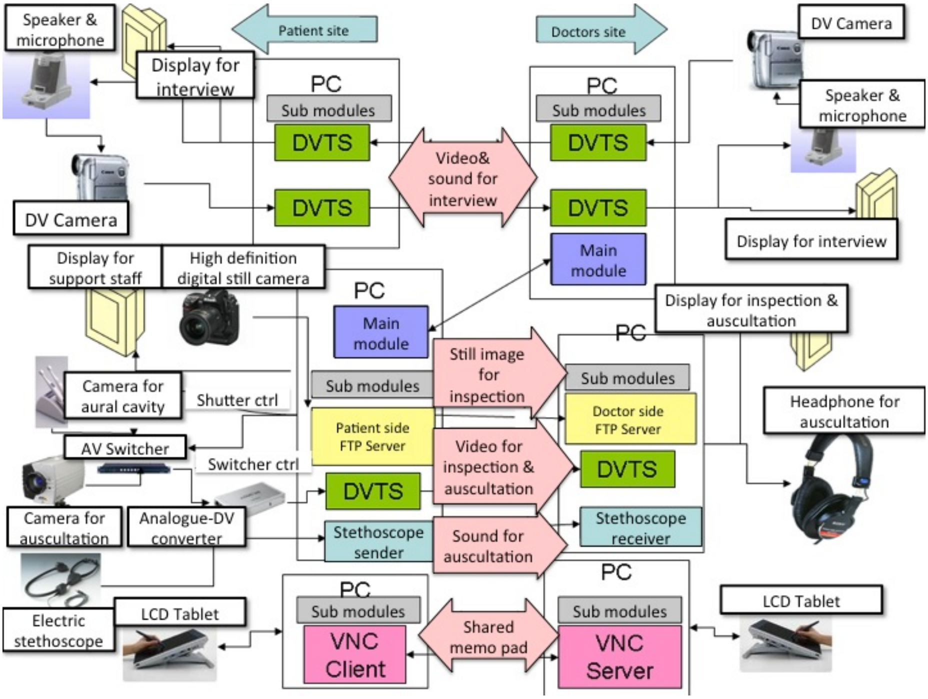
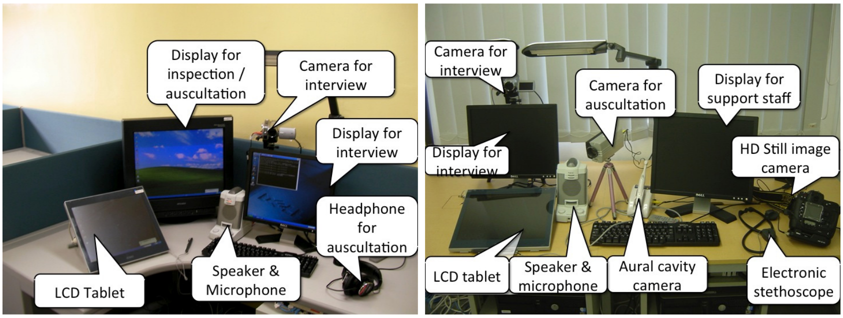


Fig. 1. The system configuration of the prototype



(a) Doctors site (b) Patient site

Fig. 2. The Interface of the prototype

Table 1. The components of the prototype

Component	Type
PC	DELL Optiplex 745
Display	DELL 19inch LCD Display
Speaker Microphone	NEC Voice Point Mini
DV camera	Panasonic NV-GS300-S
High definition still camera	Nikon D2x+WT-2
Aural cavity camera	Morita Penscope (modified)
Ausculation camera	AXIS 230 Mpeg2 Network Camera
Electric Stethoscope	Cardionics E-scope
LCD Tablet	Wacom DTI520
Headphone	Sony MDR-CD900ST

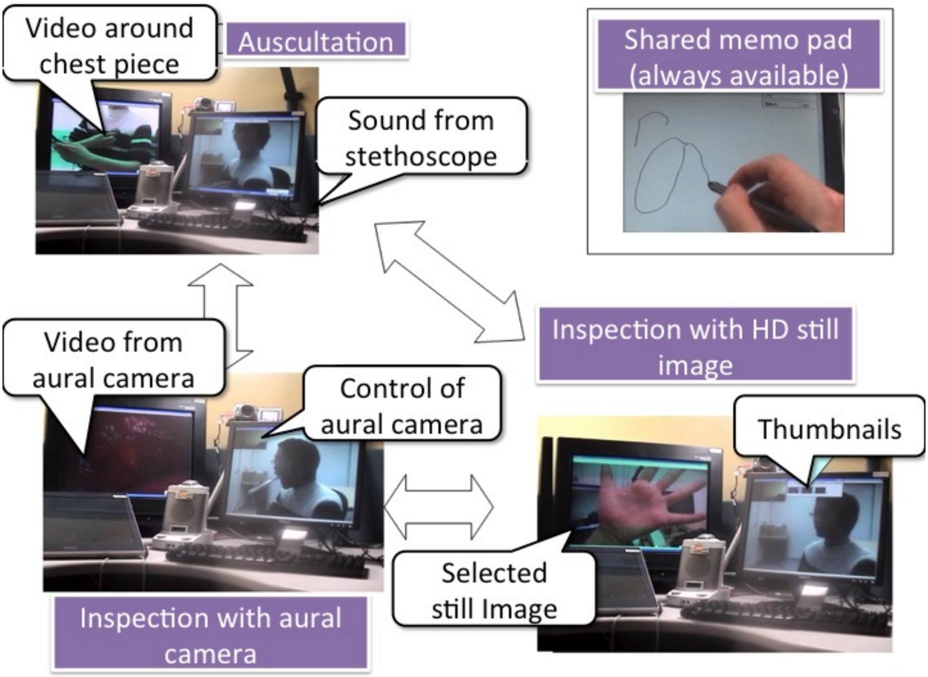


Fig. 3. The status transition diagram of the prototype

main office of the Kyoto University Health Service. Two terminals are connected to a private VPN on the campus information network of Kyoto University named KUINS-3[11], whose backbone is 1Gbps.

3 Results and Discussions

The initial introduction of the prototype clears that the most of the clinical staffs of the healthcare service division can handle the prototype after a few minutes of instruction. They can even restart whole system by themselves after some technical trouble. They confirmed that they can perform basic consultation using the prototype. However, the medical doctors claim that the prototype makes auscultation difficult due to its low sound quality and ill synchronization with video stream.

Along the results of initial introduction, the authors performed deeper analysis about problems on auscultation and difference between telemedicine and face-to-face consultation.

3.1 Problems of Auscultation

The auscultation is multimodal process. The medical doctor feels the movement of thorax through their own hand to know expiration and inspiration. The doctor also feels slight movements of the chest piece of the stethoscope and ignores the noise caused by the chest piece rubbing the thorax. Under telemedicine, the doctor cannot handle the chest piece by his own hand, and, consequently, cannot feel such movements. Although some foregoing researches try to introduce maser-slave type robots into telemedicine, to ask a supporter next to the patient to place the chest piecec is more realistic technically and legally. Thus, we need to utilize visual key instead of haptic key.

Synchronization between multiple media is one of the most important factors for successful communication under multimodal telecommunication. The authors evaluated how the synchronization error effects on the auscultation. Two medical doctors performed pseudo tele-auscultation under several conditions. Fig.4 and 5 shows the scene of the pseudo tele-auscultation. Here, one doctor plays patient and supporter, that is, he places the stethoscope on his own chest and breathes as another doctor asks. Tab.2 shows the conditions and the results. In the table, the two-tuple number in each cell gives evaluation of two doctors. Tab.3 shows the evaluation scores. The result clearly shows that the error of synchronization may harm auscultation. Especially, the auscultation of breath sound becomes impossible when video and sound differs 10% of the respiratory cycle. The medical doctors claims that even 100 milli-second differences may cause medical doctors to misunderstand expiration and inspiration.

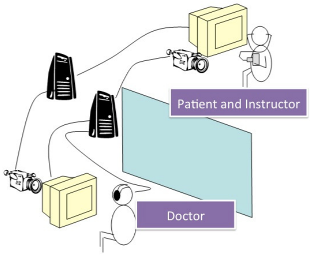


Fig. 4. Sketch of pseudo tele-auscultation


(a) Remote (doctor) site (b) patient/supporter site

Fig. 5. Snapshot of pseudo tele-auscultation

Table 2. Result of pseudo tele-auscultation

Delay between video and audio Status of sound	100 ms	200 ms	300 ms
Heart sound (normal 60 80/min)	5:5	4:3	3:3
Breath sound (slow 12 15/min)	5:5	5:4	2:2
Breath sound (fast 30 40/min)	4:5	1:1	1:1

(Doctor 1's score):(Doctor 2's score)

Table 3. Evaluation score

Score	Evaluation
5	Doesn't notice
4	Notice but doesn't matter
3	Troublesome but doesn't affect diagnosis
2	May affect diagnosis
1	Impossible to diagnose

3.2 Comparing Telemedicine and Face-to-Face Consultation

For more detailed evaluation, the authors tried to expose the difference of telemedicine and conventional face-to-face consultation using protocol analysis[9].

Medical doctors are asked to perform pseudo consultation under the following scenario under telemedicine and face-to-face condition.

- Scenario 1: Diagnose a patient with suspected pneumonia by a respiratory specialist introduced by a general practitioner (GP).
    - GP at the patient's site diagnoses the patient.
    - GP introduces specialist at the remote site.
    - The specialist inspects the patient.
    - The specialist auscultates the patient.
    - The specialist explains the diagnostic result
- Sceario 2: Diagnose a patient with hand eczema when no doctor is at the patients site.
    - A nurse introduces a doctor at the remote site.
    - The doctor interviews the patient.
    - The doctor inspects the patient.
    - The doctor explains diagnosis and tells to visit dermatologist.

Whole session videotaped and analyzed. After the analysis the authors and medical doctors had a retrospective report session. Some typical results are shown in Tab.4 and 5.

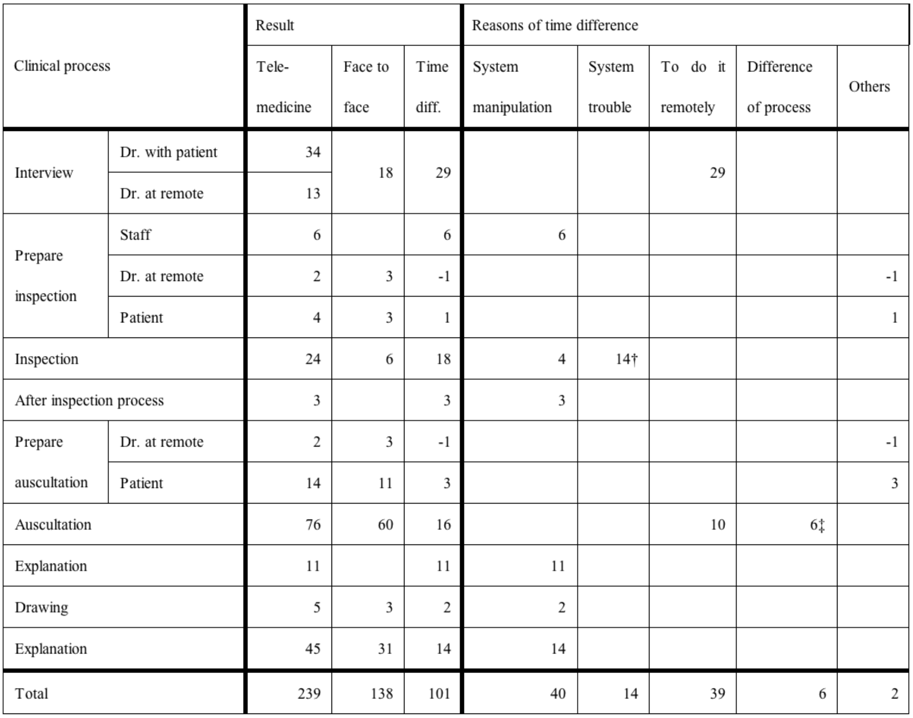
Although the clinical process itself didn't change, the result clearly shows that telemedicine requires more time to perform clinical process.

One main reason is inevitable actions to perform telemedicine, denoted "to do it remotely" in the tables. The clinical staffs at the patient's site need to tell the condition of the patient to the remote site, and the clinical staffs need to take pictures or videos to transmit still image or some additional data to the remote site. As all the data exchanged under telemedicine is need to be digitized[2], these additional tasks are inevitable for detailed process.

Another reason is the ineffectiveness of the prototype, denoted "system manipulation" in the tables. The clinical staffs sometimes have some trouble to find correct switches to start up required subsystems, and the doctors need to concentrate drawing a picture due to slippery surface of the tablet. Such problems can be solved by tunings and trainings.

The other problem is the communication difficulties over TV-phone. Under conventional consultation, the doctor just needs to take a paper out from his desk to tell the patient that he start to explain the diagnostic result. But under telemedicine, the doctor needs to speak out what he will do and ask the patient to look onto the LCD tablet. The clinical staffs need to speak out to know how the process is going at another site. As a matter of course, experience to perform telemedicine may dramatically decrease such additional tasks. Actually, the clinical staffs required fewer checkups at the end of the session. However, as patients are always new, the messages or orders for the patients won't be reduced. Additional technologies to share the atmosphere or non-verbal communication such as share AR technology may be required to smoothen the communication.

Table 4. Test result of scenario 1



Clinical process		Result			Reasons of time difference				
		Tele-medicine	Face to face	Time diff.	System manipulation	System trouble	To do it remotely	Difference of process	Others
Interview	Dr. with patient	34	18	29			29		
	Dr. at remote	13							
Prepare inspection	Staff	6		6	6				
	Dr. at remote	2	3	-1					-1
	Patient	4	3	1					1
Inspection		24	6	18	4	14†			
After inspection process		3		3	3				
Prepare auscultation	Dr. at remote	2	3	-1					-1
	Patient	14	11	3					3
Auscultation		76	60	16			10	6‡	
Explanation		11		11	11				
Drawing		5	3	2	2				
Explanation		45	31	14	14				
Total		239	138	101	40	14	39	6	2

※ unit: (second)

† Due to restarting system after system halt.

‡ Doctor auscultated two more points in telemedicine.

Table 5. Test result of scenario 2

Clinical process	Result			Reasons of time difference				
	Telemedicine	Face to face	Time difference	System manipulation	System trouble	To do it remotely	Difference of process	Others†
Explain patient condition	22		22			22		
Interview	12	8	4	4				
Prepare inspection	10		10			10		
Photo taking	17		17			17		
Photo quality check	9		9			9		
System manipulation	6		6	6				
Inspection	8	11	-3					-3
Interview	12	17	-5					-5
Inspection	13	5	8					8
Interview	9	5	4					4
Inspection	4	3	1					1
Interview	3	22	-19					-19
Inspection	6	9	-3					-3
Explanation	19	28	-9					-9‡
Total	190	108	42	10		48		-26

※unit: (second)

† Difference caused by personality of doctors, so it may counted as difference of process, too.

‡ In face-to-face process, the doctor explains diagnosis to patient during inspecting the affected area for eight seconds.

4 Conclusions

This paper developed a multimedia communication system to support realtime telemedicine and evaluated communication problems under telemedicine. The results tell that the conventional multimedia communication system may support providing clinical services at the same level as conventional face-to-face consultation. However, the evaluation tells that the erroneous synchronization and the lack of multimodal communication may harm telemedicine. As the requirements for communication quality is quite dependent on clinical process and purpose of each media, a specially designed application-level QoS (Quality of Service) control communication toolkit, as the one Mori et al[5] proposed, may be indispensable. For smooth telemedicine, the system needs to provide certain alternative methods to transmit atmosphere or non-verbal information, as the one Suenaga et al[10] proposed.

Acknowledgements. The authors would like to thank our brillient students, Mr. Takashi Nakai and Mr. Takashi Tsukasa, for their precious support to develop and introduce the prototype into real clinical environment. The authors also thank Morita MFG. Corp. who developed the aural cavity camera. The authors also thank Dr. Kazuo Nakazawa and Dr. Ryo Haraguchi of National Cerebral and Cardiovascular Center for their advice to develop LCD tablet-based shared memo pad. At last, the authors would like to thank all the staffs of the Kyoto University Health Service for their support for all the evaluations. This paper partly funded by Kyoto University Rector's Funding Program and Initiatives for Attractive Education in Graduate Schools program of JSPS.

References

1. Goldberg, M.A.: Teleradiology and telemedicine. Radiolo. Clin. North Am. 34(3), 647–665 (1996)
2. Hori, K.: Information support for telesurgery. In: Kumar, S., Marescaux, J. (eds.) Telesurg., pp. 101–111. Springer (2008)
3. Li, J.S., Zhou, T.S., Chu, K., Araki, K., Yoshihara, H.: Design and development of an international clinical data exchange system: the international layer function of the dolphin project. J. Am. Med. Inform. Assoc. 18(5), 683–689 (2011)
4. Marescaux, J., Leroy, J., Gagne, M., Rubio, F., Mutter, D., Vix, M., Butner, S.E., Smith, M.K.: Transatlantic robot-assisted telesurgery. Nature 413, 379–380 (2001)
5. Mori, M., Tsukasa, T., Hori, K., Kuroda, T., Yoshihara, H.: Network QoS control system application for telemedicine. Trans. Jap. Soc. Med. Biolog. Eng. 42(4), 140–146 (2004)
6. Murphy, R.L.H., Bird, K.T.: Telediagnosis: a new community health resource: observations on the feasibility of telediagnosis based on 1000 patient transactions. Am. J. Pub. Health 64, 113–119 (1974)
7. Nakai, T., Tsukasa, T., Hori, K., Takemura, T., Kuroda, T., Ando, M., Kawamura, T., Yoshihara, H.: Application of real-time multimedica communication technology into health care service. Jap. J. Telemed. Telecare 4(1), 35–40 (2008) (Japanese)

8. Ogawa, A., Kobayashi, K., Sugiura, K., Nakamura, O., Murai, J.: Design and implementation of DV based video over RTP. In: Packet Video Workshop (2000)
9. Rubin, J.: Handbook of Usability Testing: How to Plan, Design, and Conducting Effective Tests. John Wiley & Sons (1994)
10. Suenaga, T., Kuroda, T., Manabe, Y., Chihara, K.: A tele-instruction system for real-time telemedicine. In: Proc. Comp. Aid. Radiol. Surg., pp. 498–503 (2000)
11. Takakura, H., Ebara, Y., Miyazaki, S., Sawada, A., Nakamura, M., Okabe, Y.: Structure and security of secure gigabit network system, KUINS-III. IEICT Trans. Comm. J86-B(8), 1494–1501 (2003) (Japanese)
12. Weinstein, R.S., Bhattacharyya, A.K., Graham, A.R., Davis, J.R.: Telepathology: a ten-year progeress report. Hum. Pathol. 28(1), 1–7 (1997)

Japanese Family via Webcam: An Ethnographic Study of Cross-Spatial Interactions

Chiho Sunakawa

Department of Anthropology, The University of Texas at Austin,
1 University Station C3200
Austin, Texas 78712
`sunakawa@utexas.edu`

Abstract. This paper investigates an ethnographic understanding of family relationships across spaces and the management of discursive practices using a webcam. Through turn-by-turn analyses of video-recorded webcam-mediated conversations between Japanese families who live in the United States and their extended family members in Japan, I analyze how the reciprocal expectations of showing and watching each other's spaces are woven into unfolding processes in webcam-mediated interactions. I specifically focus on the organizational features of a 'show-and-narrate' activity in which and through which aspects of everyday lives are introduced. I examine how show-and-narrate activities are discursively marked, how children's interactional behaviors are structured, and how participants are socialized into this technologically mediated family space.

Keywords: Japan, family, ethnography, video-mediated communication, socialization, multimodal interaction.

1 Introduction

As communication technologies develop, the webcam is becoming a prevalent tool for geographically distributed family members. With a webcam, families can co-construct emotionally close interactional spaces. This emerging communication technology, which permits the temporary juxtaposition of physically distant spaces, creates a context for families to reconcile local and cultural differences, and establish a center for emotional attachment [1]. In this view, a webcam is not merely a virtual technology detached from the "real" world that replaces other means of family interaction such as telephones, letters, or face-to-face meetings. Rather, it is a "cultural product" [2] situated in local interactional spaces through which participants manage interactions within and across spaces.

Drawing on approaches in linguistic anthropology, the goal of this paper is to provide an ethnographic understanding of family relationships across spaces and the management of discursive practices using a webcam. Audio- and video-recorded data is derived from interviews and webcam-mediated conversations between Japanese families who live in the United States and their extended family members in Japan.

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 264–276, 2012.

Through turn-by-turn analyses I analyze how the reciprocal expectations of *showing* and *watching* children are achieved in unfolding processes of talking in webcam-mediated interactions. I specifically focus on the organizational features of temporally formed, collaborative exchanges that I call "show-and-narrate" activities in which and through which aspects of children's everyday lives are introduced to their grandparents. I examine how show-and-narrate activities are discursively marked, how children's interactional behaviors are structured, and how participants are socialized into this technologically mediated family space.

As a point of analytical departure, I situate talking-in-webcam-interaction within a built domestic environment. While canonical concerns in linguistic anthropology emphasize the symbolic nature of linguistic systems and the orderliness of talking across speakers, relations between language use and the surrounding material world have not yet been paid much attention except for a few studies focusing on the phenomenal field of dexis and referential practices [3], [4] and the inextricable relation between language use—from honorifics to greetings—and the place in which participants position themselves in order to accomplish communicative activities [5], [6], [7]. Recent interdisciplinary interest in multimodality and embodied interactions that address the inherently multimodal nature of discursive practices allow for a better understanding of how human interaction adapts to built-environments [8], [9], [10], [11], [12], [13], [14]. It has been argued that language, gestures, properties of space, and relevant material objects are coordinated in order to achieve a particular goal of the activity in which participants engage. Given that webcam interactions always require material objects or "cognitive artifacts" [15]—in this case, a computer, a webcam, and an internet access—with which participants calibrate their behaviors, it is crucial to examine what consequences webcam affordances have for the achievement of mediated family interactions. Following the idea that activity is a central frame within which language operates [16], [17], [18], I analyze how face-to-computer spaces are managed in a show-and-narrate activity by coordinating a range of resources including language, bodily behaviors, material objects, and spatial properties.

2 Data and Method

The data for this study is derived from three years of ethnographic fieldwork conducted in Japan and in the United States. It consists of ethnographic observation, video-recorded webcam conversations, and semi-structured interviews about the use of communication technologies. For this paper I focus on recordings made in two different families. In each family setting, webcam conversations lasted for about 40 to 50 minutes, and took place at each participant's respective house. Both family settings include young children who are the center of attention in the course of conversation.

When I recorded the conversations, I attempted to capture the screen activities as well as the discursive practices taking place in one of the two physical spaces. While recent virtual ethnographers [19] innovatively question the traditional notion of a field site as a localized space isolating "virtual" and "real" spaces, I take into consideration that the local contexts where webcam-mediated conversations occur provide ethnographic contexts for how participants accomplish their interactions. For diasporic Japanese families, visual contact with kin members in remote places provides new opportunities to maintain familial as well as cultural connections. As one of the participants told me in an interview, they appreciate having the means to actually see and interact with their family members, as well as to see spaces such as parts of the houses and rooms in which they grew up. This suggests that these pieces of visual information, in addition to interactions with family members, have an impact on participants' identity formation. Additionally, children participating in conversations tend to be the focus of the video frame. All the participants acknowledged in the interviews that they are motivated to use a webcam because of children within their family. With the webcam, family members in distant locations are not only able to interact with the children but also share the experience of observing children's growth with adult kin members.

3 Show-and-Narrate Activities

Before analyzing interactional features of show-and-narrate activities, I describe the reciprocal expectations that are taken for granted among family members. Regardless of the large amount of talk, participants describe webcam experiences by using phrases such as *miru* ('to watch') and *miseru* ('to show'). Webcam does not replace either *talking* or *meeting,* but it facilitates the acts of *showing* and *watching.* Given that the important conceptualization of visual access to others is realized in Japanese expressions for 'to visit', *kao o miru* ('watch face [of the visited]') and 'to be visited', *kao o miseru* ('show face [to the visitor]'), it is not surprising that participants' understandings of webcam interactions highlight the aspect of showing and watching faces of family members.

As children grow older, caregivers actively involve them in the very process of showing and consequently have them describe what they are showing. This particular communicative activity that I call 'show-and-narrate' leads to an important opportunity for children, grandparents, and other relatives to interact across spaces. Even though show-and-narrate activities also happy in face-to-face settings, the focus of this paper is not to compare such interactions with webcam-mediated ones, but to explore the special affordances webcams have for such interactions.

A prototypical show-and-narrate activity consists of three components: (1) showing an embodied focal point, (2) providing a verbal description of it, and (3) receiving a response from audiences. In the following excerpt, for example, three-year-old Hugo is showing his favorite toy cars to his grandparents through a webcam. In line 1, he holds a toy baggage tractor and brings it toward the webcam embedded in the frame of a laptop. Immediately after this movement, Mother provides the name of tow tractor. This narration is followed by Grandpa's short responses of *hai ha::i* ('yes, yes').

Example 1: Three components of show-and-narrate activity

1.	Hugo:	`((brings a toy car to the webcam))`	Showing 
2.	M:	`Korewa:: tooingu torakutaa de::su` This is a tow tractor.	Narrating
3.	GP:	`Hai ha::i` Yes, yes.	Responding

The organization of these three components is contingent upon contexts. Embodied focal points can be material objects, performances, and narratives about aspects of the child's everyday activities. A narration can lead to a forthcoming event, overlap with what is being shown on the screen, and shadow what is being done. The order of the three components may vary, but regardless of its complexity, a show-and-narrate activity is temporally formed, repeatedly embedded, and strategically orchestrated in order to elicit children's participation in webcam interactions. Similar to family narratives at dinner tables [20], adults structure children's interactions in certain ways to share stories and experiences while socializing children into a technologically mediated space. Children not only learn to launch and construct discursive practices, but also learn to transform communicative habitus [21] from a co-present space to a mediated space.

Questions arise**:** How are show-and-narrate activities embedded in the stream of a webcam conversation? What discursive resources do participants rely on to launch and achieve the activity? How are children socialized into this activity? In the rest of this paper, I explore these questions by focusing on a range of multimodal resources including the adult use of *desu/masu* polite sentence final particles, the quotative marker *tte,* repetitions, prosodic features, and gestures.

3.1 The Use of *Desu/Masu* Forms

As exemplified in Example 1, Hugo and Mother talking to Hugo's grandparents co-construct a highly cooperative, paired sequence consisting of Hugo's bodily act of showing a toy and Mother's providing its name. These two acts are encapsulated in adjacent or overlapping turns and they make a distinctive unit of sequence from other types of discursive practices such as talk between Mother and the grandparents. In Example 1, the use of a polite sentence final form and its intonation contour contribute to marking this show-and-narrate sequence.

Line 2 in Example 1:

```
M:     Kore wa:: tooingu torakutaa      de::su
       This.TOP.  tow(ing) tractor  POL.
```

This is a tow tractor.

Immediately after Hugo shows a toy tow tractor, Mother provides a descriptive utterance using a *desu* at the end with a stretched middle vowel, *e::*. *Desu* and *masu* forms are typically known as addressee honorific sentence final forms [22], [23] that index politeness toward the addressee. However, the introduction of *desu/masu* forms here is not a reflection of politeness toward the addressee. Instead it spotlights a mother-child activity encapsulated in a show-and-narrate sequence. This type of *desu/masu* form suggests that the formal nature of the polite form is associated with the public nature of the show-and-narrate activity.

While the use of plain forms is the norm in interactions between Japanese caregivers and children, research with sociolinguistic and ethnographic approaches reveals that the use of polite forms or the switch between the plain and polite forms are commonly observed in family settings [24], [25], [26]. This is because *masu* forms are not just honorific expressions, but can be resources for indexing social contexts, norms, and identities incorporated in conversations. Drawing on two ideological notions of *uchi* (‘inside/in-group) and *soto* (‘outside/out-group’), Cook [26], [27] suggests that while a *masu* form indexes politeness towards different types of addressees and reflects social norms in a *soto* context, it represents a speaker’s public self in an *uchi* context. For instance, *desu/masu* usages in speech between caregivers and children who belong to an *uchi* indicate a caregivers’ responsibility such as serving food, an authority figure such as the child’s doctor, and an “on-stage display” [26: 704] of a particular social role such as a presenter of a guessing game. Similar to this, the use of *desu* in line 2 creates a mediated ‘stage’ where fragmented aspects of a family’s everyday life are publicly introduced, spotlighted, talked about, and shared.

3.2 Socialization to the Act of “Showing”

To elicit the child’s participation in a show-and-narrate activity, caregivers actively and strategically manage children’s interactions with various strategies including prompting, asking, and providing a narration. For example, in Example 2, Mother provides a narration statement as she says “to Grandpa and Grandma, ‘this is a passenger step’-*tte*”. *Tte* is a quotative marker and it suggests that Mother’s utterance does not mean “say ‘this is a passenger step’ to Grandpa and Grandma,” but it invites Hugo’s gesture of showing a passenger step. By providing the name of the not-yet-seen toy car, and marking the name with a quotative *tte*, Mother secures a forthcoming turn for Hugo to provide an act of showing the toy. Structured in this path, Hugo learns how to appropriately respond to Grandpa and Grandma. This type of highly structured sequence of interaction between a mother and a child can be found in a range of ethnographic studies on language socialization. For instance, Schieffelin’s work in Kaluli [28] illustrates that mothers use a particular discourse marker, *ɛlɛma* (“say like that”), in order to provide children with specific utterances made to specific addressees. While in Kaluli, children are expected to repeat the utterances after the *ɛlɛma*, Hugo in this example is expected to *act* properly rather than *repeat* what Mother says.

Example 2: Elicitation of participation

1. M: Hugo-kun, Hugo-kun jiiji to baaba ni::
Hugo, Hugo, to Grandpa and Grandma

2. (0.5)

3. kore wa::=
This is

4. GP: =hai misete
Yes show it

5. M: Passenja:: suteppu dayo **tte**
"This is a passenger step" QT

During a show-and-narrate activity, children's courses of interaction are constantly attended and corrected by co-present caregivers. The following excerpt exemplifies a process in which Hugo appropriately and successfully shows his toys to Grandpa and Grandma via a webcam while learning to distinguish different webcam affordances, particularly the difference between the location of the webcam, and visual representations of self and others.

Example 3: Socialization to webcam affordances

1. M: hhhh soko ja nai soko ja nai soko wa chotto chigau (.)
It's not there, it's not there. It's a bit wrong.

kocchi kocchi (.)kore kore kore
Here, here, this one, this one, this one
((points to the webcam; holds the arm position))

2. Hugo: ((holds two cars; brings them to the webcam))

Successful arm posture

Prior to the excerpt, Mother prompts Hugo and directs his attention to the webcam by repeatedly pointing to the webcam location and the self-image and describing differences. In line 1, since Hugo still keeps bringing his car close to the self-image, Mother this time emphasizes the right direction as she says "it's not there, it's a bit wrong. Here, here, this one, this one, this one." Mother's pointing gesture is held until Hugo brings his arms to the right position in line 2. Since Mother sits Hugo down on her lap facing toward the computer screen, her pointing gesture creates a particularly structured space in which Hugo makes the correct arm extension. This suggests that the Mother's pointing gesture in line 2 does not simply point to the right direction, but creates a perceptual field for Hugo. Goodwin [29] calls this "highlighting," a communicative strategy to elucidate particular knowledge inscribed in the built environment. In his work on expert-novice interaction among archeologists, Goodwin describes how archeologists draw a line with a trowel in the dirt to delineate an area in which professionally meaningful dirt features emerge in the course of interaction. Similar to this highlighting process, Mother's held gesture and its delineated space help Hugo successfully make his toy cars visible to his grandparents.

Hugo's understanding of webcam affordances is significant not simply in terms of achieving a show-and-narrate activity. It is ethnographically important because it provides a rich site for the grandparents to share a family experience of witnessing the child's growth and education. As they repeatedly told me in interviews, in a webcam conversation Hugo's grandparents always found things Hugo had come to be able to do with computers. Like language, technological skills such as manipulating a computer and navigating the internet are becoming common signs of children's socialization. The process delineated in this example is one of those socialization moments.

3.3 Showing a Performance of speaking

Similar to Hugo's grandparents, Sota's grandparents are interested in seeing a moment of socialization. In the next example, Sota and his mother co-construct a rhythmic sequence in order to *show* Sota performing his improved English. Focal points featured in show-and-narrate activities do not always consist of a material object and its description such as Hugo's toy and its name. They can be a range of performative aspects in everyday lives including singing a song the child learns at school and speaking in English. Sota is a four-year-old boy living in the United States with his parents, a Japanese father, and a Japanese-American mother. Since he was born, Sota has communicated with his grandparents and relatives from both sides by a webcam, but has not yet met in-person with his grandparents from his father's side. Sota speaks Japanese most of the time, but since he started to go to preschool, he has sometimes spoken English at home as well.

During a webcam interaction, Sota's parents organize the webcam space, provide an accompanist's role, and correct their children's discursive practices. As Sota is usually running all over the house in the course of webcam interaction, his parents prepare an "on-stage space" around the sofa area in order for Sota to stay still when he is asked to say, do, and show something for his grandparents (Figure 1). In other words, while Hugo's mother incorporates the public nature of the activity by using a polite sentence final form, Sota's parents do so by creating a public space around the sofa.

Fig. 1. "On-stage" space

Prior to Example 3, Mother and Father tried to have Sota demonstrate his improved English by singing in English and listing his teachers' names. Since Sota sang too fast, and he gave an incoherent list of names, Mother modifies the task and asks about friends' names instead of teachers' names.

Example 4: Repetition

1. M: Sota, sota, otomodachi no namae nani?
Sota, Sota, what are your friends' names?

2. F: a otomodachi no namae iou
Yes, let's say friends' names.

3. Sota: ↑Jo::n Pie:rre:↑
John Pierre
Showing a performance of speaking English

4. M: ↑Jean Pierre:::↑ hokawa?
"Jean Pierre." Who else?
Narrating

5. Sota: Mi::- ↑E::ika::↑
Mi-, Erika.
Showing a performance of speaking English

6. GM: ((leans forward to ‘see’ Sota’s performance of speaking English))

Responding (bodily)

7. M: ↑Erika::↑ hoka wa dare ga iru?
“Erika,” and who else do you have?
Narrating

8. Sota: ↑Ma::na::↑
Mana
Showing a performance of speaking English

9. M: ↑Ma::na::↑
hokawa dare ga ita kke?
“Mana.” Who else do you remember?
Narrating

10. Sota: ↑Je:ise::n↑
Jason
Showing a performance of speaking English

11. M: ↑Ja::so::n↑
hokawa dare ga iru?“
Jason,” and who else do you have?
Narrating

12. Sato: ↑Jo::n Pi::ee:↑
Jean Pierre
Showing a performance of speaking English

13. M: ”Jean Pie::rre”
Jean Pierre
Narrating

14. (0.5)
15. Sensei no namae wa?
And your teacher’s name?

16. Sota: ↑Misu Ja::ne↑
Miss Jane
Showing a performance of speaking English

17. M: ↑Miss Jane↑
mouhitori no sensei no namae wa?
“Miss Jane.” Another teacher’s name?
Narrating

18. Sota: `↑Mis Magi::↑`
Miss Maggie

Showing a performance of speaking English

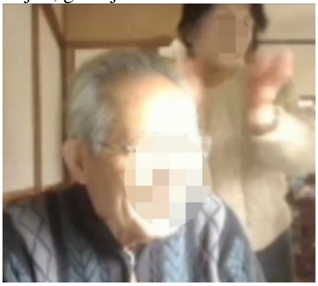
19. M: `↑Miss Maggie::↑`
`ieta ieta::`
“Miss Maggie!” You said it, you said it (all).

Narrating

20. F: `Ieta ne ippai`
You said a lot of names.

Responding

21. GM: `jo:o:zu jo::zu ((clapping her hands))`
Good job, good job.



22. `oba::chan wakaran kattakedo jo::zu jo::zu`
I couldn’t understand them, but you did it well, did it well.

23. M: `Sota, arigato:::tte`
Sota, say “thank you.”

24. Sota: `Ariato`
Thank you.

This time, Mother collaborates with Sota by repeating the names and by shadowing Sota’s prosodic contours, a rising intonation and a stretched middle vowel. A slight difference is that when she echoes Sota, Mother adds an American accent to articulate names correctly. After the repetition, she adds a facilitating question of “who else?” to solicit more names. Once Mother and Sota start listing names, Grandma leans towards the screen as if she tries to “see” this coordinated performance of listing names in English (line 6).

Prosodic features of Mother’s repetitions have intriguing implications. From a linguistic socialization point of view, mother’s repetition is a type of elicited imitation or “glossing after the fact” [30] where the child’s utterance is repeated in a form that is slightly modified by being paraphrased, translated, or explained by adults. This modification aims to familiarize the child with adult patterns of discursive negotiations [31]. However, Sota’s mother’s repetition does not simply model for the child the appropriate pronunciations of names in American English, but also helps represent the improvement of Sota’s English for the grandparents. Mother’s repetition emphasizing the English pronunciation supplies an interpretation of Sota’s performance of speaking in English.

4 Conclusion

In this paper, I provided ethnographic accounts of visual co-presence in webcam-mediated family interactions. The reciprocal expectation of showing and watching children was reflected in how participants described webcam experiences, and how they managed face-to-computer spaces involving various resources for interaction including language, the body, spatial properties, and webcam affordances.

The importance of acts of showing and watching each other's space was sequentially incorporated in a collaborative exchange that I called a 'show-and-narrate' activity. This activity, which was marked by the use of polite sentence final form, prosodic features, and repetitions, had three components: the acts of showing, narrating, and responding. Participants collaboratively and temporarily arranged these elements in many ways in order to demonstrate and describe selected aspects of their everyday lives. Particularly, children were encouraged to participate in this activity and socialized to the mediated presence and expectations of others as their discursive and bodily behaviors were structured in particular ways. The show-and-narrate activity had powerful consequences for sharing experiences and establishing family relationships and rapport.

Based on these findings I have suggested that the webcam, a relatively new tool, created a site for establishing a communication field where conventionalized understandings of family relationships were juxtaposed with emerging habits of communication. An important contribution of this study is that my analyses reconsider what multimodality means in a sociotechnical environment. As participants learn to navigate in a webcam-mediated space, their conventional ways of using a range of semiotic means are contested and reconfigured. Creating visual joint foci, for example, does not only consist of coordinating eye gazes, gestures, and speech, but also creating a new type of communicative practice that occurs in a performance space made by the use of the webcam. In other words, meanings from one modality such as vision are transferred into a mediated space by organizing and incorporating webcam features into the process of constructing interactions.

Families highlight the process of becoming familiar with webcam interactions, in which communicative practices and their social meanings are transformed and shaped in interesting ways. While emerging habitus can be learned over time and can contribute to "the production and interpretation of predictable and coherent next actions" [32: 332], family members across generations reproduce significant values of traditional family practices, such as the responsibility of parents to connect children to grandparents, by way of webcam.

References

1. Kondo, D.: Crafting Selves: Power, Gender, and Discourses of Identity in a Japanese Workplace. University of Chicago Press, Chicago (1990)
2. Duranti, A.: Linguistic Anthropology. Cambridge University Press, Cambridge (1997)
3. Hanks, W.F.: Referential Practice: Language and Lived Space among the Maya. University of Chicago Press, Chicago (1990)
4. Hanks, W.F.: Explorations in the deictic field. Current Anthropology 46(2), 191–220 (2005)

5. Duranti, A.: Language and bodies in social space: Samoan ceremonial greetings. American Anthropologist 94(3), 657–691 (1992)
6. Duranti, A.: Universal and culture-specific properties of greetings. Journal of Linguistic Anthropology 7(1), 63–97 (1997)
7. Keating, E.: Power Sharing. Oxford University Press, New York (1998)
8. LeBaron, C., Streeck, J.: Gestures, Knowledge, and the World. In: McNeill, D. (ed.) Language and Gesture: Windown into Thought and Action, pp. 118–138. Cambridge University Press, Cambridge (2000)
9. Suchman, L.: Technologies of accountability: Of lizards and aeroplanes. In: Button, G. (ed.) Technology in Working Order: Studies of Work, Interaction, and Technology, pp. 113–126. Routledge, London (1992)
10. Mondada, L.: Working with Video: How Surgeons Produce Video Records of Their Actions. Visual Studies 18, 58–73 (2003)
11. Stivers, T., Sidnell, J.: Introduction: Multimodal Interaction. Semiotica 156, 1–20 (2005)
12. Goodwin, M.H.: The Hidden Life of Girls: Games of Stance, Status, and Exclusion. Blackwell Publishing, Malden (2006)
13. Keating, E.: Space shifting: New technologies, new opportunities. In: Proceedings of the Symposium about Language and Society, Austin, TX (2008)
14. Streeck, J., Goodwin, C., LeBaron, C. (eds.): Embodied Interaction: Language and Body in the Material World. Cambridge University Press, Cambridge (2011)
15. Hutchins, E.: The distributed cognition perspective on human interaction. In: Enfield, N., Levinson, S. (eds.) Roots of Human Sociality: Culture, Cognition and Interaction, pp. 375–398. Berg, Oxford (2006)
16. Vygotsky, L.S.: Mind in Society: The Developoment of Hihger Psychological Processes. Harvard University Press, Cambridge (1978)
17. Keating, E.: Homo prostheticus: problematizing the notions of activity and computer-mediated interaction. Discourse Studies 7(4-5), 517–545 (2005)
18. Wittgenstein, L.: Philosophical investigations. Blackwell Publishing, Malden (1953/2001)
19. Hine, C.: Virtual Ethnography. Sage Publications Ltd., London (2000)
20. Ochs, E., Capps, L.: Living Narrative: Creating Lives in Everyday Storytelling. Harvard University Press, Cambridge (2001)
21. Bourdieu, P.: Outline of a Theory of Practice. Cambridge University Press, Cambridge (1977)
22. Martin, S.: Speech levels in Japan and Korea. In: Hymes, D. (ed.) Language in Culture and Society, pp. 407–415. Harper and Row, New York (1964)
23. Comrie, B.: Linguistic politeness axes: Speaker-addressee, speaker-referent, speaker-bystander. Pragmatics Microfiche 1(7), 1–12 (1976)
24. Sukle, R.: Uchi/Soto: Choices in directive speech acts in Japanese. In: Jane, J., Bachnik, M., Quinn, C.J. (eds.) Situated Meaning: Inside and Outside in Japanese Self, Society, and Language, pp. 113–142. Princeton University Press, Princeton (1994)
25. Maynard, S.: Discourse Modality: Subjectivity, Emotion and Voice in the Japanese Language. Benjamins, Amsterdam (1993)
26. Cook, H.M.: The role of the Japanese masu form in caregiver-child conversation. Journal of Pragmatics 28, 695–718 (1997)
27. Cook, H.M.: Socializing Identities through Speech Style: Learners of Japanese as a Foreign Language. Multilingual Matters Ltd., Bristol (2008)
28. Schieffelin, B.: The Give and Take of Everyday Life: Language Socialization of Kaluli Children. Fenestra Books, Tuscon (1990/2005)
29. Goodwin, C.: Professional vision. American Anthropologist 96(3), 606–633 (1994)

30. Scollon, R., Scollon, S.: Narrative, Literacy, and Face in Interethnic Communication. Ablex, Norwood (1981)
31. Scollon, S.: Reality Set, Socialization and Linguistic Convergence, Ph.D. Dissertation. University of Hawaii, Hawaii (1982)
32. Keating, E., Sunakawa, C.: Participation cues: Coordinated activity and collaboration in complex online gaming worlds. Language in Society 39(3), 331–356 (2010)

Appendix: Transcription Convention

[	A left bracket indicates a start of overlap
:	A colon is used to indicate the prolongation or stretching of the sound just preceding it.
=	Equals signs connect two lines with no discernible silence between them or the following utterance latches onto the first utterance.
(0. 0)	Numbers in parentheses indicate silence represented in tenths of a second.
(.)	A dot in parentheses indicates a micropause less than 0.2 second.
↑	An upper arrow marks relatively sharper intonation rises than the surrounding utterances.
(( ))	Double parentheses include researcher's descriptions of non-linguistic events.

Designing a Future Space in Real Spaces: Transforming the Heterogeneous Representations of a “Not Yet Existing” Object

Katsuya Takanashi[1,2] and Takeshi Hiramoto[2]

[1] PRESTO, Japan Science and Technology Agency
[2] Academic Center for Computing and Media Studies, Kyoto University
Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501 Japan
takanasi@ar.media.kyoto-u.ac.jp

Abstract. Based on ethnography in a scientific museum, this article analyzes a single case of successive group problem-solving interactions for designing an exhibition as a future space. In the process of multimodal interaction in multiple spaces, heterogeneous resources in real spaces became representations of objects not yet existing and were continuously transformed with reference to a framework for collaborative problem-solving.

Keywords: Representation, Object not yet existing, Group problem solving, Multimodality, Space.

1 Introduction

It is inevitable for analyses of multimodal interaction to incorporate the material aspects of the environment in which the interaction is conducted [12], though the material world itself becomes semiotic resources [3] only when it is referred to in the course of interaction. In this article, we will examine the use of materials in human cooperative activities in terms of representations. Such representations are embedded in both interactional contexts and material environments, and serve as bridges between them.

2 Representing “Not Yet Existing” Objects

One of the most important things for participants in meetings for planning future activities, designing things, such as museum exhibitions, or founding new venture companies is how to negotiate and share images [9] of “not yet existing” objects among participants who have different occupations with different knowledge, specialties, experiences, and interests. This is in contrast to, for instance, conferences for medical and nursing care in which most reports are about events that have already happened. When people discuss objects that do not yet exist, it is necessary to represent them in certain ways since it is impossible to observe the referents themselves directly.

M. Okumura, D. Bekki, and K. Satoh (Eds.): JSAI-isAI 2011, LNAI 7258, pp. 277–290, 2012.

In a story told in a conversation, a teller often “re-plays” a past event using his/her own body [4]. In such a situation, what is represented had existed before the replay. In contrast, when collaborating on building an exhibition in a science museum, which is the research field of this article, members must and do represent several aspects of the referents that will exist in the future by using various resources. In general, among the Peircean classification of signs, an icon can represent a referent (designatum) by virtue of resemblance. Compared to representing past things, however, there is an inherent difficulty in representing an object that does not yet exist because members cannot ascertain the preciseness of a representation by relying on resemblance to a referent being represented. In addition, members building future things can even modify the referent itself to any degree depending on their own decisions in the discussion.

It is often the case that a problem for a group is found and shared by members during a meeting and what Murphy [9] terms “collaborative imaging” plays a significant role in the process. He defines collaborative imaging as a social, jointly-produced activity in which the objects of thought are created and manipulated by relying upon a number of semiotic resources in a shared space of face-to-face interaction, and he analyzed the activity of a team that was designing a scientific laboratory building in this regard. However, merely imaging collaboratively does not necessarily mean *resolving* the problems, and, more importantly, it is sometimes more difficult to *discover* a problem than it is to resolve it. As for the future, such problem discovery depends on how to predict unobservable and undesirable troubles that might occur, and resolving them corresponds to *protections* against them.

In addition, it is not unusual that a problem presented in a meeting cannot be solved within the same meeting for some spatiotemporal reasons. This article contains a case in which resources required to solve a problem do not exist in the meeting room, and members must therefore resolve the problem in another space on other day. It also be shown that assessment measures of resemblance are different for several different representations according to the phase at which each representation is used within the problem-solving framework of “problem presentation,” “examination,” and “resolution.”

3 Research Field and Its Spatial Characteristics

In this section, we will introduce the research field of our study and its spatial characteristics as background knowledge for understanding the case analysis in section 4.

3.1 Characteristics of the Exhibition

The National Museum of Emerging Science and Innovation (Miraikan) constructed a new, permanent installation named, “Songs of ANAGURA: Missing Researchers & Their remaining Devices” [11] in the “Information Science and Technology for Society” division and opened it to the public from August 21, 2011. This installation is an experience and interaction-based one in which visitors entering the space interact with information terminals and experience animations and songs.

This installation was constructed by assembling various heterogeneous technologies. The locations and movements of visitors are captured by an infrared radiation sensor system. These data are sent to a system operated by team that designed the content, and transformed into an animation that appears under the feet of the visitors called “*Me*” as his/her alter ego, which is projected from projectors installed on the ceiling [Fig. 1]. Among the fixtures, a terminal named “Deai (encounter)” assigns a visitor his/her own ID at the entrance of the exhibition and displays a “*Me*.” “*Me*” follows visitors wherever they walk, but if some problems occur in the sensed data, it might be “hijacked,” namely, an ID for a “*Me*” might be interchanged with that for another or lost completely. This “hijack” problem remained a big problem among members during construction.

Fig. 1. “Me”

3.2 Designing a Space and Space for Design

Building this installation meant designing a new space. This installation was constructed through the cooperation of three sub-groups: the sensing, content creation, and space design groups, each having different specialties and interests. In other words, if a problem occurred during construction, it could not always be resolved in only a single subgroup, so it required being resolved through negotiation between the subgroups, depending on its nature. For instance, positioning computer terminals and other devices in the space not only affects the flow lines of visitors, but the terminals also become obstacles to capturing with the sensing system or projecting visual content on the floor and walls, thus requiring a detailed arrangement.

For cooperation and negotiation during the construction period, roundtable meetings (RT) were held every Friday afternoon for seven months before the completion of the exhibition. In RTs, members reported the week’s progress, schedules for the near future, and negotiated with each other if needed. Here, we can observe a typical activity cycle: problem presentation -> examination or work -> resolution (see Sec. 2), in which members negotiated in regards to what the future objects were to be by relying on various heterogeneous representations. In addition, in this cycle, whereas meetings were held in a meeting room distant from the worksite, examinations were sometimes done in the space that would become the exhibition itself. This is mostly because an examination generally requires the most resemble representations, and the one that existed at the time when the installation had not been completed yet was the installation-under-construction itself. Thus, the activities of designing a new space were conducted in the present real spaces.

4 Case Analysis

In this section, we will focus on an episode of validation of the sensing system and an arrangement of the apparatuses, held on the worksite on May 10. This validation can be regarded as a simulation. Though people often run a simulation in order to predict a future problem, the problem is that we sometimes do not know when and how a simulation will be needed and what the simulation must simulate. For this reason, we will focus not only on the validation of May 10 but will also examine the details of a discussion before the validation in which we attempt to comprehend how the necessity of validation was realized by members and how the way to validate was determined.

4.1 Arranging the Schedule in Preceding Meetings

During the twelfth RT on April 28, it was decided to validate the sensing system and arrange the location of the apparatuses on May 10. During the next RT on May 6, concrete steps and procedures for this validation and arrangement were negotiated and finished as follows [Fig 2]:

1. From 13:30: An examination into the risk of "hijacks" around a "*Deai*" terminal by using four of the system's sensor devices
2. From 15:30: An arrangement of locations for the "*Deai*" terminal and the other apparatuses around it by placing Miraikan staff members in front of all the terminals

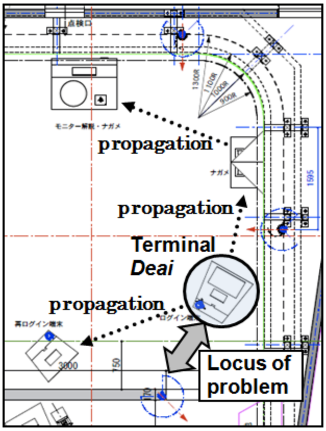


Fig. 2. Positions of apparatuses and the locus of the problem

How best to handle Task 2 depended on the results of Task 1. In this sense, Task 1, which was primarily a task for the sensor subgroup, had some propagation effects on the space design group and therefore had to be done before Task 2.

4.2 Using Representations in "Problem Presentations" in Meetings

We will trace here how the focus of and the way of validating the sensing system and the procedure for reflecting the results on the arrangement of apparatuses were negotiated in the RT on May 6.

In the RT [Excerpt 1], when Yamada, a manager of exhibition techniques who belongs to Miraikan, requests that he must receive information for determining the locations of apparatuses immediately after the validation (lines 01 and 03), Akagi, a person in charge of the sensing system, said, "Perhaps What is the biggest concern is here," as he starts pointing with his pen to an area around the "*Deai*" terminal in the scale model of the exhibition (lines 08, 10, 11). The active reaction of other participants to this pointing gesture, standing up and looking into the model, is elicited not solely by the gesture itself but also by the verbal expression "ichiban kininaru no wa (what is the most worrying is)" (lines 08-10), which is a fixed expression that displays concern about an undesirable future event. In this sense, this pointing does not only focus on a location in the model but also "presents a problem" to the group in the ongoing interaction at the same time [6]. In addition, using a scale model is effective for locating a point from an overhead view.

Following that, Akagi moves his pen in the model and represents the movement of a visitor in the installation while saying, "Perhaps what is the most worrying is, when a person is standing here, while someone else is passing through behind him, how often hijacks or some other problems will occur" (lines 16 and 18)(representation 1)[Fig. 3]. This is the first time the movement of "passing-through" is represented by a member in the course of the collaborative problem-solving. This representation is an icon that expresses the spatial configuration of the installation and the location of the apparatuses in it at a smaller scale, and it depicts the movement of a visitor by using the movement of a pen, which cannot be expressed only by the static model [9].

What is more important here is that this gesture accompanies an utterance in which the fixed expression "ichiban kininaru no wa (What is the biggest concern is)" (line 16) here again projects a negative prediction about the future. In addition, "toki ni (if)" (line 18) marks the subjunctive mood and the suffix "-chau" (line 23) shows the speaker's negative attitude toward the event described by the verb "oki (occur)." By virtue of these expressions, this utterance is not understood as a *description* of what Akagi has observed before but as a *prediction* of what might happen in the future. As evidence, Yamada's response to it (line 19) is not the receipt of a report but a display of his empathy for Akagi's concern.

Excerpt 1: Representation 1

Yamada: 01: .hh Jyuuki no ichi wo kettei suru made no:,
'.hh (To) fix the location of fixtures'
Akagi: 02: hai 'Uh-huh'
Yamada: 03: kettei suru tame no:, etto zairyoo wa:, sono ba de hosii
'we need evidence to::, fix the location of fixtures at the site.'
Akagi: 04: ((Akagi moves the plastic bottle away from his mouth))
.hh ¥Sono ba de ho¥sii 'At the site.'

Yamada: 05: de saisyuu teki ni, etto mukou no sensaa wo korosu noka, kochi no
sensaa wo korosu noka tteiuno no handan wa:,
‘And, to decide which sensor to choose and which one to discard,’

Yamada: 06: mochikaette kentou site (.) bunseki site itadaita hou ga ii youna ki ga suru.
‘You should take the data back with you and analyze it, I think.’

07 (0.3)

Akagi: 08: .hh eeto sou suruto jyuuki no ¥ichi tte naruto:¥ tabun ichiban (0.4) eeto::
‘Let me see, then, in terms of the location of fixtures, perhaps what is most,’
((Akagi stands up and starts pointing inside the model with his pen))

09 (2.0)

Akagi: 10: kininaru no wa (0.7) <ko>no ichi nan: (.) dato omoun desu yo ne (1.8)
‘worrying is: (0.7) supposed to be <h>ere.’
((Yamada and other members stand up and look into the model))

Akagi: 11: kono ichi [to:, ‘here and:, ’

Yamada: 12: [a, sou sou sou [sore desu sore desu.
‘Oh, right right right there.’

Akagi: 13: [kono ichi-kono ichi ga koko de tekitou nano ka
‘here, whether it’s appropriate to put this fixture here.’

Yamada: 14: sou ‘Yeah.’

Akagi: 15: tte iu hanashi nan da to omou n desu yo .hhhh de:::: kore wa::, etto::::: (0.3)
‘would be a problem. .hhhh and::::this::, well:::::’ (0.3)

-> Akagi: 16: tabun ima ichiban kini shiten no wa koko ni hito ga tatte ite:,
‘Perhaps our biggest concern is, when a person is standing here,’
((Akagi points at the certain spot on the model with his pen))

Representation of “passing-through” 1

Yamada: 17: un ‘Uh-huh.’

-> Akagi: 18: kono ushiro ni [tootta toki ni:,
‘while someone else is passing through behind him,’
((Akagi moves his pen to represent the movement of a person))

Yamada: 19: [sou desu ne ‘Right.’

Yamada: 20: un ‘Yes.’

Akagi: 21: “haijakku” nannari ga:,
‘how often hijacks or some other problems’

Yamada: 22: un ‘Uh-huh.’

Akagi: 23: don dake oki chau n daro[u tte iu koto nan desu ne
‘will occur.’

Yamada: 24: [un ‘Uh-huh.’

Yamada: 25: un ‘Yes.’

Akagi: 26: sore wa tabun (0.3) nankai ka yatte mite:, aidyii ga dondake sen’i suru
‘Probably (0.3) we have to test the case several times and check
how easily the IDs are switched’

Yamada: 27: un[un un ‘Yeah yeah yeah.’

Akagi: 28: [jyaa (0.7) mou su-mou ¥soko wa nanka¥ minna de .hh sono ba de

Akagi: 29: ¥soudan suru sika nai¥ noka naa tte omotte iru n desu kedomo .hh
‘Then (0.7) I think we should discuss that at the site.’

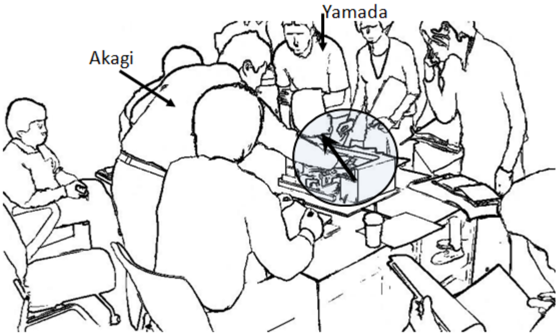


Fig. 3. Representation of “passing-through” 1 (line 18)

Yamada then summarizes that if hijacks do not occur, then they can fix the locations of the apparatuses, and if they do occur, then they will need to move “*Deai*” a little back and rearrange the locations of the other fixtures according to this change.

After that, Arisawa, a colleague of Akagi, begins participating actively in the discussion [Excerpt 2]. In replying to a request to confirm what Yamada said (lines 01, 02, 04, 08), Arisawa states his prediction that a situation in which a hijack really occurs will be one in which a visitor ignores another visitor standing and passes through behind him (lines 09, 11, 12). In saying so, he picks up Akagi’s plastic bottle of tea and puts it in front of him to resemble a standing person and slides the palm of his hand on the side of the bottle to draw the trajectory of the passing person (representation 2) [Fig. 4]. This representation is an elaboration of representation 1 in that it describes the same situation as that of representation 1 but expands the scale and makes more easily imaginable the distance between two persons and the quality of the movement, such as the speed and so forth. Creating this icon by using a plastic bottle, which does not normally represent a standing person, also structures the environment [5], which provides the configuration of semiotic resources for the following interaction (see also Barber’s discussion [1] about “preferential seeing”). Furthermore, similar to excerpt 1, this gesture accompanies an utterance that makes a negative prediction of the future. “Toki ni (if)” marks the subjunctive mood and “mushi shite (ignoring)” and “sudoori (pass though)” describe negative behaviors. In addition, “sudoori” does not merely describe the “passing-though” movement but also implies the quickness of the movement, and the accompanying gesture is expressed by a quick and strong movement (Sec. 5).

Excerpt 2: Representation 2

Yamada: 01: ma-eeto, yatte mite, mazu ano:sonnani hijack okon nai n dattara
‘If the test result does not show that hijack occurs so frequently,’
Yamada: 02: anshin site koko made iku to.
‘keeping this line wouldn’t be a problem.’
Akagi: 03: sou desu ne hai. ‘Right.’
((Arisawa and Akagi laughing))

Yamada: 04: de::: haijakku ga okoru you dattaraba, toriaezu sageru
'On the other hand if it shows that hijack occurs, we must move it backwards.'
05 (1.2)
Arisawa: 06: un::. 'Yes.'
07 (1.1)
Yamada: 08: tori[aezu sageru t[te iu:: 'Move it backwards.'
-> Arisawa: 09: [sore wa- [sore wa hontou ni
'Is that, Is that actually'
((Arisawa opens and pushes his right hand toward Yamada))
Yamada: 10: Un 'Uh-huh.'
-> Arisawa: 11: kou ita toki ni tatteru hito ni taisite,
'You mean the case in which a person stands like this and the other person'
((Arisawa holds the Akagis' plastic bottle and put it on the table))
-> Arisawa: 12: kono hito wo mushi shite sudoori suru keesu desu yo ne
'pass though behind him without paying any attention to him, right?'
((Arisawa represents the movement of a person who walk through by sliding his hand on the side of the plastic bottle quickly and strongly))
Representation of "passing-through" 2
Yamada: 13: sou desu sou desu sou desu. 'Yes that's right.'

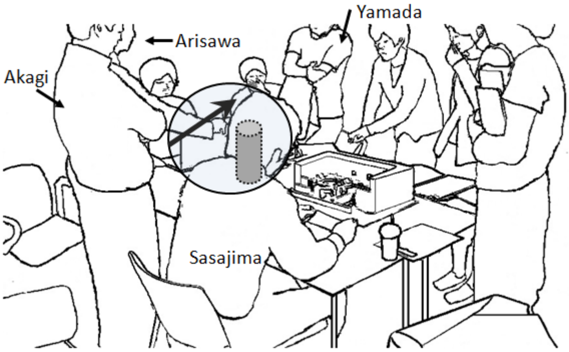


Fig. 4. Representation of "passing-through" 2 (line 12)

(The bottle is hidden behind Sasajima from this angle)

4.3 Using Representations in an "Examination" at the Construction Site

During the validation of the sensing system at the work site, which started at 13:30 on May 10 [Excerpt 3], participants are trying to find the distance between a person standing by the "*Deai*" terminal and one passing through behind him at the time when hijacks will occur. Okada, a colleague of Akagi and Arisawa and in charge of the sensing system, lets his colleague Suzuki stand by the model of the "*Deai*", which is made of cardboard with the same scale and shape of the real object to be created in the near future, and walks behind him. While doing so, Yamada is monitoring a display that shows the trajectory and the changing of personal ID numbers. The conditions of the trials are 3 by 3, namely, distance (30/40/50 cm) and speed (fast/normal/slow). When results of the trials for every condition are acquired, Yamada notes them on his memo pad in matrix format.

Excerpt 3: Representations 3 & 4

((Okada walks to the terminal "*Deai*" with Suzuki and has him stand in front of it. He then marks the floor with the packing tape.))
Okada: 01: yamada san ((Yamada looks at Okada)) sanjyuu senchi: (0.5)
Okada: 02: gurai hanareta tokoro wo (0.4) cyotto aruite mimasu
Declaration
'Mr.Yamada I'm going to walk through on the tape at 30 cm away from him.'
((Okada starts walking toward the entrance and Yamada turns and looks at the screen))
Yamada: 03: <ha>i 'Yes.'
04 (9.0)((Okada pass through between the corrugated cardboard made fixtures))
Trial=Representations of "passing-through" 3&4
Yamada: 05: a (1.3) ((Yamada looks at Okada)) kuttui<te>, i¥rekawari ma¥sita
'Oh, (1.3) the IDs were combined once and then switched.'
Report
Okada: 06: irekawari masita¿ 'Switched¿'
07 (0.4)((Yamada nods and looks at the screen))
Okada: 08: wakari masita (0.9) 'All right.'
Okada: 09: mou ikkai yatte mimasu ((Okada walks toward the entrance))
Repeat
'I will try (that) again.'
10 (11.4)((Okada pass through between the corrugated cardboard made fixtures))
Yamada: 11: kuttuite, irekawari masita
'They got combined and then switched.' (((Yamada looks at Okada))
12 (0.5)
Okada: 13: irekawari masita¿ 'Switched¿'
Yamada: 14: un (1.0) ¥irekawarimasune¥ 'Right. (1.0) IDs are switched.'
Okada: 15: irekawari masu yo ne. sanjyuu senchi de.
Evaluation
'Isn't it witched. At 30cm.'

In this validation, a typical action sequence was observed, which constitutes the "examination" phase in the higher level framework for problem-solving (sec. 2.3).

1. Declaration: Okada declares the conditions (distance and speed) which he will test from now on (lines 01 and 02).
2. Trial by division of labor: Okada walks according to the condition, and Yamada monitors (the changing of) IDs on the display (line 04), and they repeat this twice.
3. Report: Yamada announces the result and shares it with Okada (lines 05 - 08).
4. Evaluation: Stating the evaluation of the worth of the result (line 14-15).

What is to be noticed is that, the real-scale model of “*Deai*,” the person playing the role of the user of the terminal and the one passing through behind, even if they are located in the right location in the to-be exhibition space and are almost the same size as the intended referents, should not be the future object, namely the exhibition itself, but rather must be regarded as representations of them (representation 3)[Fig. 5: left], because the actual future objects do not exist yet, and, more importantly, because only some of their features are relevant for the current aim of the activity.

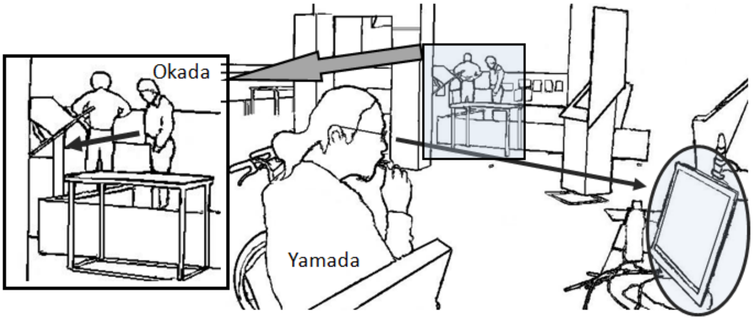


Fig. 5. Trial: Representations of “passing-through” 3 (left) and 4 (right)(line 04)

The focus of the attention of the participants in this validation was specified by a declaration before each trial into only some of the features of the trial, and the movement of “passing-through” was transformed into numerical data by the sensing system and shown on a two-dimensional display as another representation (representation 4)[Fig. 5: right]. It is no coincidence that what was focused on by declaration and what was shown on the display included almost identical information. Rather, it was a necessary consequence of the deliberate achievement of “seeing as a professional activity” [2]. Furthermore, because the display was installed on a desk distant from “*Deai*,” the person passing-through behind could not monitor the result in real time by himself, so it was necessary to divide labor with the other person.

Once the results on all conditions are acquired, Yamada stands up and walks to Okada with his memo pad and inscribes the results on the memo pad in a 3 by 3 matrix format (representation 5)[Fig. 6] while saying, ”good information.”

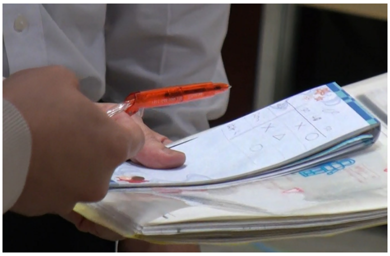


Fig. 6. Memo on validation results (representation of “passing-through” 5)

This representation is also a representation of the movements of "passing-through" several times, but it includes only information on the changing of IDs at each distance and speed, and excludes other details of bodily movements and spatial characteristics. This selection of what should be inscribed is based on and follows what are crucial in representation 4. In such a way, concrete representations are transformed continuously and result in a professional document [10].

4.4 "Resolution": Propagation Effects of the Result of the Examination

The arrangement of apparatuses beginning at 15:30 was led by Hiraizumi, another manager of Miraikan, who had floor staff stand by each apparatus and discuss their arrangement with Sasajima, the chief of exhibition production, and two members in charge of space design. After Sasajima reported the conclusion on the location of "*Deai*" (move it a little back), which resulted from the validation, they rearranged the locations of the apparatuses around the terminal. In the middle of the process, Yamada came and interrupted the team to explain the result to Hiraizumi while referring to his memo inscribed on the validation stage (representation 5) and pointing to "*Deai*" [Fig. 7]. Thus, more abstract representations such as the memo or other documents were easily available in different situations. Hiraizumi also pointed to the terminal and asked Sasajima whether the present location of it was based on the result of the validation. Thus, the result of the validation by one subgroup had a propagation effect on the arrangement of the other apparatuses by the other subgroup.

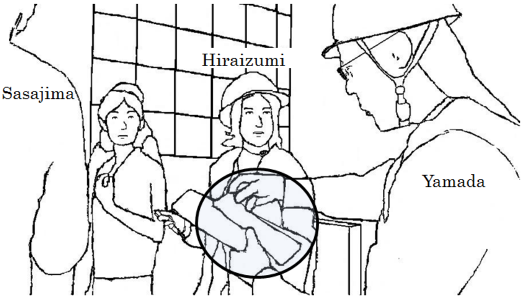


Fig. 7. Recycle of representation of "passing-through" 5

5 Discussion

In the previous section, we overviewed how various heterogeneous representations were used to present the problem of deciding the location of the "*Deai*" terminal in the preceding meetings and to resolve it on the construction site [Fig. 8].

This overall problem-solving activity by the group was achieved through the transformation of representations. In other words, the appropriateness of each representation should be assessed with reference to which phase it was used to achieve what purpose in the problem-solving process.

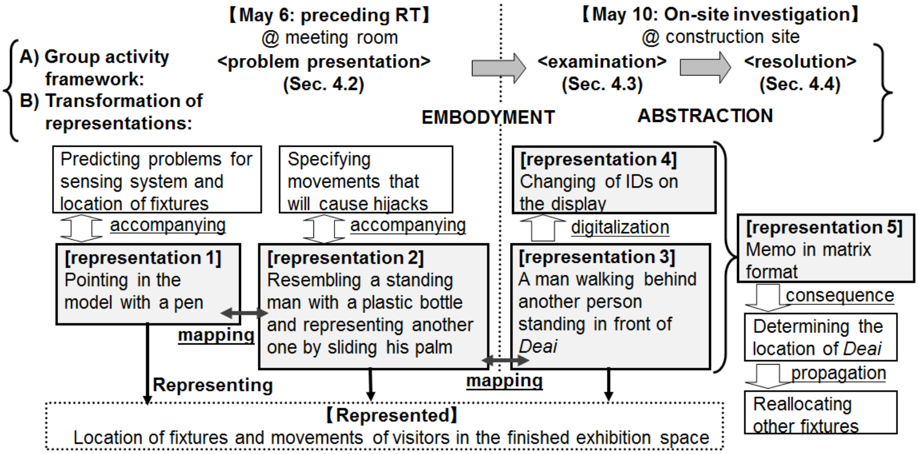


Fig. 8. Activity framework and transformation of representations

Representations 1 - 5 commonly represented the location of the "*Deai*" terminal and the movements of visitors around it but were different from each other in the phases in the problem-solving activity in which they occurred. Representations 1 and 2 were used in the "problem presentation" phase. Since the problem presentation occurred during the meeting in the meeting room, available resources for representing were limited. While it is common that some resources such as scale models and drawing sheets are referred to in order to discuss referents that do not yet exist in a meeting room distant from a to-be exhibition space, these representations omit several otherwise important features such as the size of the represented objects and the behaviors of people around them [9]. Therefore, they must be compensated for with talking and pointing or other kinds of gestures accompanying the talk [7]. In spite of the limitations of resources available in the meeting, the acts of presenting the problem by using these representations were situated in the interactive context of the meeting, which then facilitated the discovery and sharing of the problem among the participants. Had this sharing of the problem not occurred, examination by using representation 3 - 5 would not have been done, either.

Though representations 1 and 2 expressed very similar objects and events, there was a difference. While representation 1 was more panoramic and designated the location where the problematic behavior might happen, representation 2 was expressed in larger scale and specified the features of the behavior. It might be by virtue of this characteristic of representation 2 that "speed" was added to the conditions of validation in the examination phase as well as "distance."

In contrast to representations 1 and 2, representations 3 - 5 were observed at the construction site. Among them, representation 3 corresponded to "trials" as the core of the "examination" phase and therefore needed to be an embodied representation that resembled what should be represented as much as possible. Using cardboard models and human bodily movements of the same size and shape as those of the represented referents at a construction site was quite suitable for this purpose.

It is important to remember, however, that the appropriateness of even these representations should be evaluated in terms of the relevance to the task of determining the location of the terminal, and other details irrelevant to this task should be ignored. As the evidence, representation 3 was transformed immediately by the sensing system into representation 4, namely the trajectories and the changing of IDs on the display, and this information was further organized in terms of distance and speed, which are most relevant to the current problem-solving, and aggregated finally into the matrix on the memo pad (representation 5). Thus, through the abstraction from representation 3 to 5, the problem-solving activity progressed from the "examination" to the "resolution" phase. Due to the abstractness, representation 5 was in turn made available as the guide for the rearrangement of the apparatuses around the "Deai" terminal.

The process of transformation examined above might be similar to the "roof" gestures analyzed by LeBaron [8] in the sense that several representations of a single referent change through interaction between members. However, the foci of LeBaron's discussion are seen in a process in which gestures representing a "roof" become increasingly refined and "symbolic" and in the fact that a gesture that becomes a symbol is available in a subsequent meeting for another purpose. In contrast, what is characteristic in the case taken up in this article is that a series of the representations cannot necessarily be regarded as being refined on a single objective criterion but that the appropriateness of each of them should be judged according to which phase in the problem-solving activity they contribute to. Furthermore, it can be said that members flexibly choose the space itself for their activity with reference to the goals of the activities and the nature of the representations they will rely upon in the activities. In this way, each representation is embedded in both the interactional context and the material environment and serves as a bridge between them.

In order to consider adequately the nature of resemblance of representations of things that do not exist yet, it is necessary to make reference to the activity context of the group in which these representations are used. In addition, because such an activity context does not always get completed within a single meeting in a single space but sometimes stretches over a succession of meetings and other several interactions conducted in multiple spaces, organizational ethnographic approach, as well as microethnography [8], is indispensable to understanding such continuous group activities over a period of time.

When we develop a groupware that supports the activities of a group by virtue of information processing technology, it should be effective in order to enable the members of the group to share the records of their successive activities. However, visualizations of such records should not merely display events in chronological order but rather organize the structure of activities with reference to a proper activity framework such as "problem solving." Tracing the process of the transformation of representations is a way to uncover such activity frameworks, to which members of an interaction must orient themselves.

Acknowledgments. We would like to express our deep gratitude to The National Museum of Emerging Science and Innovation and the members of ANAGRA. We also wish to thank Tomoko Ikeda and Eiko Yasui for their exact comments on the draft. This research was supported by the "Development of a method of conversation analysis for understanding multiparty interaction," PRESTO, JST.

Transcription Symbols

(.)	brief pause (less than 200 milliseconds)
(numbers)	pause, represented in tenth of a second
[	point of overlap onset
.hh	audible in-breath
¥words¥	voice with inaudible laughter
(words)	unclear talk, with words in parenthesis represents candidate hearing
((words))	descriptions of events
.	falling intonation contour
,	low rising intonation contour
¿	middle rising intonation contour
?	high rising intonation contour
:	stretched voice
words	stressed talk
<words>	relatively slowed talk
>words<	relatively rapid talk

References

1. Barber, A.I.: Chaining Spaces. Journal of Linguistic Anthropology 15, 194–217 (2005)
2. Goodwin, C.: Professional Vision. American Anthropologists 96(3), 606–633 (1994)
3. Goodwin, C.: Action and Embodiment within Situated Human Interaction. Journal of Pragmatics 32, 1489–1522 (2000)
4. Goodwin, C.: Time in Action. Current Anthropology 43, S19–S35 (2002)
5. Goodwin, C.: The Body in Action. In: Justine Coupland, J., Gwyn, R. (eds.) Discourse, the Body and Identity, pp. 19–42. Palgrave/Macmillan, New York (2003)
6. Goodwin, C.: Pointing as Situated Practice. In: Kita, S. (ed.) Pointing: Where Language, Culture and Cognition Meet, pp. 217–241. Lawrence Erlbaum, Mahwah (2003)
7. Goodwin, C.: Participation, Stance, and Affect in the Organization of Activities. Discourse and Society 18(1), 53–73 (2007)
8. LeBaron, C.D.: Building Communication: Architectural Gestures and the Embodiment of New Ideas. The University of Texas at Austin (1998) (unpublished dissertation)
9. Murphy, K.M.: Collaborative Imagining: The Interactive Use of Gestures, Talk, and Graphic Representation in Architectural Practice. Semiotica 156, 113–145 (2005)
10. Smith, D.: Textually Mediated Social Organization. International Social Science Journal 36(1), 59–75 (1984)
11. `http://www.miraikan.jp/anagura/en/index.html`
12. Streeck, J., Goodwin, C., LeBaron, C.: Embodied Interaction in the Material World: an Introduction. In: Streeck, J., Goodwin, C., LeBaron, C. (eds.) Embodied Interaction, pp. 1–26. Cambridge University Press, Cambridge (2011)

Author Index