Ken Satoh
Akihiro Inokuchi
Katashi Nagao
Takahiro Kawamura (Eds.)

LNAI 4914

New Frontiers in Artificial Intelligence

JSAI 2007 Conference and Workshops
Miyazaki, Japan, June 2007
Revised Selected Papers

Springer

Lecture Notes in Artificial Intelligence 4914
Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Ken Satoh   Akihiro Inokuchi
Katashi Nagao   Takahiro Kawamura (Eds.)

New Frontiers in Artificial Intelligence

JSAI 2007 Conference and Workshops
Miyazaki, Japan, June 18-22, 2007
Revised Selected Papers

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

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

Akihiro Inokuchi
The Institute of Scientific and Industrial Research (ISIR)
Osaka University, Osaka, Japan
E-mail: inokuchi@ar.sanken.osaka-u.ac.jp

Katashi Nagao
Graduate School of Information Science
Nagoya University, Nagoya, Japan
E-mail: nagao@nuie.nagoya-u.ac.jp

Takahiro Kawamura
Research and Development Center
TOSHIBA Corporation, Kawasaki, Japan
E-mail: takahiro@eel.rdc.toshiba.co.jp

Library of Congress Control Number: 2008921032

CR Subject Classification (1998): I.2, H.2.8, H.3, F.1, H.4, H.5.2, I.5, J.1, J.3, K.4.3

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-78196-X Springer Berlin Heidelberg New York
ISBN-13 978-3-540-78196-7 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12228591 06/3180 5 4 3 2 1 0

Preface

The technology of artificial intelligence is increasing its importance thanks to the rapid growth of the Internet and computer technology. In Japan, the annual conference series of JSAI (The Japanese Society for Artificial Intelligence) has been playing a leading role in promoting AI research, and selected papers of the annual conferences have been published in the LNAI series since 2003.

This book consists of award papers from the 21st annual conference of JSAI (JSAI 2007) and selected papers from the four co-located workshops. Seven papers were awarded among more than 335 presentations in the conference and 24 papers were selected from a total of 48 presentations in the co-located workshops: Logic and Engineering of Natural Language Semantics 2007 (LENLS 2007), the International Workshop on Risk Informatics (RI 2007), the 5th Workshop on Learning with Logics and Logics for Learning (LLLL 2007), and the 1st International Workshop on Juris-informatics (JURISIN 2007).

The award papers from JSAI 2007 underwent a rigorous selection process. Firstly, recommendations were made from three people (Session Chair, session commentator and one PC member) in each session, and then recommended papers were carefully reviewed and voted for by PC members for final selection.

Papers from four workshops were selected by the workshop organizers via a selection process other than the workshop paper selection, and half of the workshop papers were selected on average. LENLS 2007 focused on dynamic semantics of natural language to shed new light on the incremental comprehension of sentences and discourse interacting with the context. The main themes of RI 2007 were data mining and statistical techniques to detect and analyze the risks potentially existing in the organizations and systems and to utilize the risk information for their better management. LLLL 2007 aimed at making techniques of machine learning and machine discovery more powerful and flexible by analyzing a role of computational logic and giving new insight into computational logic by analyzing activities of machine learning and machine discovery from the viewpoint of logic. The purpose of JURISIN 2007 was to discuss both the fundamental and practical issues for 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.

We believe that this book is a good introduction of the current status of AI research in Japan and contributes to advances in AI.

December 2007 Ken Satoh

Organization and Editorial Board

The award papers were selected by the Program Committee of the annual conference of JSAI (Japan Society for Artificial Intelligence) 2007. The paper selection of each co-located international workshop was made by the Program Committee of each workshop. Based on the decisions of the paper awards and the paper selections, each chapter was edited by the Program Chairs of the 21st annual conference of JSAI 2007 and the co-located international workshops. The entire contents and structure of the book were managed and edited by the chief editors.

Volume Editors

Primary Volume Editor: Ken Satoh (National Institute of Informatics, Japan)
Akihiro Inokuchi (Osaka University, Japan)
Katashi Nagao (Nagoya University, Japan)
Takahiro Kawamura (TOSHIBA Corporation, Japan)

Chapter Editors (Program Chairs)

Award Paper Chapter	Katashi Nagao (Nagoya University, Japan)
LENLS 2007 Chapter	Kei Yoshimoto (Tohoku University, Japan)
RI 2006 Chapter	Takashi Washio (Osaka University, Japan)
	Shusaku Tsumoto (Shimane University, Japan)
LLLL 2006 Chapter	Akihiro Yamamoto (Kyoto University, Japan)
	Kouichi Hirata (Kyushu Institute of Technology, Japan)
JURISIN 2007 Chapter	Katsumi Nitta (Tokyo Institute of Technology, Japan)
	Ken Satoh (National Institute of Informatics, Japan)
	Satoshi Tojo (Japan Advanced Institute of Science and Technology, Japan)

Table of Contents

Part I Awarded Papers

Part II Logic and Engineering of Natural Language Semantics

Part III Risk Informatics

Part IV Learning with Logics and Logics for Learning

Part V Juris-Informatics

Part I

Awarded Papers

Overview of Awarded Papers: The 21st Annual Conference of JSAI

Katashi Nagao

Center for Information Media Studies, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan
nagao@nuie.nagoya-u.ac.jp

This chapter features seven awarded papers, selected from JSAI 2007 - the 21st annual conference of Japanese Society for Artificial Intelligence. These awarded papers are truly excellent, as they were chosen out of 335 papers, with the selection rate just about two per cent, and the selection involved approximately ninety reviewers. The selection rate was smaller than usual years that used to be three to four per cent.

Synopses of the seven papers follow, with short comments for the award.

Nakano et al. consider a probabilistic model of fine-grained timing dependencies among multimodal communication behaviors (direct manipulation and speech) between an animated agent and its user. There model quite accurately judges whether the user understand the agent’s utterances and predicts user’s successful direct manipulation. It is valuable to suggest that the model is useful in estimating user’s understanding and can be applied to determining the next action of an agent.

Uchihira et al. report a product-based service design methodology that consists of five steps beginning with the generation of service concepts and ending with the description of service business plans. They also developed visualization tools that can help stakeholders identify new opportunities and difficulties of the target product-based service. They applied their method to a pilot case study and illustrated its effectiveness. It is expectable to make their tools more applicable to real world problems.

Minematsu and Nishimura regard speech as timbre-based melody that focuses on holistic and speaker-invariant contrastive features embedded in an utterance and describe a novel model of speech recognition using the holistic sound pattern of an input word. Their model is based on the observation that infants acquire spoken language through hearing and imitating utterances mainly from their parents but never imitate their parents' voices as they are. It is expectable to open up a new world of speech understanding after precise investigation of the proposed model.

Murakami et al. propose metrics “unexpectedness” and “unexpectedness” for measuring the serendipity of recommendation lists produced by recommender systems. Their metrics is based on the idea that unexpectedness is the distance between the results produced by the method to be evaluated and those produced by a primitive prediction method. They also evaluated their method through the results obtained by three prediction methods in experimental studies on television program

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 3–4, 2008.

recommendations. This work will be extended to become a new model of human emotion and attention.

Nakajima et al. describe a time-variant beam forming method that can extract sound signals from moving sound sources. Their method precisely equalizes the amplitude and frequency distortions of moving sound sources so that these sound signals can be extracted. It is very useful for successful natural human-robot interaction in a real environment because a robot has to recognize various types of sounds and sound sources. Several difficulties still remain to apply this work real robot applications but it is one of sound methods.

Masuda et al. propose a video scene tagging system which uses online video annotations including user commentaries on video scenes and scene-quoted video blogs and a video scene retrieval system which uses tags automatically extracted from annotation contents. Moreover, they performed open experiments using the systems for real-world evaluations. Their proposal of the practical method and of the systems for video scene retrieval will be more useful in the near future.

Nishida et al. address a “spatio-temporal semantic mapping system,” which is a general framework for modeling human behavior in an everyday life environment. It consists of a wearable sensor for spatially and temporally measuring human behavior in an everyday setting together with a Bayesian network modeling method to acquire and retarget the gathered knowledge on human behavior. It is valuable that their proposed system enables us to quantitatively observe and record everyday life phenomena and thus acquire some kind of reusable knowledge from the large-scale sensory data.

On behalf of the JSAI 2007 program committee, I would like to thank all the chairpersons, commentators, discussants, and attentive audience who contributed to selecting these exciting papers, and of course the authors who contributed these papers.

JSAI 2007 Program Committee

Katashi Nagao, Chair
Makoto Yokoo, Vice-chair
Ryutaro Ichise, Akihiro Inokuchi, Atsushi Iwasaki, Yuiko Ohta,
Daisuke Katagami, Ken Kaneiwa, Takahiro Kawamura, Tatsuyuki Kawamura,
Hidekazu Kubota, Takanori Komatsu, Ken Satoh, Masahiro Terabe,
Yoshiyuki Nakamura, Hiromitsu Hattori, Masahiro Hamasaki, Ryouhei Fujimaki,
Kazunori Fujimoto, Shigeo Matsubara, Masaki Murata

Modeling Human-Agent Interaction Using Bayesian Network Technique

Yukiko Nakano[1], Kazuyoshi Murata[1], Mika Enomoto[1], Yoshiko Arimoto[2], Yasuhiro Asa[3], and Hirohiko Sagawa[3]

[1] Tokyo University of Agri-culture and Technology
2-24-16 Nakacho, Koganei-shi, Tokyo 184-8588, Japan
{nakano,kmurata,menomoto}@cc.tuat.ac.jp
[2] Tokyo University of Technology
1404-1 Katakura, Hachioji, Tokyo 192-0981, Japan
ar@mf.teu.ac.jp
[3] Central Research Laboratory, Hitachi, Ltd.
1-280, Higashi-koigakubo Kokub-unji-shi, Tokyo 185-8601, Japan
{yasuhiro.asa.mk,hirohiko.sagawa.cu}@hitachi.com

Abstract. Task manipulation is direct evidence of understanding, and speakers adjust their utterances that are in progress by monitoring listener's task manipulation. Aiming at developing animated agents that control multimodal instruction dialogues by monitoring users' task manipulation, this paper presents a probabilistic model of fine-grained timing dependencies among multimodal communication behaviors. Our preliminary evaluation demonstrated that our model quite accurately judges whether the user understand the agent's utterances and predicts user's successful mouse manipulation, suggesting that the model is useful in estimating user's understanding and can be applied to determining the next action of an agent.

1 Introduction

In application software, help menus assist the user when s/he does not understand how to use the software. Help functions are usually used in problematic situations, so their usefulness and comprehensibility are critical for overall evaluation of the software. More importantly, if the advice provided by the help function is not helpful, that may confuse the user. Therefore, help functions that give useful advice at the right time are desirable.

To solve this problem, as a design basis of conversation-based help system, this paper proposes a human-agent multimodal interaction model using the Bayesian Network technique. This model predicts (a) whether the instructor's current utterance will be successfully understood by the learner, and (b) whether the learner will successfully manipulate the object in the near future.

Clark and Schaefer [1] defined that the process of ensuring that the listener shares an understanding of what has been said is grounding. Thus, the first issue, (a), can be said as the judgment of grounding: the judgment whether the instructor's utterance will be successfully grounded or not. If the predictions by the Bayesian network are accurate enough, they can be used as constraints in determining agent actions. For

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 5–12, 2008.

example, if the current utterance will not be grounded, then the help agent must add more information.

2 Background

2.1 Monitoring Listener's Behaviors and Adjusting Utterances

Analyzing conversations where the speaker and the listener share a workspace, Clark and Krych [2] found that speakers dynamically adjust their utterances that are in progress according to the listener's feedback expressed in multimodal manners, such as spoken language, nonverbal behaviors (e.g. gestures and facial expressions), listener's task manipulation, and change of the task environment caused by the manipulation. In particular, monitoring a listener's task performance seems to be an effective way of organizing such multimodal conversations.

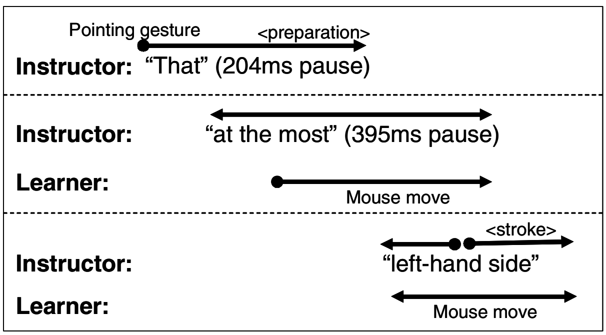


Fig. 1. Task manipulation dialogue

A software instruction dialogue in a video-mediated situation (originally in Japanese) is shown in Fig. 1. The speaker (instructor) is referring to the position of the "TV program searching button" by giving identification utterances in small pieces. Note that her gesture stroke follows the learner's mouse movements. This suggests that the speaker monitors the listener's mouse movement and adapts her verbal/nonverbal behaviors according to the listener's task manipulation. By virtue of such multimodal communication, the instructor smoothly coordinates the conversation even though there is no verbal response from the learner.

2.2 Nonverbal Information for Utterance Grounding

To accomplish such interaction between human users and animated help agents, predicting the task and conversation situation and modifying the content of instruction according to the prediction is necessary. For example, if the user does not seem to understand the agent's instruction, additional explanation is necessary. If the system predicts that the user fully understands the instruction and will conduct a proper operation in the near future, waiting for the user's operation without giving unnecessary annoying instruction would be better.

By applying a Bayesian Network technique to a dialogue-management mechanism, Paek and Horvitz [3] built a spoken-dialogue system, which takes account of the uncertainty of mutual understanding in spoken conversations. A similar technique was also applied to building user models in help systems [4]. However, there has been little study about timing dependencies among different types of behaviors in different modalities, such as speech, gestures, and mouse events, in predicting conversation status, and using such predictions as constraints in selecting the agent's next action.

Based on these discussions, this paper uses a probabilistic reasoning technique in modeling multimodal dialogues and evaluates how accurately the model can predict a user's task performance and judgment of utterance grounding [1].

3 Data Collection and Corpus

This section describes our corpus that is used for constructing a dialogue model. First, to collect dialogue data, we conducted an experiment using a Wizard-of-Oz method. In the experimental setting, an agent on an application window assists a user in operating a PC-TV application, a system for watching and recording TV programs on a PC.

3.1 Data Collection

A subject who joins the experiment as a user (hereafter, "user") and an instructor, who helps the user conducting the task and plays a role as a help agent, were in separate rooms. The equipment setting is shown in Fig. 2. The output of the PC operated by the user was displayed on a 23-inch monitor in front of the user and projected on a 120-inch big screen, in front of which the instructor was standing. The

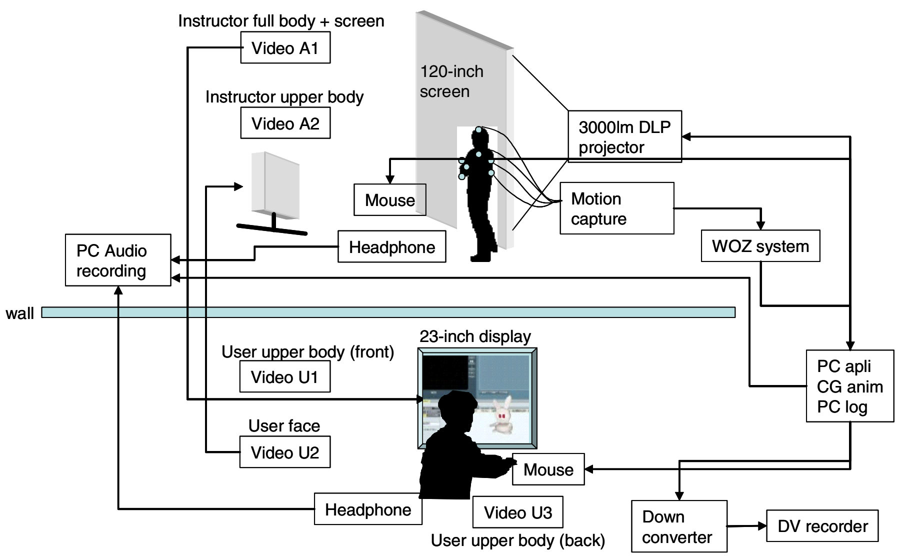
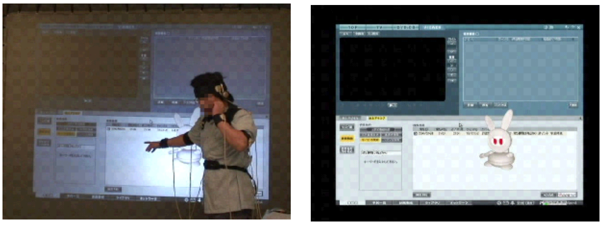


Fig. 2. Data Collection Environment



(a) Instructor (b) PC output

Fig. 3. Wizard-of-Oz agent controlled by instructor

instructor talked to the user while looking at the user's face (Video U2), which was monitored on a small display.

In addition, 10 motion sensors were attached to the instructor's body (Fig. 3 (a)) to capture the instructor's motion. The motion data was sent to the Wizard-of-Oz system and used to control a rabbit-like animated agent, which was overlaid on the user's PC-TV application. Thus, Fig. 3 (b) was displayed on the user's monitor as well as the big screen behind the instructor.

Both the user and the instructor wore headsets, and talked to each other through the headsets. The instructor's voice was changed through a voice transformation system, Herium, to make the voice sound artificial. Each participant's speech data was recorded by a headset microphone and saved in a Windows PC using a USB audio capture device. The audio was saved in the WAV audio format.

3.2 Task and Experimental Design

Each user was assigned one of two situations: recording a TV program or burning a DVD. The number of users was balanced between the situations. With the instructor's help, the user worked on two tasks for each situation.

Ten instructors and twenty users participated in the experiment. Each instructor had two sessions with two different users. Thus, we collected conversations from twenty pairs.

4 Corpus

The agent's (actually, instructor's) speech data was split by pauses longer than 200ms. We call each speech segment an inter-pausal unit (IPU), and use this as a unit of transcription. We assigned the following tags to 25 conversations using the Anvil video-annotating tool [5].

4.1 Utterance Content Tags

Focusing on the characteristics of the task, the utterance content of each IPU was categorized as follows.

- Identification (id): identification of a target object for the next operation
- Operation (op): request to execute a mouse click or a similar primitive action on the target

- Identification + operation (idop): identification and operation in one IPU
- State (st): referring to a state before/after an operation
- Function (fn): explaining a function of the system
- Goal (gl): utterance content for determining a purpose or a goal to be accomplished
- Acknowledgment (ack): agreement to or acknowledgement of the partner's speech

The inter-coder agreement for this coding scheme is very high, K = 0.89 (Cohen's Kappa), suggesting that the assigned tags are reliable.

4.2 Agent's Gestures and Motions

(1) Instructor's nonverbal behaviors
We annotated the shape and phase of the instructor's (Wizard-of-Oz agent's) gestures.
(1-1) Gesture Shape
- Pointing: pointing at one place on the monitor, or a hand motion that circles multiple objects.
- Trace: drawing a line to trace the words and phrases on the display.
- Other: other gestures

(1-2) Gesture Phase
- Preparation: arm movement from the beginning of a gesture to the stroke.
- Stroke: the peak of a gesture. The moment that a stroke is observed.
- Hold: holding a stroke hand shape, such as continuing to point at one position without moving a hand.
- Retract: retracting a hand from a stroke position.
- Partial retract: partially retracting an arm to go to the next stroke.
- Hesitate: a pause between a partial retract and a preparation or a pause between strokes.

(2) Agent Motions
We also annotated the positions and the gestures of the agent, which is actually controlled by the instructor.
(2-1) Agent movement: Duration of agent's position movement. If the agent does not move for longer than the time of 15 frames, that is counted as the end of the movement.
(2-2) Agent touching target as pointing (att): Duration of agent touching the target object as a stroke of a pointing gesture.

4.3 Mouse Operations

Using an automatic logging tool, we collected the following three kinds of log data as user's mouse operations.

- Mouse movement: movement of the mouse cursor
- Mouse-on-target: the mouse cursor is on the target object
- Click target: click on the target object

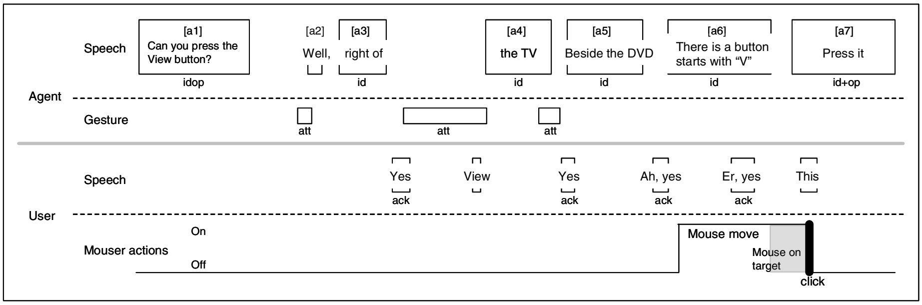


Fig. 4. Dialogue between Wizard-of-Oz agent and user

4.4 Corpus Data

An annotated corpus is shown in Fig. 4. The upper two tracks illustrate the agent's verbal and nonverbal behaviors, and the other two tracks illustrate the user's behaviors. At the first IPU, the instructor said, [a1] "Could you press the View Button?" The user did not respond to this instruction, so the instructor changed the explanation strategy: giving a sequence of identification descriptions [a2-5] by using short utterance fragments between pauses. Although the user returned acknowledgement, the user's mouse did not move at all. Thus, the instructor added another identification IPU [a6] accompanied by another pointing gesture. Immediately after that, the user's mouse cursor started moving towards the target object. After confirming that the user's mouse cursor reached the target object, the agent finally requested the user to click the object at [a7]. Note that the collected Wizard-of-Oz conversations are very similar to the human-human instruction dialogues shown in Fig. 1. While carefully monitoring the user's mouse actions, the Wizard-of-Oz agent adjusts the content of the instruction and its timing.

5 Dialogue Modeling Using Bayesian Network

In this section, a probabilistic dialogue model is constructed from the corpus data by using the Bayesian Network technique, which can infer the likelihood of the occurrence of a target event based on the dependencies among multiple kinds of evidence.

We extracted conversational data from the beginning of an instructor's identification utterance about a new target object to the point when the user clicks on the object. Each IPU was split at 500 ms intervals, and 1395 intervals were obtained. As shown in Fig. 5, the network consists of 9 properties concerning verbal and nonverbal behaviors for the past 1.5 seconds, current, and future interval(s).

As a preliminary evaluation, we tested how accurately our Bayesian network model can predict an instructor's grounding judgment and the user's mouse click. The following five kinds of information were given to the network as evidence. For the previous three intervals (1.5 sec), we used (1) the percentage of time the agent touched the target (att), (2) the number of the user's mouse movements. Evidence for

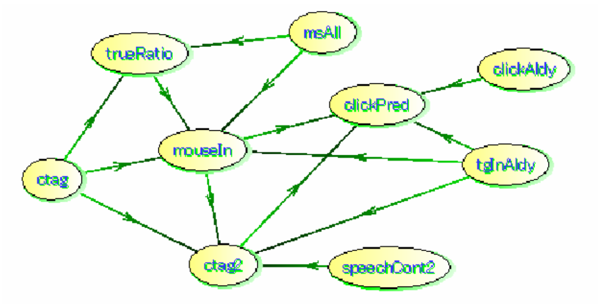


Fig. 5. Bayesian network model

the current interval is (3) content type of current IPU's, (4) whether the end of the current interval will be the end of the IPU (i.e., whether a pause will follow after the current interval), and (5) whether the mouse is on the target object.

5.1 Predicting Grounding Judgment

We tested how accurately the model can predict whether the instructor will go on to the next leg of the instruction or will give additional explanations using the same utterance content type (the current message will not be grounded).

The results of a 5-fold cross-validation are shown in Table 1. The prediction of "same content" is very accurate (F-measure is 0.90) because 83% of the data are "same content" cases. However, finding "content change" is not very easy because that occurs with less frequency (F-measure is 0.68). Testing the model using more balanced data would be better.

Table 1. Evaluation results

	Precision	Recall	F-measure
Content change	0.53	0.99	0.68
Same content	1.00	0.81	0.90

5.2 Predicting User's Mouse Clicks

As a measure of the smoothness of task manipulation, the network predicted whether the user's mouse click would be successfully performed within the next five intervals (2.5 sec). If a mouse click is predicted, the agent should just wait without annoying the user with an unnecessary explanation. Randomized data is not appropriate to test mouse click prediction, so we used 299 sequences of utterances that were not used for training. Our model predicted 84% of the user's mouse clicks: 80% of them were predicted 3-5 intervals before the actual occurrence of the mouse click, and 20% were predicted 1 interval before. However, the model frequently generates wrong predictions. Improving the precision rate is necessary.

6 Conclusion

Aiming at building a conversational help agent, first, this paper reported our experiment and verbal and nonverbal behavior annotation. Then, we proposed a probabilistic model for predicting grounding judgment and a user's successful mouse click. Adding more data, we will conduct more precise statistical analysis to demonstrate the co-constructive process of multimodal conversations. Moreover, our next step is to implement the proposed model in a conversational agent and evaluate the effectiveness of the proposed model.

References

1. Clark, H.H., Schaefer, E.F.: Contributing to discourse. Cognitive Science 13, 259–294 (1989)
2. Clark, H.H., Krych, M.A.: Speaking while monitoring addressees for understanding. Journal of Memory and Language 50(1), 62–81 (2004)
3. Paek, T., Horvitz, E.: Uncertainty, Utility, and Misunderstanding: A Decision-Theoretic Perspective on Grounding in Conversational Systems. In: Brennan, S.E., Giboin, A., Traum, D. (eds.) Working Papers of the AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems, American Association for Artificial Intelligence, Menlo Park, California, pp. 85–92 (1999)
4. Horvitz, E., et al.: The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. In: Proceedings of Fourteenth Conference on Uncertainty in Artificial Intelligence, pp. 256–265 (1998)
5. Kipp, M.: Gesture Generation by Imitation - From Human Behavior to Computer Character Animation, Boca Raton, Florida: Dissertation.com (2004)

Analysis and Design Methodology for Product-Based Services

Naoshi Uchihira[1], Yuji Kyoya[1], Sun K. Kim[2], Katsuhiro Maeda[1], Masanori Ozawa[1], and Kosuke Ishii[2]

[1] Toshiba Corporation, Kawasaki 212-8582, Japan
naoshi.uchihira@toshiba.co.jp
[2] Stanford University, CA 94305-4022, U.S.A.

Abstract. Recently, manufacturing companies have been moving into service businesses in addition to providing their own products. However, engineers in manufacturing companies do not find creating new service businesses easy because their work-related skills, understanding of design processes, and organizational skills have been developed and optimized for designing products and not services. To design product-based services more effectively and efficiently, systematic design methodologies suitable for engineers are necessary. We have designed a product-based service design methodology called DFACE-SI. This methodology consists of five steps beginning with the generation of service concepts and ending with the description of service business plans. Characteristic features of DFACE-SI include visualization tools that can help stakeholders identify new opportunities and difficulties of the target product-based service. We also applied DFACE-SI to a pilot case study and illustrated its effectiveness.

1 Introduction

In light of the ongoing transformation from a traditional industrial society to a knowledge-based society, many manufacturing companies have started introducing services along with their core products. Van Looy, Van Dierdonck, and Gemmel offered two reasons why companies are moving into the service industry [1]. One is to meet customer requirements in which the value-creating process for a product (e.g. driving) is more highly regarded than the product itself (e.g. a car). The other is to differentiate products by providing product-based services. Services can provide ongoing value and revenue through a life cycle of products (e.g. proactive maintenance with remote monitoring). However, when engineers, who are accustomed to designing products, try to design services, they encounter difficulties. As one might expect, these difficulties arise because the engineer's work-related skills and experiences, mental models, understanding of design processes, and organizational skills have been developed and optimized for designing products and not services. Engineers thus need systematic design methodologies for effectively and efficiently designing product-based services.

We have developed a service design methodology called "DFACE-SI", which is a specialized version of DFACE for Service Innovation. DFACE [2] is a design

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 13–25, 2008.

methodology based on Design for Six Sigma (DFSS) and widely used at Toshiba Corp. DFACE-SI consists of five steps starting with how to generate service concepts and ending with how to describe a service business plan. The main purpose of DFACE-SI is to make it easier for stakeholders to recognize potential opportunities and difficulties related to a target service. To achieve this, DFACE-SI provides charts and tools for identifying opportunities (concept generation) and difficulties (risk analysis). DFACE-SI also provides service function templates, service design patterns, and failure mode checklists that were designed during a service case analysis for use in designing and evaluating service concepts and schemes.

2 Opportunities and Difficulties of Service Businesses for Manufacturers

2.1 Opportunities

There are many opportunities for manufacturers to enter the service businesses [3]. However, how manufacturing companies can best exploit these opportunities is still unclear. If tried and true tactics for moving from product to service-based businesses were in place, then manufacturers could implement better strategies for becoming service businesses. We have been looking for commonly used transition patterns from the viewpoint of the customer contact point expansion based on 40 case studies of successful practices used by service businesses in manufacturing companies in Japan and the US[1]. A product-based service is here defined as a "value co-creation process of a product through collaboration between a customer and manufacturer" in which the customer-manufacturer contact point plays an important role. The original contact point is the buying and selling of products. Our model (the "Customer Contact Expansion Model") includes three types of expansion (adjustment, commitment, and territory expansion) from the original contact point of a product-based service (Fig. 1). These types of expansion can provide additional service values such as better product quality, customer reassurance, and product convenience.

Adjustment Expansion (better product quality): Modify products to meet the specific needs and usages of each customer in order to maximize the quality and functions of the products (e.g. maintenance, customizing, and consulting services).

Commitment Expansion (reassurance): Raise the level of commitment in managing customer risk (e.g. by providing rental, leasing, and outsourcing services).

Territory Expansion (convenience): Offer additional product functions that help the customer more effectively achieve objectives and provide a service platform for third parties to offer such functions (e.g. one-stop solutions, service platform providers).

[1] This case study project was led by the JAIST MOT team (leader: Prof. Akio Kameoka), which is supported by the New Energy and Industrial Technology Development Organization (NEDO) of Japan[4].

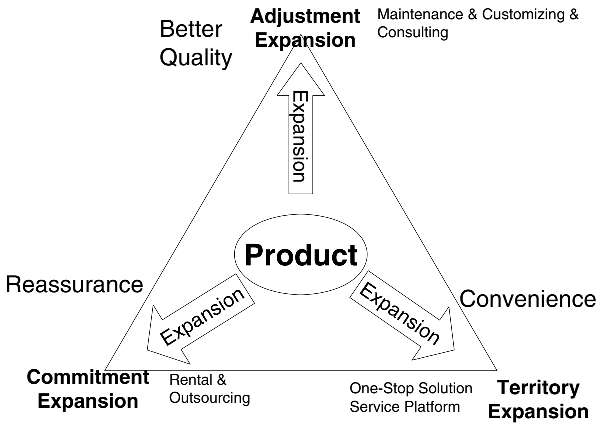


Fig. 1. Customer Contact Expansion Model

We have characterized the 40 successful practices using these three types of expansion. In the following sections, we separate these expansion types onto eight service function templates (Table 1) in order to provide a concrete means of identifying opportunities for product-based service businesses.

2.2 Difficulties

Even when opportunities are recognized, manufacturing companies cannot always easily and successfully manage the transition from a product- to service-based business. Calthrop and Baveja showed that only 21companies have successfully achieved their service strategies [6]. To find out why this percentage is so low, we analyzed difficulties in product- to service-based business transitions in the case studies. Figure 2 is a cause-effect diagram (fishbone diagram) showing a business model for identifying organizational problems. Whereas business model difficulties related to areas such as critical mass, social acceptability, and pricing are common to all service areas, organizational difficulties are specific to manufacturing companies. Since stakeholders need to identify these problems during the planning stage, DFACE-SI enables risk analysis through the use of Project Failure Mode and Effects Analysis (Project FMEA) and by providing failure mode checklists derived from the cause-effect diagram.

3 DFACE-SI

3.1 Basic Concept

The main purpose of DFACE-SI is to establish a shared recognition of opportunities and difficulties when starting a product-based service among stakeholders. Although there is no magic wand for starting a successful service business, we

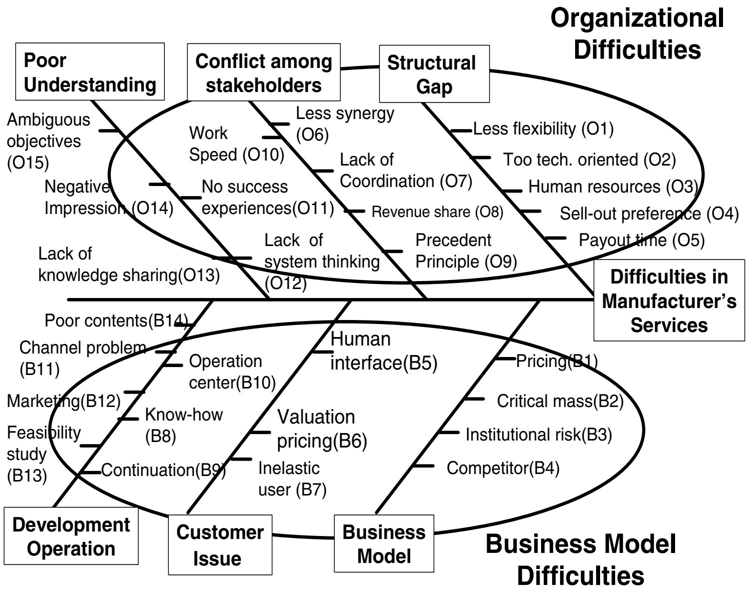


Fig. 2. Cause-Effect Diagram for Identifying Difficulties

think that shared recognition can be effective for successfully starting a service business. DFACE-SI thus provides design tools and design charts with service function templates, service design patterns, and failure mode checklists, which were developed during the case analyses. DFACE-SI has three main phases (Fig. 3), described as follows.

– Service concept generation using idea generation support tools (Customer Contact Expansion Model, Scenario Graph, and Value Graph) for identifying opportunities. Service function templates of the Customer Contact

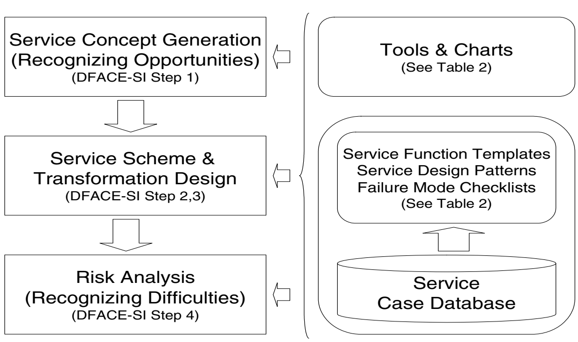


Fig. 3. Basic Concept of DFACE-SI

Expansion Model are used for generating service functions that configure the service concept.

– Service scheme and transformation design using design charts (Entity/Activity Chart, CVCA, and Scenario Chart) and service design patterns.
– Risk analysis using failure mode checklists and Project FMEA for identifying problems.

3.2 Procedure

An input of DFACE-SI is the target product and an output is the product-based service business plan. DFACE-SI consists of the following five steps (Fig. 4). An illustrative explanation of these five steps with examples is provided in Section 4.

Step1: Service concept generation
This step starts with the target product. Using Scenario Graph and Value Graph, the designer specifies 5W (Who, When, Where, Why, and What) features regarding a product-based service. The designer then finds 1H (How) features related to engineering metrics (EM) while using the Customer Contact Expansion Model, which recommends eight possible directions for service functions. Several suites of service functions form candidates of a final service concept. The final concept is selected using concept selection tools (Strategy Canvas and QFD).

Step2: Service scheme design
A service scheme shows a business model, procedure, and plans for turning a service concept into an actual service. The scheme can be described using several charts such as Entity/Activity Chart, CVCA, and Scenario Chart. During this step, the designer can describe the scheme by modifying service design patterns extracted from the case database by attribute pattern matching. Typical design patterns include maintenance services (adjustment expansion), rental services (commitment expansion), and content distribution and updating services (territory expansion). Most services consist of a combination of these patterns. Also during this step, a rough earnings model is calculated using commonly used methods such as the discounted cash flow method.

Step3: Transformation design
To implement a service scheme, the organization must be transformed from what it is (As-Is) to what it intends to be (To-Be). The designer must design organizational transformation scenarios. Examples of organizational transformation include establishment of a customer contact center, development of a service channel and an agency, and training of service managers and operators.

Step4: Risk analysis
Project FMEA is applied to the service scheme and organizational transformation scenarios. Here, failure modes can be determined using the failure mode checklists.

Step5: Evaluation and Refinement
The designer evaluates the rough earnings model and risks. If problems are found, the designer returns to previous steps in order to improve the service concepts,

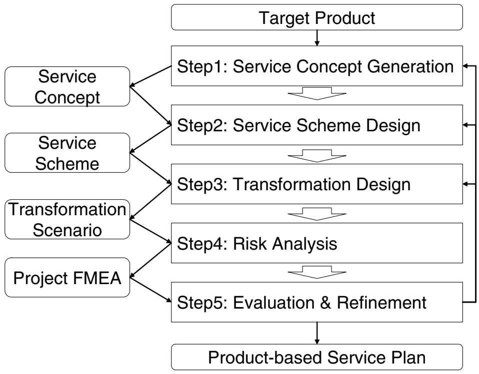


Fig. 4. Procedure of DFACE-SI

service schemes, and transformation scenarios. A final plan (a service concept, a service scheme, transformation scenarios, and Project FMEA) is then described with a detailed explanation, and then a decision-maker decides if the plan is a GO or a NOGO.

3.3 Tools

In the five steps of DFACE-SI, we use several tools and charts: Customer Contact Expansion Model, Scenario Graph[7], Value Graph, Entity/Activity Chart, CVCA, Scenario Chart, Project FMEA, service function templates, service design patterns, and failure mode checklists. Scenario Graph, Value Graph, CVCA, and Project FMEA are introduced at Stanford University Course ME317: Design for Manufacturability (DfM)[8]. In this section, we explain how the Customer Contact Expansion Model and related service function templates are used for generating concepts and how the Entity/Activity Chart is used for designing service schemes.These tools are unique to DFACE-SI and play important roles. Customer Contact Expansion Model provides 3 expansion axes. We can retract the following eight elemental service function templates from these axes (Table 1). The designer can create actual service functions (EM: engineering metrics) by associating What items (CR: customer requirements) of Value Graph with these eight elementary function templates as association keys (Fig. 5).

A service design scheme consists of a structure model (Entity/Activity Chart), a flow model (CVCA), and a behavior model (Scenario Chart). Entity/Activity Chart is an undirected bipartite graph. Entity nodes represent products, users, organizations, and information and activity nodes represent a function of the entities. An edge between an entity and an activity represents that the activity is executed with the linked entities. CVCA shows value flows (money, information, product, claim, etc.) among entities and Scenario Chart shows an execution sequence of activities (control flow). CVCA and Scenario Chart are derived from Entity/Activity Chart (Fig. 6).

Table 1. Eight Elementary Service Function Templates in Customer Contact Expansion Model

Expansion Type	Elementary Function	Explanation
Adjustment Expansion	Consulting	Consulting services to teach customers how they can make better use of the product
	Customizing	Customizing services to improve the product so that customers can make better use of it.
	Downtime and Risk Reduction	Maintenance services to reduce downtime and related risks by using monitoring information of the product.
Commitment Expansion	Financial Risk Reduction	Risk reduction services to take over financial risks (e.g. repair cost and investment risk) in place of customers.
	Social Risk Reduction	Risk reduction services to take over social risks (social responsibility) in place of customers.
	Operational Efficiency	Operation services to operate the product efficiently in place of customers.
Territory Expansion	Seamless Services	Related services necessary to solve customers' problems with the product, which are seamlessly provided.
	Rich Content	Content delivery and updating services by a platform connected to the products, where the content is processed in the product.

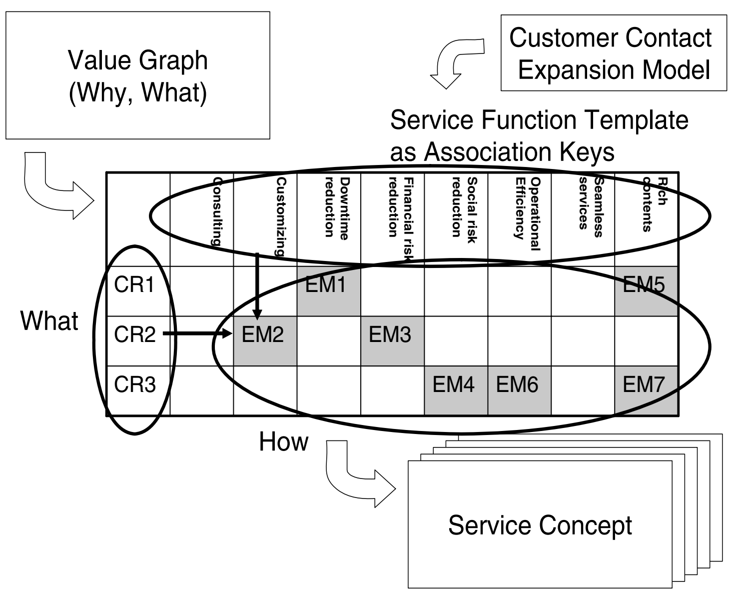


Fig. 5. Creating Service Functions Using Customer Contact Expansion Model

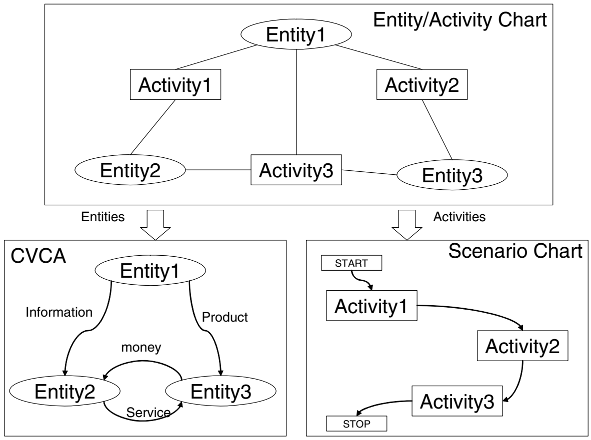


Fig. 6. Three Charts Representing Service Scheme

4 Service Business for Digital Video Camera (DVC)

Using the five DFACE-SI steps, we made a service business plan for a DVC business scenario. In this case study, the target product was a DVC, which has a hard disk and can be connected to a PC and the Internet.

Step1: Service Concept Generation

In this step, the designer considers several target service scenarios using the Scenario Graph, and after responding to three question types (Who, Where, and When) he/she then selects one of the scenarios. In the case of our scenario, a user takes pictures at sightseeing spots and enjoys these pictures at home (Fig. 7).

After considering the essence of the target service from two viewpoints (Why and What), the designer identifies key functions (How) to realize them using the Customer Contact Expansion Model. In this case, several service functions (EM: Engineering Metrics) are recognized in Fig. 8 and Table 2 (camera rental, camera shake adjustment, video editing software, etc.) that realize Customer Requirements (CR). For example, one function "camera shake adjustment" is created in an intersection of "recording beautiful scene" and "customizing photos" as shown in Table 2. In Table 2, row items are derived from CR in Value Graph, and column items are eight elementary function templates as association keys.

Finally, a final service concept (DVC rental service) is selected. This concept consists of the following three functions. (1) Lending a DVC to a user at sightseeing spots and housing the video taken by the user, (2) downloading the video at the user's PC and editing the video with additional location content (e.g. sightseeing spot information, background music), and (3) sharing the video (e.g. uploading it to blogs) and chatting about it with friends via the Internet (Fig. 9).

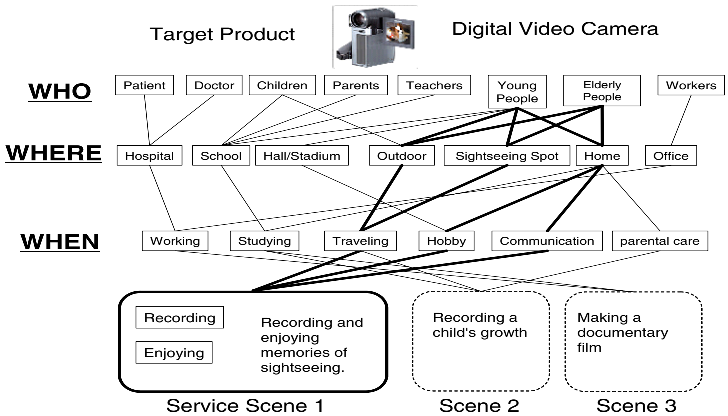


Fig. 7. Scenario Graph for Digital Video Camera

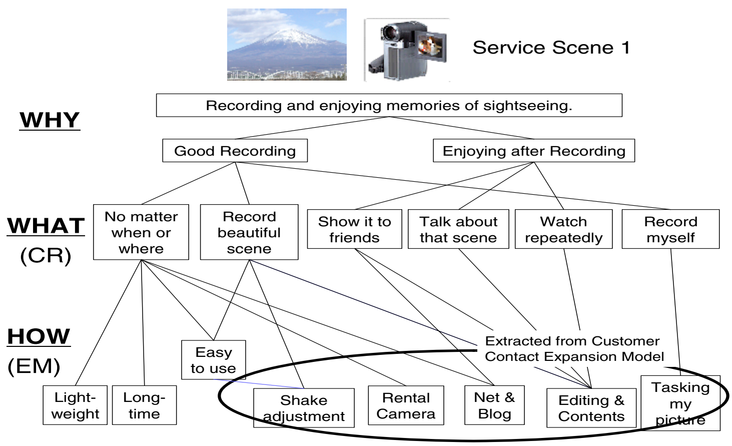


Fig. 8. Value Graph for DVC Rental

Step2: Service Scheme Design
In this step, the designer describes an Entity/Activity Chart representing a structure model of the service (Fig. 10). Here, the designer can describe the scheme by modifying two service design patterns (rental and content service types). A flow model (CVCA) is constructed by extracting entities from the Entity/Activity Chart (Fig. 11(a)), and a behavior model (Scenario Chart) is constructed by extracting activities from the Entity/Activity Chart (Fig. 11(b)).

Step3: Transformation Design
To produce a service scheme, the following organizational transformations from an As-Is organization to a To-Be organization are required.

Table 2. Customer Contact Expansion Model for DVC Rental

	Adjustment Expansion			Commitment Expansion			Territory Expansion	
	Customizing	Downtime Reduction	Consulting	Financial Risk Reduction	Social Risk Reduction	Operational Efficiency	Seamless Services	Rich Contents
No matter when or where	N/A	N/A	N/A	Camera rental	N/A	N/A	Download from net	N/A
Record beautiful scene	Camera shake adjustment	N/A	N/A	N/A	N/A	Supporting to edit	Supporting to edit	Supporting to edit
Show it to friends	N/A	N/A	N/A	N/A	N/A	N/A	Sharing with net	Sightseeing contents
Talk about that scene	N/A	N/A	N/A	N/A	N/A	N/A	Connecting to Blog	Sightseeing contents
Watch repeatedly	N/A	N/A	N/A	N/A	N/A	N/A	N/A	New additional contents
Record myself	N/A	N/A	N/A	N/A	N/A	Taking my picture	N/A	N/A

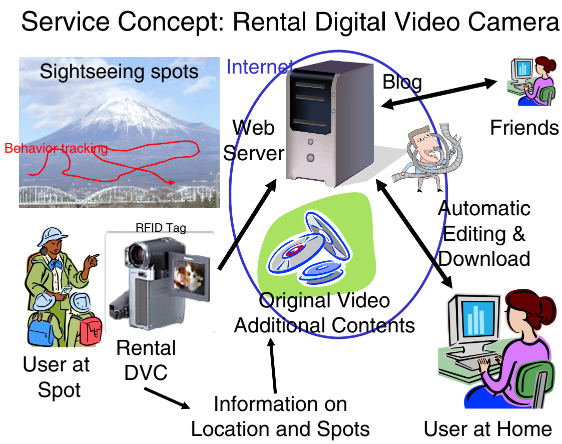


Fig. 9. Service Concept for DVC Rental

– Installation of operation center (content server center)
– Construction of DVC rental shop chain
– Service personnel training
– Content provider exploitation (tourist information and video decoration content)

The way to achieve these transformations is explained using a set of transformation scenarios.

Step4: Risk Analysis

Project FMEA is applied to a service scheme and organizational transformation scenarios (Fig. 12). Possible failure modes (B7, O5, B14, B11, O6, O8, B7, and B4) are derived from a failure mode checklist (Fig. 2).

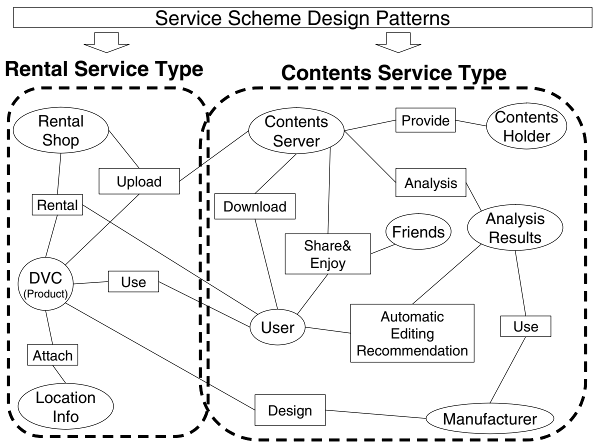


Fig. 10. Structure Model (DVC Rental)

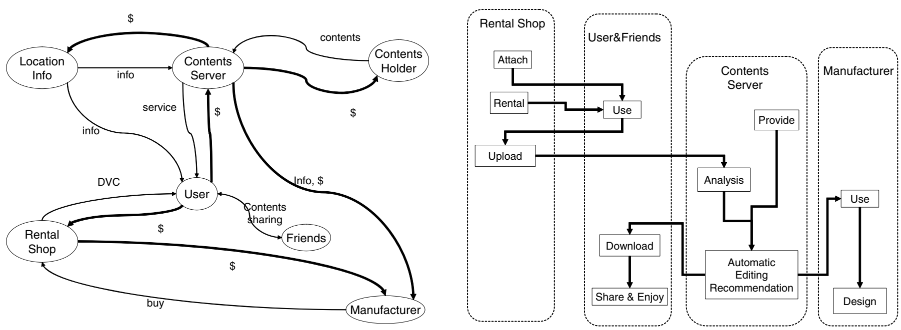


Fig. 11. Flow Model (a) and Behavior Model (b) (DVC Rental)

Step5: Evaluation and Refinement

The designer evaluates a rough earnings model and risks, and describes the final plan for a decision-maker. The final plan consists of a service concept, a service and an operation scheme, transformation scenarios, and Project FMEA.

5 Comparison with Related Works

Cooper & Edgett [9] and Edvardsson & Gustafsson [10] have presented new service development methodologies. These approaches are rather analytic since they provide no specific design charts or tools. Shostack [11] proposed useful and widely applicable service modeling tools (molecular modeling and a service blueprint). These tools were designed for general purposes in the service industry and do not take into account characteristic features of product-based services. Wise &

Failure Mode Checklist

Stage	ID	Failure Mode	Cause	Effect	Action	Rank
Development	B7	No service operation partner is obtained.	The forecast of a business model is too optimistic.	Withdrawal	Early withdrawal if no approval	2
	O5	Limited payout time	Unexpected large system investment	Replanning	Clear accountability	3
	B14	Poor Contents	Tourist agency's old attitudes	Less attractiveness	Substitute contents providers	4
Operation	B11	Limited to one sightseeing spot	Application is limited.	Restricted business scale	High margin model at specific spot.	3
	O6	Less Synergy with a DVC	Poor charm of DVC functions	Less DVC div. commitment	Strengthening synergy of products and services.	5
	O8	Revenue share with partner	DVC profit is larger than service revenue	Partner disaffection	Design win-win structure	3
	B7	User are inelastic.	Preferable to take a photo with ones own camera	Service business loses money	Enhance attractiveness of services.	4
	B4	Strong competitors (e.g. film company)	Market entry is easy.	Cost competition and less margin	Make a barrier by de facto standardization of a format.	5

Fig. 12. Project FMEA (DVC Rental) (part)

Baumgartner [3] developed four types of business models for product-based services. Their work was thought provoking but too analytical for most engineers. Shimomura et al. [12] developed the most advanced synthetic approach (Service CAD), which most engineers are now familiar with. Compared with Service CAD, DFACE-SI focuses on product-based services and provides a service concept generation tool that is based on the Customer Contact Expansion Model in which eight abstract functions are used as association keys for generating ideas. Unlike other methodologies, DFACE-SI also provides a risk analysis function.

Oliva & Kallenberg [13] developed a four-step transition process model. This transition process includes our three types of customer contact expansion. Furthermore, Oliva & Kallenberg mention problems related to organizational transformation of manufacturing companies. However, their model is somewhat too analytical for engineers who require a method that focuses more on procedural design. For these reasons, we believe that DFACE-SI can benefit engineers.

6 Conclusion

Recently, services sciences, management and engineering (SSME) has attracted interest not only in academia but also in manufacturing companies. However, there are few frameworks for analyzing and designing product-based service businesses [3][13]. We developed a service design methodology for product-based service businesses, called DFACE-SI, and tested it in a pilot case study. Our approach is unique because it provides a service concept generation tool based on the Customer Contact Expansion Model and because it provides a risk analysis feature (Project FMEA) that is particularly useful for organizational transformation. In our opinion, many Japanese manufacturing companies have been unsuccessful in entering the service businesses despite having good service concepts. We believe DFACE-SI will improve the success of such companies.

References

1. Van Looy, B., et al.: Services Management: An Integrated Approach. Financial Times Management (1998)
2. Ishiyama, K.: TOSHIBA Management Innovation Deployment (in Japanese). Journal of the Japanese Society for Quality Control 33(2), 154–160 (2003)
3. Wise, R., Baumgartner, P.: Go Downstream: The New Profit Imperative in Manufacturing. Harvard Business Review, 134–141 (September-October 1999)
4. Kameoka, A. (ed.): Research Report on Science and Technology for Service Innovation in Manufacturing Industry (in Japanese), JAIST Report (2006)
5. Uchihira, N., Koizumi, A.: Classification of Services in Manufacturing Industry and Strategy of Knowledge Utilization (in Japanese). In: 21st Annual Conference of The Japan Society for Science Policy and Research Management, 1B10, pp. 33–36 (2006)
6. Calthrop, P., Baveja, S.S.: From Products to Services: Why it is not so simple, CEOFORUM (2006), http://www.ceoforum.com.au/article-print.cfm?cid=7207
7. Kim, S.K., Ishii, K., Beiter, K.A.: Scenario Graph: Discovering New Business Opportunities and Failure Modes. In: ASME Int. Design Eng. Tech. Conf., ASME DETC2007-34967 (2007)
8. Ishii, K.: Course Materials, Design for Manufacturability (ME317), Stanford University, USA (2003)
9. Cooper, R.G., Edgett, S.J.: Product Development for the Service Sector, Basic Books (1999)
10. Edvardsson, B., et al.: New Service Development and Innovation in the New Economy, Studentlitteratur (2000)
11. Shostack, G.L.: How to Design a Service. European Journal of Marketing 16(1), 49–63 (1982)
12. Shimomura, Y., et al.: Proposalof the Service Engineering (1st Report, Service Modeling Technique for the Service Engineering) (in Japanese). Transactions of the Japan Society of Mechanical Engineers. C 71(702), 669–676 (2005)
13. Oliva, R., Kallenberg, R.: Managing the transition from products to services. International Journal of Service Industry Management 14(2), 160–172 (2003)

Consideration of Infants' Vocal Imitation Through Modeling Speech as Timbre-Based Melody

Nobuaki Minematsu[1] and Tazuko Nishimura[2]

[1] Graduate School of Engineering, The University of Tokyo,
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
mine@gavo.t.u-tokyo.ac.jp
[2] Graduate School of Medicine, The University of Tokyo,
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
nt-tazuko@ams.odn.ne.jp

Abstract. Infants acquire spoken language through hearing and imitating utterances mainly from their parents [1,2,3] but never imitate their parents' voices as they are. What in the voices do the infants imitate? Due to poor phonological awareness, it is difficult for them to decode an input utterance into a string of small linguistic units like phonemes [3,4,5,6], so it is also difficult for them to convert the individual units into sounds with their mouths. What then do infants acoustically imitate? Developmental psychology claims that they extract the holistic sound pattern of an input word, called *word Gestalt* [3,4,5], and reproduce it with their mouths. We address the question "What is the acoustic definition of word Gestalt?" [7] It has to be speaker-invariant because infants extract the same word Gestalt for a particular input word irrespective of the person speaking that word to them. Here, we aim to answer the above question by regarding speech as timbre-based melody that focuses on holistic and speaker-invariant contrastive features embedded in an utterance.

1 Introduction

Many speech sounds are produced as standing waves in a vocal tube, and their acoustic characteristics mainly depend on the shape of the tube. No two speakers have the same tube, and speech acoustics vary by speaker. Different shapes cause different resonances, which cause different timbre[1]. Similarly, different vowels are produced in a vocal tube by changing the tube's shape. Acoustically speaking, both differences between speakers and differences between vowels arise from the same cause. Speech features can also be changed by other factors such as features of a microphone, acoustics of a room, transmission characteristics of a line, auditory characteristics of a hearer, etc.

Despite the large acoustic variability, humans can accurately perceive speech. How is this done? Despite the progress of speech science, the contrast between the

[1] In musicology, timbre of a sound is defined as its spectral envelope pattern.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 26–39, 2008.

variability of speech acoustics and the invariance of speech perception remains an unsolved problem [8]. Speech engineering has attempted to solve this problem by collecting large numbers of samples corresponding to individual linguistic categories, e.g. phonemes, and modeling them statistically. IBM announced that they had collected samples from 350,000 speakers to build a speech recognizer [9]. However, no child needs this many samples to be able to understand speech. Perhaps the majority of the speech a child hears is from its mother and father. After it begins to talk, about a half of the speech a child hears is its own speech.

Developmental psychology explains that infants acquire spoken language by imitating the utterances of their parents. It is a tenet of anthropology that this behavior is found in no primates other than humans [2]. Further, we can say with certainty that infants never imitate the voices of their parents and that this is a clear difference from the vocal imitation of myna birds who imitate many sounds (cars, doors, animals, etc) as well as human voices. Hearing an adept myna bird say something, one can identify its owner [10]. However, the vocalizations of a child offer no clue as to the identity of its parents. What in the parents' voices do infants imitate? Due to poor phonological awareness, it is difficult for them to decode an input utterance into a string of phonemes [3,4,5,6], so it is also difficult to convert the individual phonemes into sounds. What then do infants imitate acoustically? Developmental psychology claims that they extract the holistic sound pattern of an input word, called *word Gestalt* [3,4,5], and reproduce it with their mouths. What then is the acoustic definition of word Gestalt? It must be speaker-invariant because infants can extract the same Gestalt irrespective of the person talking to them. To the best of our knowledge, no researcher has yet succeeded in defining it [7].

We recently formulated the above problem as a mathematical one and found a holistic and speaker-invariant representation of speech [11,12]. Here, an utterance is regarded as timbre-based melody. We did an experiment to test this representation. Acoustic models built with samples from only eight speakers based on the representation showed a slightly better recognition rate than Hidden Markov Models (HMMs) built with 4,130 speakers [13,14]. When we formulated the variability-invariance problem, we referred heavily to old and new findings in studies of linguistics, anthropology, neuroscience, psychology, language disorder, and musicology. Based on these findings and our results, we believe that infants extract the holistic and speaker-invariant contrastive features of speech.

2 Absolute Sense and Relative Sense of Sounds

2.1 Perception of Different Sounds as Identical

Figure 1 shows two utterances of /aiueo/, produced by two speakers. The one on the left was generated by a 200-cm-tall speaker and the one on the right was generated by an 80-cm-tall speaker. Although there is a large acoustic difference between the utterances, it can easily be perceived that they carry the same linguistic content, that is, /aiueo/. How do we perceive this equivalence in different stimuli? Do we perceive the equivalence after converting the two utterances into

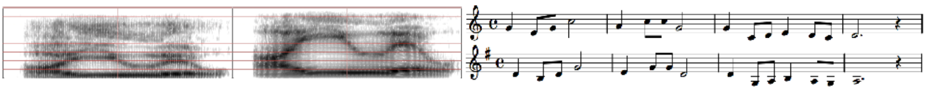


Fig. 1. Linguistic content of /aiueo/ uttered by two different speakers and the same piece of music played in two different keys

two phonemic sequences (sound symbol sequences) and finding the string-based equivalence between the two?

We think that the current speech recognition technology requires an answer of yes to the above question because the technology is based on a sound-to-symbol conversion technique that identifies separate sounds as single units among the linguistic sound categories, i.e. phonemes. However, this strategy dictates that acoustic models of the individual phonemes have to be made with samples from many speakers because a symbol corresponds to a variety of sounds.

Young children can also perceive the equivalence between the two utterances. Developmental psychology claims that infants do not have good phonemic awareness or a firm grasp of sound categories. This means that it is difficult for them to symbolize a separate sound and that invariance in perception is not due to string-based comparison. As explained in Section 1, infants first learn the holistic sound patterns in utterances and the individual phonemes later. This implies that invariance in perception must be based on comparison of holistic patterns. The question then is "What is the acoustic definition of the holistic and speaker-invariant pattern in speech?"

2.2 Relative Sense of Sounds in Music – Relative Pitch

Figure 1 also shows two pieces of the same melody performed in two different keys; C-major (top) and G-major (bottom). Although the acoustic substance of the two performances is very different, humans can easily perceive the equivalent musical content. When one asks a number of people to transcribe the two pieces as sequences of Do, Re, Mi, etc, three kinds of answers can be expected. Some will answer that the first one is So-Mi-So-Do La-Do-Do-So and the second one is Re-Ti-Re-So Mi-So-So-Re. These people are said to have absolute pitch (AP) and Do, Re, and Mi are pitch names for them, i.e. *fixed Do*. Others will claim that, for both pieces, they hear in their minds the same internal voices of So-Mi-So-Do La-Do-Do-So. They are said to have relative pitch (RP) and can verbalize a musical piece. For them, Do, Re, and Mi are syllable names, i.e. *movable Do*. The last group will not be able to transcribe the music, singing only "La-La-La-La La-La-La-La" for both. They also have RP but cannot verbalize a musical piece. They perceive the equivalence without sound-to-symbol conversion and only with a melody contour comparison. It should be noted that the RP people, the second and third groups, cannot identify a separate tone as one among the tonal categories.

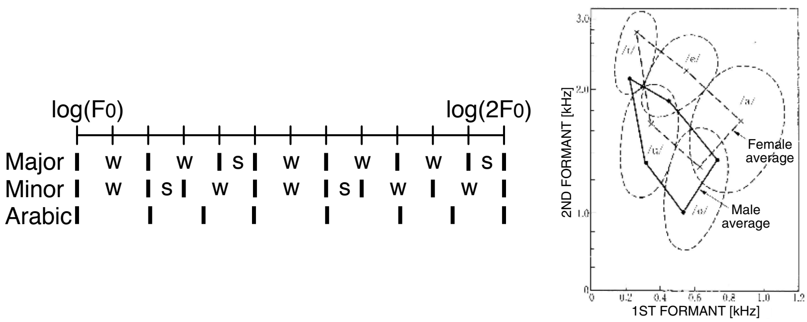


Fig. 2. Musical scales in octaves and a Japanese vowel chart using F_1 and F_2

AP people can memorize the absolute height of tones and use the heights to name musical tones. RP people capture the difference in the height between two tones (musical interval). If one explicitly defines the acoustic height of the Do sound, all the RP people, including the "La-La" people, can verbalize a given melody based on that definition. The difference between the second and third groups is that the former do not need a helper to define Do acoustically. How can they name the individual tones with no memory of the absolute height of tones? This ability is due to the tonal scale embedded in music, and, because this scale structure is key-invariant, the second group of people can easily identify the incoming tones independently of key.

Figure 2 shows three musical scales, all of which consist of octaves, eight tones in a group. The first two are well-known Western music scales, called major and minor. The third one is an Arabic music scale. For major and minor scales, an octave is divided into 12 semitone intervals, and eight tones are selected and arranged so that they have five whole tone intervals and two semitone ones. If C is used for tonic sound (the first sound) in a major scale, the key is called C-major and the tonal arrangement is invariant with key. The second group of people keeps the major and minor sound arrangements in memory and, based on these key-invariant arrangements, can identify the individual tones [15]. This is why they cannot symbolize a separate tone but can identify tones in a melody independently of key. Therefore, they find it difficult to transcribe a melody immediately after modulation in key. In contrast, people with AP can naturally transcribe the individual tones as pitch names even immediately after modulation. They sometimes do not notice the key change at all, but their identification is key-dependent, i.e., not robust at all.

RP people are capable of key-invariant and robust identification of tones because they dynamically capture the key-invariant sound arrangement [15]. In the following section, a similar mechanism is considered for speech perception.

2.3 Relative Sense of Sounds in Speech – Relative Timbre

A mother's voice is higher and a father's voice is lower because, in general, male vocal chords are heavier and longer [16]. A mother's voice is thinner and a father's

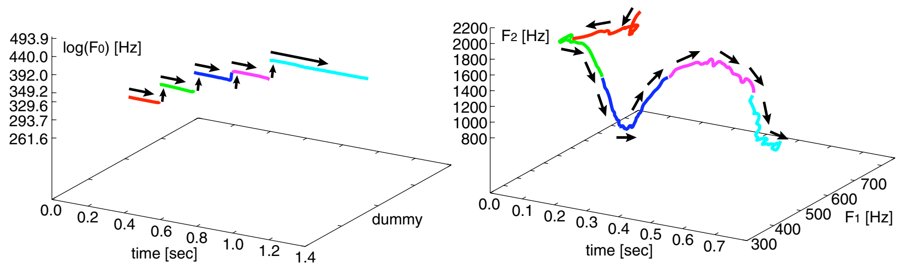


Fig. 3. Dynamic changes in pitch in CDEFG and changes in timbre in /aiueo/

voice is deeper because, in general, male vocal tracts are longer [16]. The former difference is one of pitch and the latter is one of timbre. The importance of RP is often discussed with regard to the former difference. People with strong AP tend to take a longer time to perceive the equivalence between a musical piece and a transposed version of it [17]. They often have to translate one symbol sequence consciously into another one to confirm the equivalence. Considering these facts, a similar mechanism, i.e. relative timbre, is hypothesized to explain why infants can perceive the equivalence between the two utterances in Figure 1 but cannot distinguish discrete symbols in the utterances.

As far as we know, however, all the discussions of sound identification in speech science and engineering have been based on absolute identification. How is it possible to discuss the relative identification of speech sounds based on invariant sound structure? Music consists of dynamic changes in pitch. Similarly, speech consists of dynamic changes in timbre. In Figure 3, a sound sequence of CDEFG played on a piano and a speech sound sequence of /aiueo/ are shown. Dynamic changes are visualized in a phase space. Pitch is a one-dimensional feature of F_0 and timbre is tentatively shown as a two-dimensional feature of F_1 and F_2. Cepstrum coefficients can also be used to expand the timbre space. Tonal transposition of a melody translates the dynamic changes in $F0$ but the shape of the dynamics is not altered. If the non-linguistic factors of speech such as speaker, microphone, etc, do not change the shape of the speech dynamics, the relative identification of speech sounds can be implemented on machines.

3 Robust and Structural Invariance Embedded in Speech

3.1 Mathematical Derivation of the Invariant Structure

As shown in the Japanese vowel chart in Figure 2, the male vowel structure can be translated into the female vowel structure. If the translation is accurate enough, the timbre dynamics can be easily formulated as invariant because differences in speakers do not change the sound arrangement and only transpose it multidimensionally. However, every speech engineer knows that this idea is too simple to be effectively applied to real speech data.

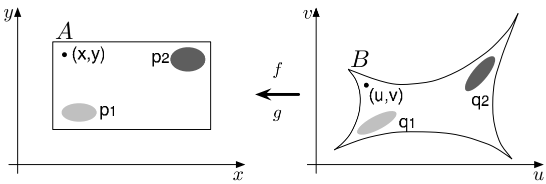


Fig. 4. Linear or non-linear mapping between two spaces

What kind of function can map the acoustic space of speaker A into that of speaker B: linear or non-linear? This question has been frequently raised in speaker adaptation research on speech recognition and speaker conversion research on speech synthesis. Figure 4 shows two acoustic spaces, one each for speakers A and B. Acoustic events p_1 and p_2 of A are transformed to q_1 and q_2 of B, respectively. If the two spaces have a one-to-one correspondence and point (x, y) in A is uniquely mapped to (u, v) in B and vice versa, transform-invariant features can be derived [18,13]. Every event is characterized as distribution:

$$1.0 = \oiint p_i(x, y)dxdy, \quad 1.0 = \oiint q_i(u, v)dudv. \tag{1}$$

We assume that $x{=}f(u, v)$ and $y{=}g(u, v)$, where f and g can be non-linear. Any integral operation in space A can be rewritten as its counterpart in B.

$$\iint \phi(x, y)dxdy = \iint \phi(f(u, v), g(u, v))|J(u, v)|dudv \tag{2}$$

$$= \iint \psi(u, v)dudv, \tag{3}$$

where $\psi(u, v) = \phi(f(u, v), g(u, v))|J(u, v)|$. $J(u, v)$ is Jacobian. Then, we get

$$q_i(u, v) = p_i(f(u, v), g(u, v))|J(u, v)|. \tag{4}$$

Physical properties of p_i are different from those of q_i. p_1 may represent /a/ of speaker A and q_1 may represent /a/ of B. We can show that the Bhattacharyya distance (BD) between two distributions is invariant with any kind of f or g.

$$BD(p_1, p_2) = -\log \oiint \sqrt{p_1(x, y)p_2(x, y)}dxdy \tag{5}$$

$$= -\log \oiint \sqrt{p_1(f(u, v), g(u, v))|J| \cdot p_2(f(u, v), g(u, v))|J|}dudv \tag{6}$$

$$= -\log \oiint \sqrt{q_1(u, v)q_2(u, v)}dudv \tag{7}$$

$$= BD(q_1, q_2) \tag{8}$$

The BD between two events in space A and that between their corresponding two events in space B cannot be changed. Substances can change easily, but their

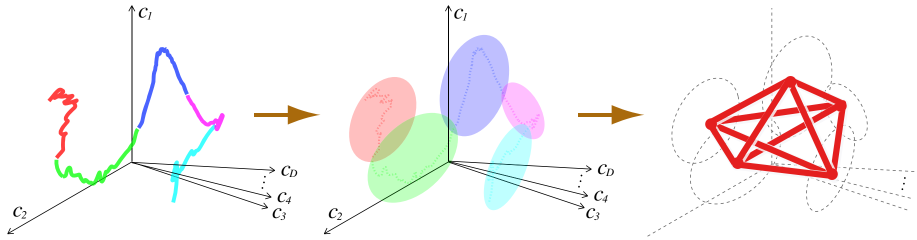


Fig. 5. Speaker-invariant structure in a cepstrum space without time axis

contrasts cannot be changed by any static transformation. The invariance is also satisfied with other distance measures such as the Kullback-Leibler distance. In this study, after some preliminary experiments, we adopted the BD.

The shape of a triangle is uniquely determined by fixing the lengths of all three sides. Similarly, the shape of an n point geometrical structure is uniquely determined if the lengths of all the ${}_nC_2$ segments, including the diagonal ones, are given. In other words, if a distance matrix is given for n points, the matrix completely determines the shape of the n-point structure. As stated above, the BD is robustly transform-invariant. When n distributions are given, their BD-based distance matrix represents its robustly-invariant structure. An invariant structure can be extracted from an utterance. Figure 5 shows this procedure in a cepstrum space. After converting an utterance into a sequence of distributions, all the BD-based timbre contrasts between any two distributions are calculated.

3.2 Discussions of Structural Representation

Figure 6 shows Jakobson's geometrical system of French vowels [19]. He claimed that the same vowel system could be found irrespective of the speaker. It is well-known that Jakobson was inspired by the assertions of Saussure, the father of modern linguistics, who claimed that language is a system of conceptual and phonic differences and that the important thing in a word is not the sound alone but the phonic differences that make it possible to distinguish that word from all others [20]. The proposed invariant, holistic, and contrastive representation of an utterance can be regarded as a mathematical and physical interpretation of Saussure's claims and Jakobson's claims [11,12]. We discard sound substances and extract only phonic contrasts from an utterance because the former are very fragile and the latter are robustly invariant.

If Western music is played with the Arabic scale shown in Figure 2, it will take on a different color, i.e. Arabic accented Western music. This is also the case with speech. If the vowel arrangement of an utterance is changed, it will be a regionally accented pronunciation. Figure 6 also shows the vowel structures of two accented pronunciations of American English, plotted after vocal tract length normalization [21]. The vowel arrangement can change the color of pronunciation. There is good functional similarity between the sound structure in musical tones and that in vowel sounds. The difference may be just observations.

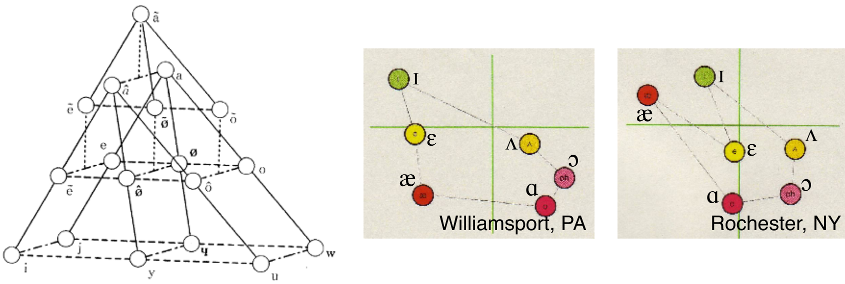


Fig. 6. Jakobson's geometrical structure of French vowels [19] and two accented pronunciations of American English [21]

Figure 3 shows a dynamic pattern or trajectory of pitch and of timbre. The pitch (melody) contour is often defined as a sequence of local pitch movements (ΔF_0), that is key-invariant. Similarly, the timbre contour can be defined as a sequence of local timbre movements (Δcepstrum), that is strongly speaker-dependent. It was mathematically and experimentally shown that vocal tract length differences change and rotate the direction of the timbre contour [22]. For example, with a frequency warping technique, a speech sample uttered by a male adult can be modified to sound like that of a boy. The warping operation shifts formant frequencies higher, that is, it makes them sound like the speaker is shorter. The direction of the Δcepstrum of a frame in the original speech and that of the corresponding frame in the modified speech was calculated. It was found that the timbre direction of a 170-cm-tall speaker and that of a 100-cm-tall speaker were approximately orthogonal. The directional difference became larger as the speaker's simulated height was increased or decreased. Further, the rotation of the timbre contour was not dependent on phonemes [22]. These results clearly indicate that the direction of local timbre dynamics is strongly dependent on the speaker. This is why, as shown in Figure 5, the directional components of the local dynamics are discarded and only the contrasts are extracted as scalar quantities. It should be noted that the proposed framework captures both local timbre contrasts and temporally distant contrasts.

4 Investigation Using Speech Recognition Research

4.1 Structural Acoustic Matching between Two Utterances

When two utterances are represented as different structures, how is the matching score between the two utterances to be calculated? As shown above, no transform can change the structure, which means that any transform can be interpreted as one of two geometrical operations, rotation or shift. As shown in Figure 7, the matching score for the two utterances is calculated as the minimum of the total distance between the corresponding two distributions (points in the figure) after shift and rotation. It was shown experimentally in [23] that minimum distance

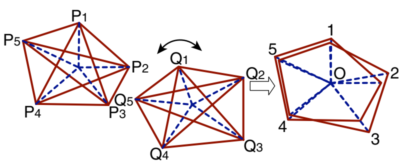


Fig. 7. Distance calculation

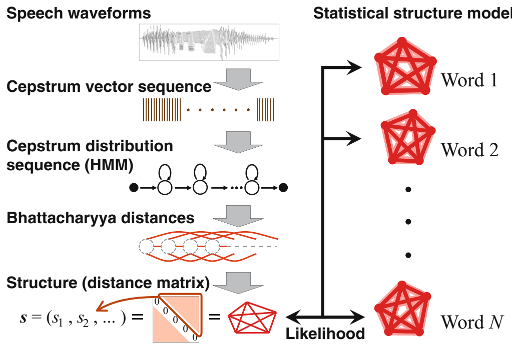


Fig. 8. Framework of structural recognition

D could be approximately calculated as a Euclidean distance between the two matrices, where the upper-triangle elements form the *structure vector*;

$$D(P, Q) = \sqrt{\frac{1}{n} \sum_{i<j} (P_{ij} - Q_{ij})^2}, \tag{9}$$

where P_{ij} is an (i, j) element of P and n is the number of distributions.

4.2 Verification of Structural Speech Recognition

To investigate the fundamental characteristics of the proposed framework, we examined automatic recognition of isolated words [13,14]. Here, a word was defined artificially as a connected vowel utterance. Since Japanese has the five vowels, /aiueo/, V_1-V_2-V_3-V_4-V_5 ($V_i \neq V_j$) utterances like /eoaui/ were used as words. The vocabulary size, i.e. perplexity, is ${}_5P_5$ (=120).

Eight male and eight female speakers recorded five utterances for each of the 120 words for a total of 9,600 utterances. Half of the samples from four males and four females were used for training and the others for testing. Since the proposed framework can eliminate differences between speakers well, only eight speakers were used for training. The recognition framework is shown in Figure 8. Vector sequences were converted into distribution sequences as MAP-based HMM training. Word templates were stored as statistical models averaged over structure vectors of each word's utterances.

As a comparison, an isolated word recognizer using tied-mixture triphone HMMs trained with 4,130 speakers [24] was built. Mel-frequencey cepstrum coefficients (MFCC) and its Δ were used with cepstral mean normalization (CMN). The word-based and vowel-based recognition rates are shown in Table 1. Although the proposed method completely discarded speech substances, it performed almost as well as the HMMs, which used both static and dynamic features. It should be noted that direct comparison is not fair because, for HMMs, the experiment was task-open, but, for the proposed framework, it was task-closed. In spite of this, we

Table 1. Recognition rates of the two methods

	HMM	Proposed
#speakers	4,130	8
word-based	97.4	98.3
vowel-based	98.8	99.3

can say that the proposed holistic and structural representation of speech identifies words well. Detailed descriptions of the recognition experiments are found in [13,14,25].

5 Investigation Using Speech Perception Research

5.1 Perception of Speaker-Variable and Size-Variable Speech

The RP people who can verbalize a given melody as a syllable name sequence have troubles transcribing it for some time immediately after the key has been modulated. We showed experimentally that this was also the case with speech [11]. Speaker-variable word utterances were generated with speech synthesis techniques and presented to human subjects. The speaker-variable utterances are those whose speaker information changes along the time axis. For example, the speaker is changed mora by mora or phoneme by phoneme. The presented stimuli were meaningless words. It was found that changing the speaker significantly degraded the transcription performance. Timbre changes due to speaker changes tended to be perceived as phoneme changes. However, the performance of a speech recognizer for the same stimuli was not degraded because the recognizer used speaker-independent HMMs trained with 4,130 speakers. A similar finding was obtained in another study [26] where size-variable speech samples were used.

In music, "La-La" people can enjoy music and can perceive the equivalence between a musical piece and a transposed version of it without symbolizing tones. We built a "La-La" machine, for which the two utterances in Figure 1 were completely identical but for which speech sound symbolization was impossible. We thought that this machine was a good simulator of infants' abilities and wondered whether this holistic speech processing could be found in adults.

5.2 Holistic Speech Processing Found in Adult Listeners

We found the answer in previous studies [27,28] done by another research group. Figure 2 shows a Japanese vowel chart. If vowel sounds of people the size of giants and those of people the size of fairies are obtained, they have to be plotted outside the ranges of existing people because formant frequencies depend on vocal tract length. Can subjects identify these vowel sounds in isolation? If they have difficulty with separate vowels, can they identify a continuous utterance more easily? The speaker-invariant holistic patterns are also embedded even in utterances of giants and fairies.

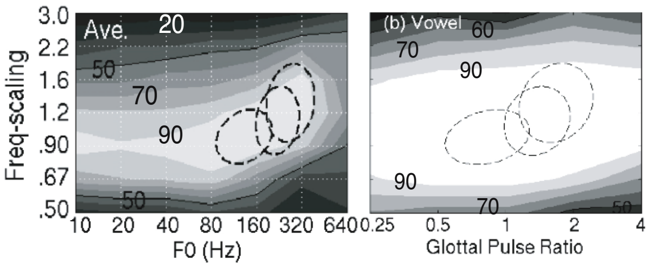


Fig. 9. Vowel identification rates without/with acoustic context [28]

Figure 9 (left) shows the identification rates of isolated Japanese vowels generated with various F_0s and body sizes. X and Y indicate F_0 and the ratio of frequency scaling in the spectrum, respectively. In the figure on the left, a vowel sample at (x, y) means that F_0 is x Hz and, roughly speaking, body height is $170/y$ cm ($5.6/y$ ft.). Three circles are the ranges of F_0 and body height for adult males, adult females, and children. Within these ranges, the identification rates are better than 90%. It is very difficult to absolutely identify separate vowels uttered by giants and fairies. For 65-cm-tall fairies (2.1 ft.) with an F_0 of 160 Hz, the performance is chance level (20%). Figure 9 (right) shows the identification rates of vowels in four-mora *unknown* words. Here, a vowel at (x, y) has an F_0 of $160x$ Hz. When giants and fairies say something in connected speech, subjects are reasonably able to identify the individual sounds even though the utterance is *meaningless*. For 65-cm-tall fairies, people can identify the sounds with about 60% accuracy. If known words are presented, performance will improve. With familiar words, it will improve drastically.

For "La-La" people, the request, "Remember the third tone in the next melody. Listen to another melody and raise your hand if you hear the same tone." is very difficult to execute. Unless tones are symbolized, people will experience difficulty. Some people may have similar difficulty with the request, "Remember the third sound in the next utterance. Listen to another utterance and raise your hand if you hear the same sound." Unless sounds are symbolized, people will have difficulty. To the best of our knowledge, two types of people have this kind of difficulty: young children and dyslexics. Their phonemic awareness is very weak [4]. Dyslexics are said to be good at seeing a whole tree but bad at seeing individual leaves [4,29]. However, both groups enjoy speech communication and can easily perceive the equivalence between the two utterances in Figure 1.

5.3 Non-robust Processing with Only Absolute Sense of Sounds

People with strong AP have some difficulty perceiving the equivalence between a musical piece and a transposed version of it [17]. Musicians with strong AP whose reference tone (A above middle C) is fixed at 440 Hz often have troubles performing music. An acoustic version of the reference tone depends on the orchestra playing it and it is sometimes 442 or 445 Hz. This small difference is difficult for AP musicians to accept. Absolute processing has to be non-robust.

If people have a strong absolute sense of speech sounds, they are likely to have difficulty perceiving the equivalence between the two utterances in Figure 1. Some autistics, who are considered to be good at seeing leaves but bad at seeing a whole tree [30,31], fall into this category. An autistic Japanese boy *wrote* that it was easy to understand his mother's speech but difficult to understand the speech of others [32]. However, it was also difficult for him to understand his mother's speech over the phone. He was able to write before he could speak, and spoken language was always difficult for him. Another autistic boy imitated voices as myna birds do, but spoken language was difficult also for him [34]. Autistics are much better at memorizing individual stimuli separately and absolutely as they are, but they are much worse at extracting holistic patterns hidden in the stimuli than non-autistic people [30,31]. It is explained in [30] that autistics lack the drive towards central coherence (Gestalt) and live in a fragmented world.

No child with normal hearing imitates voices, but myna birds and some autistics try to imitate voices as they are. Every speech synthesizer learns and imitates the voices of a single speaker. Every speech recognizer learns the voices of so many speakers by statistically modeling the acoustic features of the individual allophones separately and absolutely. However, the robustness in recognizing speech is far lower than that of humans. We cannot help considering the similarity between speech systems and autistics. In the 90's, AI researchers found that the robots sometimes behaved like autistics [33]. Both robots and autistics were bad at dealing with small environmental changes, known as the frame problem. AI researchers and therapists have recently been collaborating [33]. For them, making robots more suited to the real world and helping autistics become more accustomed to it are similar problems. Considering these facts, we think that speech engineers may have to face the same problem that AI researchers have.

6 Discussion and Conclusion

Anthropological studies showed that, basically speaking, no primates other than humans have relative pitch [35,36,37]. This is because relational processing requires a higher cognitive load than absolute processing. Therefore, other primates have difficulty perceiving the equivalence between the two musical pieces shown in Figure 1. Since pitch is one-dimensional and timbre is multi-dimensional, relative timbre processing should require an even higher load. What kind of behavior will be observed when the two utterances shown in Figure 1 are presented to chimpanzees? What if one of them is presented directly and the other is presented over the phone? If they cannot perceive the equivalence, human *spoken* language must also be very difficult. Researchers in the field of anthropology have made many attempts to teach human language to chimpanzees but most of them used visual tokens, not oral ones. Human vocal sounds failed to work as tokens even after being presented an enormous number of times [38]. However, young children can easily perceive the equivalence in different samples of speech.

We think that this invariant perception owes much to relative timbre, where an utterance is perceived as a timbre-based melody.

Finally, we want to carry out a thought experiment. Suppose that the parents of identical twins get divorced immediately after the twins are born and that one twin is taken in by the mother and the other is taken in by the father. What kind of pronunciation will the twins have acquired after ten years? The twins do not produce voices that sound, respectively, like the mother and like the father. However, there is an exceptional case in which the twins' pronunciations will be very different: when the parents are speakers of different regional accents. Timbre difference based on difference in speakers does not affect the pronunciation but that based on regional accents does. Why? The simplest explanation is that infants do not learn the sounds as they are but learn the sound system embedded in spoken language. The proposed representation extracts the invariant system embedded in an utterance. We believe that this is the answer to the question.

References

1. Kuhl, P.K., Meltzoff, A.N.: Infant vocalizations in response to speech: Vocal imitation and developmental change. J. Acoust. Soc. Am. 100(4), 2425–2438 (1996)
2. Gruhn, W.: The audio-vocal system in sound perception and learning of language and music. In: Proc. Int. Conf. on language and music as cognitive systems (2006)
3. Hayakawa, M.: Language acquisition and matherese. In: Language, Taishukan pub. vol. 35(9), pp. 62–67 (2006)
4. Shaywitz, S.E.: Overcoming dyslexia, Random House (2005)
5. Kato, M.: Phonological development and its disorders. J. Communication Disorders 20(2), 84–85 (2003)
6. Hara, K.: Phonological disorders and phonological awareness in children. J. Communication Disorders 20(2), 98–102 (2003)
7. Minematsu, N., Nishimura, T.: Universal and invariant representation of speech, CD-ROM of Int. Conf. Infant Study (2006), http://www.gavo.t.u-tokyo.ac.jp/~mine/paper/PDF/2006/ICIS_t2006-6.pdf
8. Johnson, K., Mullennix, J.W.: Talker variability in speech processing. Academic Press, London (1997)
9. http://tepia.or.jp/archive/12th/pdf/viavoice.pdf
10. Miyamoto, K.: Making voices and watching voices. Morikawa Pub. (1995)
11. Minematsu, N., et al.: Theorem of the invariant structure and its derivation of speech Gestalt. In: Proc. ISCA Int. Workshop on Speech Recognition and Intrinsic Variation, pp. 47–52 (2006)
12. Minematsu, N.: Are learners myna birds to the averaged distributions of native speaker? – a note of warning from a serious speech engineer –, CD-ROM of ISCA Int. Workshop on Speech and Language Technology in Education (2007)
13. Asakawa, S., Minematsu, N., Hirose, K.: Automatic recognition of connected vowels only using speaker-invariant representation of speech dynamics. In: Proc. InterSpeech, pp. 890–893 (2007)
14. Qiao, Y., Asakawa, S., Minematsu, N.: Random discriminant structure analysis for continous Japanese vowel recognition. In: Proc. Int. Workshop on Automatic Speech Recognition and Understanding, December 2007 (to appear)

15. Taniguchi, T.: Sounds become music in mind – Introduction to music psychology –. Kitaoji Pub. (2000)
16. Titze, I.R.: Principles of voice production. Prentice-Hall Inc., Englewood Cliffs (1994)
17. Miyazaki, K.: How well do we understand absolute pitch? J. Acoust. Soc. Jpn. 60(11), 682–688 (2004)
18. Minematsu, N., Asakawa, S., Hirose, K.: Linear and non-linear transformation invariant representation of information and its use for acoustic modeling of speech. In: Proc. Spring Meeting Acoust. Soc. Jpn., pp. 147–148 (2007)
19. Jakobson, R., Lotz, J.: Notes on the French phonemic pattern, Hunter (1949)
20. Saussure, F.: Cours de linguistique general. In: Publie par Charles Bally et Albert Schehaye avec la collaboration de Albert Riedlinge, Lausanne et Paris, Payot (1916)
21. Labov, W., Ash, W., Boberg, C.: Atlas of North American English. Walter de Gruyter, Berlin (2001)
22. Saito, D., et al.: Derectional dependency of cepstrum on vocal tract length. In: Proc. Int. Conf. Acoustics, Speech, and Signal Processing (2008, submitted)
23. Minematsu, N.: Yet another acoustic representation of speech. In: Proc. Int. Conf. Acoustics, Speech, and Signal Processing, pp. 585–588 (2004)
24. Kawahara, T., et al.: Recent progress of open-source LVCSR engine Julius and Japanese model repository. In: Proc. Int. Conf. Spoken Language Processing, pp. 3069–3072 (2004)
25. Asakawa, S., Minematsu, N., Hirose, K.: Multi-stream parameterization for structural speech recognition. In: Proc. Int. Conf. Acoustics, Speech, and Signal Processing (2008, submitted)
26. Takeshima, C., Tsuzaki, M., Irino, T.: Identification of size-modulated vowel sequences and temporal characteristics of the size extraction process, IEIEC Technical Report, SP2006-29, 13-17 (2006)
27. Smith, D.R., et al.: The processing and perception of size information in speech sounds. J. Acoust. Soc. Am. 171(1), 305–318 (2005)
28. Hayashi, Y., et al.: Comparison of perceptual characteristics of scaled vowels and words. In: Proc. Spring Meeting Acoust. Soc. Jpn., pp. 473–474 (2007)
29. Davis, R.D., Braun, E.M.: The gift of dyslexia, Perigee Trade (1997)
30. Frith, U.: Autism: Explaining the enigma. Blackwell Pub., Malden (1992)
31. Happe, F.: Autism: An introduction of psychological theory. UCL Press (1994)
32. Higashida, N., Higashida, M.: Messages to all my colleagues living on the planet. Escor Pub. (2005)
33. Nade, J.: The developing child with autism: evidences, speculations and vexed questions. In: Tutorial Session of IEEE Int. Conf. Development and Learning (2005)
34. Asami, T.: A book on my son, Hiroshi, Nakagawa Pub., vol. 5 (2006)
35. Trehub, S.E.: The developmental origins of musicality. Nature neurosciences 6, 669–673 (2003)
36. Hauser, M.D., McDermott, J.: The evolution of the music faculty: A comparative perspective. Nature neurosciences 6, 663–668 (2003)
37. Levitin, D.J., Rogers, S.E.: Absolute pitch: perception, coding, and controversies. Trends in Cognitive Sciences 9(1), 26–33 (2005)
38. Kojima, S.: A search for the origins of human speech: Auditory and vocal functions of the chimpanzee. Trans Pacific Press (2003)

Metrics for Evaluating the Serendipity of Recommendation Lists

Tomoko Murakami, Koichiro Mori, and Ryohei Orihara

Corporate Research and Development Center, Toshiba Corporation
1, Komukai Toshiba-cho, Saiwai-ku Kawasaki 212-8582, Japan
tomoko.murakami@toshiba.co.jp

Abstract. In this paper we propose metrics *unexpectedness* and *unexpectedness_r* for measuring the serendipity of recommendation lists produced by recommender systems. Recommender systems have been evaluated in many ways. Although prediction quality is frequently measured by various accuracy metrics, recommender systems must be not only accurate but also useful. A few researchers have argued that the bottom-line measure of the success of a recommender system should be user satisfaction. The basic idea of our metrics is that unexpectedness is the distance between the results produced by the method to be evaluated and those produced by a primitive prediction method. Here, *unexpectedness* is a metric for a whole recommendation list, while *unexpectedness_r* is that taking into account the ranking in the list. From the viewpoints of both accuracy and serendipity, we evaluated the results obtained by three prediction methods in experimental studies on television program recommendations.

1 Introduction

In information recommendation, one of the critical issues is to provide useful information to the user. From the viewpoint that items relevant to the user's preferences are also useful for the user, recommender systems have been evaluated by using metrics for measuring *accuracy*. The usefulness of a recommender system has a bearing not only on how accurately the system recommends items according to user preferences but also on how much serendipity it provides. Although precision and recall are well known metrics for measuring the accuracy of recommendations, a metric for evaluating serendipity is still an open problem in the research area of information recommendation.

In this paper, we propose metrics for measuring *unexpectedness* in recommendations. Unexpectedness is computed by comparing beliefs based on the idea that it represents the deviation from the result obtained by a primitive method. We describe empirical studies based on an experiment on television (TV) show recommendation. In the experiment, we evaluated recommendation lists from the viewpoints of both accuracy and serendipity and verified the validity of the proposed metrics.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 40–46, 2008.

This paper is organized as follows. We review related work in section 2 and propose metrics for measuring unexpectedness in section 3. We explain our experimental studies in section 4 and draw some conclusions in section 5.

2 Evaluation Criteria

Recommender systems have been evaluated using various metrics. From the perspective that good recommender systems recommend items that suit the preferences of the user, they are evaluated by using metrics related to accuracy [Breese 98, Billsus 98, Sarwar 00]. Two popular metrics for measuring accuracy are *precision* and *recall* [Cleverdon 68], which were developed in the research area of information retrieval. In addition to precision and recall, various predictive accuracy metrics are used in recommender systems [Shardanand 95, Breese 98]. The *mean absolute errors (MAE)* is a kind of predictive accuracy metric that measures how close the predicted results are to the user's true answer. Other popular metrics include the receiving operating characteristic (ROC) [Swets 69] and the half life utility metric [Breese 98].

However, if recommender systems achieved high accuracy by computing predictions only for easy-to-predict items, they would not necessarily be useful to the user. One of the major issues with accuracy is the inability to capture user satisfaction in existing recommender systems. User satisfaction with recommender systems is related not only to how accurately the system recommends but also to how much it supports the user's decision making. The evaluation metrics for measuring user satisfaction move beyond accuracy to include include *novelty* and *serendipity* [Herlocker 04].

A few researchers have argued that the bottom-line measure of the success of a recommender system should be user satisfaction [Herlocker 04, Ziegler 05]. Swearingern and Sinha examined how usefulness, novelty and usability are related to user satisfaction by means of a questionnaire survey and reported that they are significantly correlated [Swearingen 01]. Ziegler [Ziegler 05] tackled ways of avoiding the problem of recommendation list similarity, which means that the items in a recommendation list are similar to each other; for example, in book recommendation, a recommendation list might be filled with books by the same author or ones in the same genre. This often occurs for early-starters because the system is hardly able to obtain their preferences. Similar recommendation achieves a high performance for accuracy and a low one for serendipity. Ziegler introduced the metric *Intra-List Similarity (ILS)* to evaluate the similarity among items in a recommendation list. Ziegler achieved high user satisfaction by making topic-diversified recommendations by reducing the similarity of recommendation lists through the use of ILS. However, questions about whether topic-diversified recommendation achieves high serendipity still remain. Although user satisfaction should be evaluated from various aspects in decision making, we assume that it depends on whether a recommender system suggests unexpected items that are relevant to the user's preferences. To enable recommender systems to perform with high serendipity, metrics for measuring serendipity are required in advance.

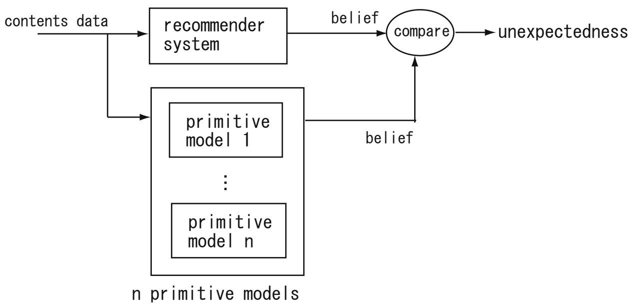


Fig. 1. Basic idea: unexpectedness is computed by comparing the belief with the result produced by an appropriate primitive model

3 Unexpectedness

We propose a metric for evaluating *unexpectedness* in a recommendation list composed of N items selected from M overall items. The basic idea of this metric is shown in figure 1. The metric is based on the idea that *unexpectedness* is low for easy-to-predict items and high for difficult-to-predict items. To estimate how easy prediction is, we introduce a primitive prediction method (PPM) that show high ratability. We expect that unexpectedness will be low for items predicted by PPM, whereas it will be high for items that cannot be predicted by it. Therefore, we assume that unexpectedness is the deviation from the result obtained by PPM and compare the beliefs produced by the recommender system's prediction method and PPM, as shown in figure 1.

Let $s_i(i = \cdots N)$ denote the i-th ranked item in the recommendation list, $Pr(s_i)$ denote s_i's belief generated by the prediction method, and $Prim(s_i)$ denote s_i's belief generated by PPM. Here, *belief* means the degree to which the recommender system confidently recommends each item. We introduce the relation to the user's preferences as $isrel(s_i) \in \{0, 1\}$, where $isrel(s_i) = 1$ means that s_i is related to the user's preferences and $isrel(s_i) = 0$ means that it is not. Although we use the relation to user preferences for all items to measure unexpectedness in this paper, it must be estimated because it is not given when we evaluate the recommendation. Since we assume that unexpectedness is the deviation from the result obtained by PPM, we define the unexpectedness of each item as follows.

$$unexpectedness = \frac{1}{N}\sum_{i=1}^{N} \max(Pr(s_i) - Prim(s_i), 0) \cdot isrel(s_i) \qquad (1)$$

Equation 1 is based on the idea that unexpectedness in recommendation is computed by the sum of the unexpectedness values for all items and that an item is unexpected only if its belief generated by the prediction method used for making recommendations is stronger than that generated by PPM. It is related to the user's preferences; otherwise, it is not unexpected.

However, evaluation using equation 1 does not take into consideration the ranking in recommendation. Thus, a recommendation list in which unexpected items lie higher in the rank is evaluated as being equal to one in which unexpected items lie lower in the rank. To avoid this, we propose another metric $unexpectedness\_r$. Let $count(i)(i = \cdots N)$ denote the number of items suited to the user's preferences lying above the i-th rank in the recommendation list.

$$unexpectedness\_r = \frac{1}{N}\sum_{i=1}^{N}\max(Pr(s_i) - Prim(s_i), 0) \cdot isrel(s_i) \cdot \frac{count(i)}{i} \quad (2)$$

Equation (2) is a metric that weights unexpected items lying higher in rank by multiplying equation(1) by $\frac{count(i)}{i}$.

4 Empirical Studies

4.1 Data Set

We checked the validity of the proposed metrics through a preliminary experiment on television(TV) show recommendation. In this experiment, we used TV show content data made from an electronic program guide and questionnaire survey data collected from 44 TV show viewers. The TV show content data was composed of detailed information on about 800 shows delivered by terrestrial and satellite broadcasting such as airdate, airtime, TV station, genre, and TV personalities. Questionnaire surveys were conducted in twice: the first survey had a duration of two weeks from February 3, 2007 and the second survey lasted for one week from March 3. The data from the first survey was used for learning the preference model of each viewer and data from the second survey was used for comparing the predicted results with the users' answers. In the first survey, we collected data on shows of interest, which included shows viewed in real time, prerecorded shows, and those that the viewers were interested in but missed recording or viewing in real time. In the second survey, we inquired about serendipity for recommended shows, which means whether the recommended shows had hardly ever been heard of before by viewers but seemed to be interesting or whether they were already known but were unexpected. We also asked viewers arbitrarily to list their favorite celebrities and genres.

4.2 Method

The recommended shows in the second survey were determined by preference models for each viewer generated by using the first survey data and delivered to each viewer once a day. There were at most 60 recommended shows made by three kinds of prediction methods, as follows.

BN_g recommendation based on Bayesian net models, which predict a viewer's preferences from his/her favorite genres

BN_{gh} recommendation based on Bayesian net models, which predict a viewer's preferences from his/her favorite genres and viewing habits

KF recommendation based on keywords weighted by Graham's method [Graham 02]

We introduced two types of Bayesian net models BN_g and BN_{gh} to compare differences in the structures of models used in the experiment. It has been reported that Bayesian nets show high prediction accuracy [Breese 98], but it depends on the model; that is, the structure of the Bayesian nets and the probability distribution on them. Here, BN_g was the simplest Bayesian net model and BN_{gh} was BN_g with viewing habit added to it.

KF is based on Graham's method [Graham 02], which is famous for filtering spam email accurately. We can apply Graham's method to recommendation because predicting a few items according to the user's preference is equivalent to discriminating emails that should be read from a large number of spam emails. The number of shows that suited a viewer's preferences was actually at most 2.5% of all the shows broadcast in one day in this experiment.

We evaluated the results by three prediction methods using precision, recall, unexpectedness (equation (1)), and unexpectedness_r (equation (2)). Precision is defined as the proportion of relevant items in the predicted items and reccall is defined as the proportion of predicted items in the relevant items. If we denote the number of relevant items by R, then precision and recall are defined as follows.

$$precision = \frac{count(N)}{N}$$

$$recall = \frac{count(N)}{R}$$

We introduced three candidates for PPM. One is a method of predicting shows based on the viewing timeframe during the first survey. The other two methods are based on a viewer's favorite genres or celebrities obtained from answers in the second survey. Thus, $Prim(s_i)$ is a binary function that takes 0 or 1, where $Prim(s_i) = 1$ means s_i is predicted by PPM. We also introduced the relation to the user's preferences as $isrel(s_i) \in 0, 1$, where $isrel(s_i) = 1$ means that s_i is related to the user's preferences and $isrel(s_i) = 0$ means that it is not.

4.3 Verification of Metrics

Some results of our evaluation of the *top 20* recommendations produced by the three prediction methods are given in table 1. They show the overall viewers' average of precision, recall, unexpectedness, and unexpectedness_r for one day in the experiment. Table 1 is composed of four subtables, where entries in the top table indicate the precision and recall in recommendation achieved by the three prediction methods. The results show that BN_{gh} was the best method for precision and recall. Thus, introducing viewing habit into the preference model is an effective way to achieve high accuracy. When PPM was based on

Table 1. Top-20 recommendation list

	BN_g		BN_{gh}		KF	
date	precision	recall	precision	recall	precision	recall
2007/3/3	0.2630	0.3583	0.4435	0.6075	0.3283	0.4524
2007/3/4	0.2087	0.2552	0.4217	0.5661	0.2793	0.3495
2007/3/5	0.3707	0.4455	0.4978	0.6201	0.3783	0.4477
2007/3/6	0.3446	0.4231	0.4924	0.6252	0.3935	0.4731
2007/3/7	0.3609	0.4509	0.4793	0.6074	0.4261	0.5037
2007/3/8	0.3565	0.4369	0.5098	0.6275	0.4065	0.4764
2007/3/9	0.3457	0.4270	0.4848	0.6019	0.4152	0.4819
average	0.3214	0.3996	0.4756	0.6080	0.3753	0.4550
PPM:genre						
	unexpected	unexpected_r	unexpected	unexpected_r	unexpected	unexpected_r
2007/3/3	0.0431	0.0236	0.1372	0.1018	0.1706	0.0921
2007/3/4	0.0304	0.0121	0.1088	0.0762	0.1422	0.0811
2007/3/5	0.0694	0.0416	0.1325	0.0980	0.1989	0.1285
2007/3/6	0.0598	0.0351	0.1273	0.0904	0.2000	0.1187
2007/3/7	0.0668	0.0414	0.1365	0.0979	0.2283	0.1393
2007/3/8	0.0622	0.0401	0.1451	0.1061	0.2022	0.1138
2007/3/9	0.0589	0.0332	0.1188	0.0797	0.2022	0.1241
average	0.0558	0.0324	0.1295	0.0929	0.1921	0.1140
PPM:celebrity						
	unexpected	unexpected_r	unexpected	unexpected_r	unexpected	unexpected_r
2007/3/3	0.0774	0.0406	0.2273	0.1606	0.3097	0.1676
2007/3/4	0.0621	0.0273	0.2265	0.1616	0.2726	0.1533
2007/3/5	0.1225	0.0707	0.2598	0.1886	0.3674	0.2217
2007/3/6	0.1184	0.0667	0.2663	0.1896	0.3804	0.2210
2007/3/7	0.1212	0.0708	0.2449	0.1767	0.4087	0.2429
2007/3/8	0.1198	0.0723	0.2780	0.2034	0.3989	0.2234
2007/3/9	0.1168	0.0642	0.2576	0.1751	0.4022	0.2323
average	0.1054	0.0589	0.2515	0.1794	0.3629	0.2089
PPM:timeframe						
	unexpected	unexpected_r	unexpected	unexpected_r	unexpected	unexpected_r
2007/3/3	0.0327	0.0177	0.0463	0.0301	0.0924	0.0581
2007/3/4	0.0244	0.0123	0.0318	0.0213	0.0629	0.0388
2007/3/5	0.0363	0.0213	0.0543	0.0380	0.1011	0.0646
2007/3/6	0.0339	0.0195	0.0411	0.0272	0.0924	0.0534
2007/3/7	0.0328	0.0219	0.0512	0.0338	0.1033	0.0628
2007/3/8	0.0371	0.0229	0.0418	0.0265	0.1043	0.0582
2007/3/9	0.0303	0.0169	0.0415	0.0271	0.0826	0.0478
average	0.0325	0.0189	0.0440	0.0292	0.0913	0.0548

the viewer's favorite genres or celebrities or viewing timeframe, the unexpectedness, and unexpectedness_r in recommendation by three prediction methods are shown in the second, third, and bottom subtables in table 1, respectively. We can see that KF was the best in unexpectedness and unexpectedness_r. Thus, KF was the most probable method among the three methods in both accuracy and serendipity.

We compared the result of unexpectedness achieved by the proposed metrics with the answer data obtained by the questionnaire survey. The number of shows that the viewers said they were unexpected in the recommendation lists predicted by BN_g, BN_{gh}, and KF were 17, 173, and 194 on overall the viewers' average, respectively. We can see that the number of unexpected shows reported in viewers' answers was reasonably close to the value of unexpectedness obtained by the proposed metrics. Therefore we consider that the proposed metrics can capture unexpectedness in recommendation to a certain degree.

5 Conclusion

In this paper, we proposed metrics *unexpectedness* and *unexpectedness_r* for measuring the serendipity of recommendation lists produced by recommender systems. The basic idea of the proposed metrics is that unexpectedness is considered to be the deviation from the result obtained by a primitive prediction method. From the viewpoints of both accuracy and serendipity, we evaluated the results produced by three prediction methods in the experimental studies on television program recommendation. In future, we plan to improve these metrics by investigating the relationship between serendipity in recommendation and user satisfaction.

References

[Breese 98] Breese, J., Herlocker, J., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. In: UAI 1998. Proc. of the 14th Conference on Uncertainty in Artificial Intelligence, pp. 43–52 (1998)

[Cleverdon 68] Cleverdon, C., Kean, M.: Factors Determining the Performance of Indexing Systems. In: Aslib Cranfield Research Project, Cranfield, England (1968)

[Billsus 98] Billsus, D., Pazzani, M.: Learning collaborative information filters. In: Proc. of the 15th National Conference on Artificial Intelligence(AAAI), pp. 46–53 (1998)

[Sarwar 00] Sarwar, B., et al.: Analysis of recommendation algorithms for E-commerce. In: EC 2000. Proc. of the 2nd ACM Conference on Electronic Commerce, pp. 285–295 (2000)

[Shardanand 95] Shardanand, U., Maes, P.: Social Information Filtering: Algorithms for Automating "Word of Mouth". In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems(ACM SIGCHI), pp. 210–217. ACM Press, New York (1995)

[Swets 69] Swets, J.: Effectiveness of information retrieval methods. Amer. Doc. 20, 72–89 (1969)

[Swearingen 01] Swearingen, K., Sinha, R.: Beyond Algorithms: An HCI Perspective on Recommender Systems. In: ACM SIGIR Workshop on Recommender Systems (2001)

[Herlocker 04] Herlocker, J., et al.: Evaluating Collaborative Filtering Recommender Systems. J. of ACM Transactions on Information Systems 22(1), 5–53 (2004)

[Ziegler 05] Ziegler, C.-N, et al.: Improving Recommendation Lists Through Topic Diversification. In: Proc. of WWW 2005, pp. 22–32 (2005)

[Graham 02] Graham, P.: A plan for spam (August 2002), http://www.paulgraham.com/spam.html

Moving Sound Source Extraction by Time-Variant Beamforming

Hirofumi Nakajima, Kazuhiro Nakadai, Yuji Hasegawa, and Hiroshi Tsujino

Honda Research Institute Japan Co., Ltd.
8-1 Honcho, Wako-shi, Saitama, 351-0114, Japan
{nakajima,nakadai,yuji.hasegawa,tsujino}@jp.honda-ri.com
http://www.honda-ri.com/

Abstract. We have developed a time-variant beamforming method that can extract sound signals from moving sound sources. It is difficult to recognize moving sound sources due to their amplitude and frequency distortions caused by the fact that the sources themselves are moving. Using our proposed method, the amplitude and frequency distortions of moving sound sources are precisely equalized so that these sound signals can be extracted. Numerical experiments showed that using our method improves moving sound source extraction. Extracting such sounds is important for successful natural human-robot interaction in a real environment because a robot has to recognize various types of sounds and sound sources.

Keywords: Beamforming, Time-variant system, Moving sound source.

1 Introduction

For successful human-robot interaction in a real environment, a robot should recognize not only a human voice but also other sounds in the surrounding environment [1]. If we separate sound sources in the environment into moving sound sources and fixed sound sources, the moving sound sources are used for predictions of temporal changes and the fixed sound sources are used for recognition of spacial information. We believe that useful human-robot interaction can be achieved by using both temporal and spacial information obtained from fixed and moving sources. There has been some research related to obtaining temporal and spacial information from fixed sound sources [2]. However, for moving sound sources, there has been less research on obtaining temporal and spacial information from moving sound sources because of problems in extracting the original sound source. It is difficult to extract a moving sound source because the source's own movement changes the amplitude and frequency of the source sound. We developed a new beamforming method that is able to precisely extract source signals by equalizing these changes caused by the source's movement. Our proposed method is different to conventional sound-extraction methods in that it digitizes the sound source positions and switches beamforming coefficients depending on these digitized source positions. This means there is no discontinuity

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 47–53, 2008.

due to switching beamforming coefficients and it is not necessary to equalize frequency changes due to the Doppler effect and amplitude changes because there is theoretically no errors during extractions.

1.1 Time-Variant System and Time-Variant Convolution

We assume that there is a moving sound source whose position and signal waveform are $\mathbf{p}(t)$ and $s(t)$, respectively. The observed signal at position $\mathbf{q}(t)$ can be derived from the wave equation for moving sources [3] as

$$x(t) = \int s(t_s)h(t - t_s, \mathbf{p}(t_s))dt_s, \tag{1}$$

where $h(t, \mathbf{p}(t_s))$ is the impulse response from the source at position $\mathbf{p}(t_s)$ to the observation point. This equation implicates that the output signal $x(t)$ can be calculated if the source signal $s(t)$ and impulse responses $h(t, \mathbf{p}(t_s))$ at all positions can be located even if the source is moving. In this paper, we define Eq. (1) as the **time-variant** convolution operation. In the discrete time system, this operation is given approximately by [4]

$$x(k) = \sum_{k\ =0}^{\infty} s(k_S)h(k - k_S, \mathbf{p}(k_S)). \tag{2}$$

Note that the sampling frequency should be set greater than twice the upper limit of the source frequency **accounting for the Doppler effect** due to the source's movement. Also Eq. (2) is represented as vectors and a matrix: Further, Eq. (2) can be represented as vectors and a matrix:

$$\begin{aligned}
\mathbf{x} &= \mathbf{Hs} \\
\mathbf{s} &= [s(1), s(2), ..., s(L_S)]^T \\
\mathbf{x} &= [x(1), x(2), ..., x(L_S + L_h - 1)]^T \\
\mathbf{H} &= \begin{bmatrix}
h(1, \mathbf{p}(1)) & 0 & \cdots & 0 \\
h(2, \mathbf{p}(1)) & h(1, \mathbf{p}(2)) & \ddots & \vdots \\
\vdots & h(2, \mathbf{p}(2)) & \ddots & 0 \\
h(L_h, \mathbf{p}(1)) & \vdots & & h(1, \mathbf{p}(L_S)) \\
0 & h(L_h, \mathbf{p}(2)) & & h(2, \mathbf{p}(L_S)) \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & h(L_h, \mathbf{p}(L_S))
\end{bmatrix}.
\end{aligned} \tag{3}$$

Here, $\mathbf{s}$ denotes the source signal vector, $\mathbf{x}$ the observed signal vector, and $\mathbf{H}$ the **time-variant** convolution matrix, which is an extension of a conventional time-invariant convolution matrix [5]. If we define the source movement pattern as position vector $\mathbf{p}(\mathbf{k})$ of the discrete time vector $\mathbf{k}$, $\mathbf{H}$ is determined by the pattern $\mathbf{p}(\mathbf{k})$ and the observation point $\mathbf{q}$. Here, we assume the discrete time origin to be 0, the source signal length as L_S, and the the impulse response length as L_h.

2 High Precision Beamforming for Moving Sound Source

Our proposed method of beamforming (BF) is the time-variant extension version of the BF [6] that can be used for any transfer function. Figure 1 shows the propagation process and the system configuration of BF. In this figure, the number of microphones is N, $\mathbf{s}$ represents a source signal vector with L_S rows, $\mathbf{x}(n)$ denotes an observed signal vector of the n-th microphone with $L_S + L_S - 1$ rows, $\mathbf{H}(\mathbf{p}(\mathbf{k}), n)$ represents the time-variant convolution matrix from the moving source with the pattern $\mathbf{p}(\mathbf{k})$ to the n-th microphone, $\mathbf{G}(n)$ defines a BF coefficient matrix connected to the n-th microphone, and $\mathbf{y}$ is the output signal of the system. The row number and column number of the coefficient matrix $\mathbf{G}(n)$ correspond to the times of observation signal and source signal, respectively. The BF output $\mathbf{y}$ is calculated as

$$\begin{aligned} \mathbf{y} &= \mathbf{G}^T \mathbf{x} \\ \mathbf{G} &= [\mathbf{G}(1), \mathbf{G}(2), ..., \mathbf{G}(N)]^T \\ \mathbf{x} &= [\mathbf{x}(1), \mathbf{x}(2), ..., \mathbf{x}(N)]^T. \end{aligned} \tag{4}$$

Input signal vector group $\mathbf{x}$ is represented by

$$\begin{aligned} \mathbf{x} &= \mathbf{H}(\mathbf{p}(\mathbf{k}))\mathbf{s} \\ \mathbf{H}(\mathbf{p}(\mathbf{k})) &= [\mathbf{H}(\mathbf{p}(\mathbf{k}), 1), \mathbf{H}(\mathbf{p}(\mathbf{k}), 2), ..., \mathbf{H}(\mathbf{p}(\mathbf{k}), N)]^T. \end{aligned} \tag{5}$$

To extract source signals from observed signals, $\mathbf{y}$ should equal $\mathbf{s}$. The BF coefficient $\mathbf{G}(n)$ should become one of the solutions of

$$\mathbf{H}(\mathbf{p}(\mathbf{k}))\mathbf{G} = \mathbf{I}, \tag{6}$$

where $\mathbf{I}$ is an identity matrix. If we solve the above equation by using the pseudo-inverse matrix of $\mathbf{H}(\mathbf{p}(\mathbf{k}))$, we get the minimum norm weighted delay-and-sum beamformer for moving sources. Further, if we can obtain the moving pattern $\mathbf{p}_U(\mathbf{k})$ of any obstruction sources and its convolution matrix $\mathbf{H}_U(\mathbf{p}(\mathbf{k}))$, we can

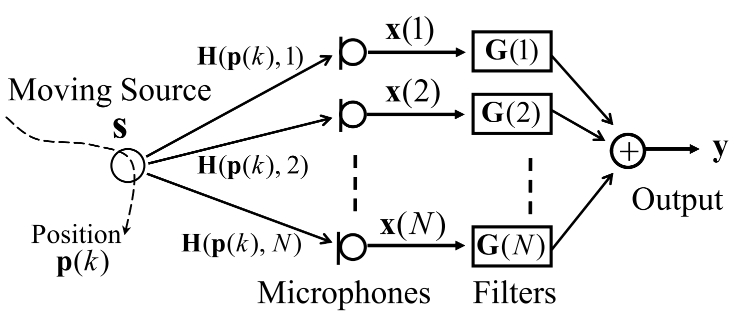


Fig. 1. System Configuration

get the null-based beamformer or the minimum side-lobe beamformer by adding the following constraints [6,7]:

$$\mathbf{H}(\mathbf{p}_U(\mathbf{k}))\mathbf{G} = \mathbf{0}, \tag{7}$$

where $\mathbf{0}$ means a zero matrix.

3 Numerical Simulation

We performed a numerical simulation (See Fig. 2 for layout). The sound sources of the experiment were a moving source S1 and a fixed source S2. We set a 125Hz sine wave for S1 (Fig. 3) and 400Hz sine wave for S2. The sampling frequency was 1 kHz and the waveform length of both S1 and S2 was set 0.5 s. The objective of the simulation was to extract the source waveform of S1 using a null-based beamformer of 3 elements: M1, M2 and M3. For comparison, we designed conventional fixed beamformers at all the source directions digitized for every 10

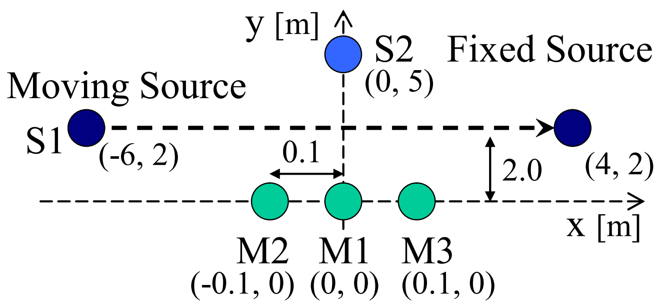


Fig. 2. Layout of numerical experiment

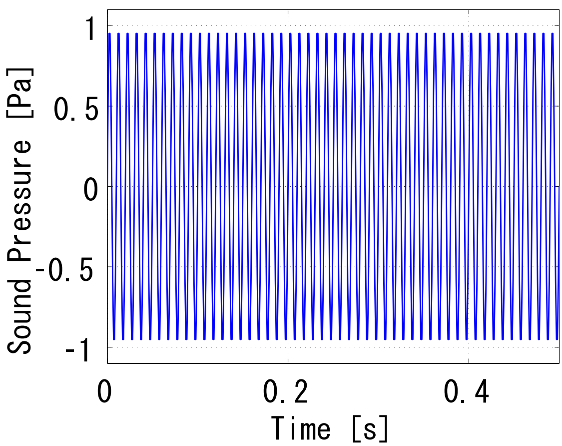


Fig. 3. Source signal S1

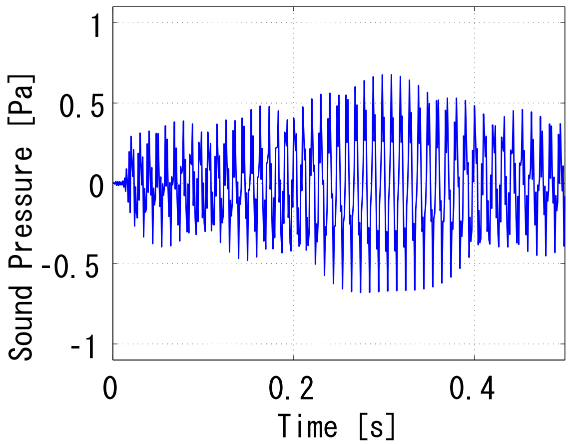


Fig. 4. Input signal at M1

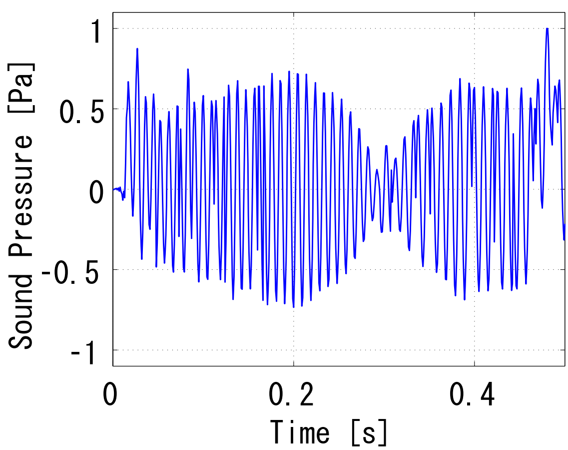


Fig. 5. Output signal of conventional BF

degrees. We selected a beamformer having the maximum power and used the beamformer's output as the final output, that was evaluated as a conventional method. For our proposed BF, a simultaneous equation derived from Eqs. (6) and (7) was solved. Figure 4 shows the observed signal at M1 synthesized by Eq. (3). We see S2 signal (high frequency) and S1 signal changing its amplitude and frequency. Figure 5 and 6 represent output signals of conventional BF and proposed BF, respectively. In Fig. 5, we found that the conventional BF cancel the high frequency component originated by S2; however, the output signal amplitude still changes for S1 because of its movement and the discontinuity resulting from BF coefficient switching. The output from our BF (shown in Fig. 6) does

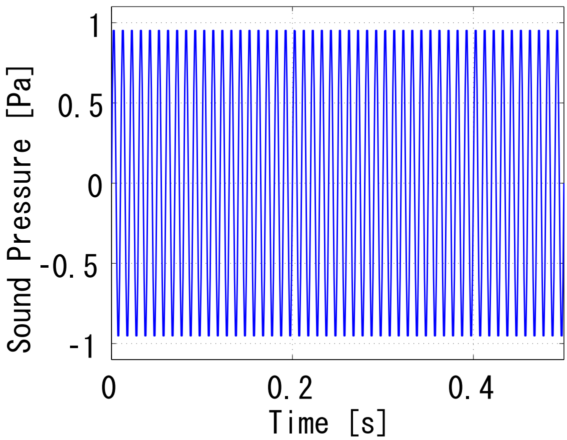


Fig. 6. Output signal of proposed BF

Table 1. Estimation error E (dB)

	without noise	with noise (-40dB)
Conventional BF	0.6	1.0
Proposed BF	-260	-14.3

not contain this amplitude changing or discontinuity. We see that our proposed BF precisely extracts the signal of the moving source S1. The reason that the conventional BF output signal at around 0.3 s becomes low is that the focused and null directions are close. Table 1 denotes the estimation error E (dB) of the conventional and our proposed BF outputs with/without measurement noises. The error E was evaluated using

$$E = 10 \log_{10} \left(\frac{|\mathbf{s} - \hat{\mathbf{s}}|^2}{|\mathbf{s}|^2} \right). \tag{8}$$

Here, $\mathbf{s}$ represents the original source signal and $\hat{\mathbf{s}}$ defines the estimated source signal. In this table, we can see that our proposed method perfectly reconstructs the moving source signal in an ideal environment with no noise and it also achieves better estimations compared to those of conventional methods in a noisy environment.

4 Summary

We have developed a new beamformer that equalizes amplitude and frequency changing caused by source movements and precisely extracts source signals. According to the numerical simulation results, our proposed method is effective compared with conventional methods even if the measurement signal contains some noise.

References

1. Nakadai, K., et al.: Active audition for humanoid. In: AAAI 2000. 17th National Conf. on Artificial Intelligence, pp. 832–839. AAAI, Stanford, California, USA (2000)
2. Yamamoto, S., et al.: Assessment of General Applicability of Robot Audition system by Recognizing Three Simultaneous Speeches. In: IROS 2004. Proc of 2004 Int. Conf. on Intelligent Robots and Systems, IEEE/RSJ, pp. 2111–2116 (2004)
3. Morse, P.M., Ingard, K.U.: Theoretical Acoustics. In: Princeton, USA, pp. 717–732 (1968)
4. Okuyama, T., Mastuhisa, H., Utsuno, H.: Calculation method of the Sound pressure at a Moving point in the Virtual space using Interpolation of the Transfer function. In: Proc of Noise and Vibration meeting, Acoust Sciety of Japan (in Japanese) (2006) N-2006-64
5. Miyoshi, M., Kaneda, Y.: Inverse Filtering of Room Acoustics. IEEE Trans. on Acoustic Speech and Signal Processing ASSP-36(2), 145–152 (1988)
6. Nakajima, H.: Minumum mean energy sidelobe beamforming by using indefinite term. The Jounal of the Acoustic Society of Japan, 726–737 (2006) (in Japanese)
7. Nakajima, H., Tanaka, N., Tsuru, H.: Minimum sidelobe beamforming based on Mini-Max criterion. Acoust. Sci. & Tech. 25(6), 486–488 (2004)

Video Scene Retrieval Using Online Video Annotation

Tomoki Masuda[1], Daisuke Yamamoto[1], Shigeki Ohira[2], and Katashi Nagao[3]

[1] Graduate School of Information Science, Nagoya University
[2] EcoTopia Science Institute, Nagoya University
[3] Center for Information Media Studies, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan
{masuda,yamamoto,ohira,nagao}@nagao.nuie.nagoya-u.ac.jp

Abstract. In this paper, we propose an efficient method for extracting scene tags from online video annotation (e.g., comments about video scenes). To evaluate this method by applying extracted information to video scene retrieval, we have developed a video scene retrieval system based on scene tags (i.e., tags associated with video scenes). We have also developed a tag selection system that enables online users to select appropriate scene tags from data created automatically from online video annotation. Furthermore, we performed experiments on tag selection and video scene retrieval. We found that scene tags extracted by using our tag selection system had better cost performance than ones created using a conventional client-side video annotation tool.

1 Introduction

In recent years, a lot of video contents have been delivered and shared on the Web through the development of internet technology and the spread of broadband access lines. With the appearance of video sharing services, such as YouTube[1], the amount of video contents, which includes not only commercial but also user-generated contents, has been increasing explosively. Therefore, the demand for applications such as video scene retrieval and video summarization is rising and expected to rise even more in the future.

To make these applications, we must acquire meta information corresponding to the content of a video, which we call annotation [1]. We have developed an online video annotation system called Synvie in recent years [2]. We call data extracted from users' natural knowledgeable activities online video annotation. And we opened a public experimental service of Synvie and are accumulating annotation data now[2]. Online video annotation data have both advantages and disadvantages. They may contain video information from various people's viewpoints. And because they are accumulated from users' natural activities, the annotation costs are very small. However, they contain useless information, so we must screen them for use in practical applications.

[1] YouTube: http://www.youtube.com/
[2] Synvie Beta: http://video.nagao.nuie.nagoya-u.ac.jp/

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 54–62, 2008.

In this paper, we propose an efficient screening method and an application based on online video annotation. Specifically, we propose an efficient method for extracting scene tags from online video annotation. To evaluate the method by applying extracted information to video scene retrieval, we developed a video scene retrieval system based on scene tags.(i.e., tags associated with video scenes). Moreover, we performed experiments on screening tags and video scene retrieval. Through these experiments, we verified the usefulness of online video annotation and our screening method and application of it.

2 Creating Scene Tags (Tagging Video Scenes)

Creating scene tags means relating keywords to an arbitrary time code in a video. Scene tags contain nouns, verbs, and adjectives, but do not contain particles or auxiliary verbs. Unknown words are treated as nouns. We created scene tags by using three methods for 27 videos that were registered in Synvie. The length of the used videos was about 349 seconds on average: the longest video was 768 seconds and the shortest was 76 seconds. We used various kind of videos, e.g., educational videos, stories, and entertainment videos. In next chapter, we compare the usefulness of tags created by each method.

2.1 Tagging Using an Annotation Tool

One annotator added scene tags by using a tool that enables a user to add tags at an arbitrary time in a video. The annotator who was not a creator and did not have any special knowledge about the videos added objective information acquired from images and sounds to video scenes as scene tags in detail and exhaustively. This method is a kind of conventional client-side video annotation [3]. We defined the human cost for creating scene tags as the time that an annotator spent adding them. This was 1480 seconds on average: the longest time was 3692 seconds and the shortest was 582 seconds.

2.2 Automatic Extraction of Scene Tags from Online Video Annotation

Synvie is a video sharing system that lets users comment on video scenes and quote them in a weblog. We have been running a public experimental service since July 1, 2006 and analyzing data accumulated from July 1 to October 30, 2006. We gathered 97 registered users and 94 videos. From the accumulated annotation data, we could acquire text data related to time data. Through some processing, we created scene tags automatically from annotation data. The tag creation process is shown below.

1. Using morphological analysis (using a Japanese morphological analyzer "Cabocha" [4]).
2. Removing stop words.

3. Extracting nouns, verbs, adjectives, and unknown words.
4. Relating words to time data and saving them in a database.

These processes can be performed automatically and annotation data can be accumulated through natural communication by humans on the web, so it can be said that the human costs for creating scene tags is extremely small. 153 scene tags were created on average for 27 videos by this method.

2.3 Scene Tag Extraction Using an Online Tag Screening System

We can easily predict that annotation data may include useless data such as data having no relation to the video. In comments or weblogs, users do not necessarily refer to the contents of the video. Therefore, scene tags created from the online video annotation automatically, as described in section 2-2, have a high probability of including useless tags. Indeed, we found scene tags that were obviously unsuitable for scenes being viewed. They included tags that were not meaningless but were unsuitable for the scenes and tags that lost their meaning as a result of the morphological analysis processes. For these reasons, the quality of tags created automatically from online video annotation will not be high. So we must screen tags in order to use them in practical applications. If we succeed in screening them appropriately, we will obtain higher-quality tags.

Because it would be ideal for this screening to be achieved successfully through automatic processing, we tried to do it by various different methods. First, we used the well-known technique TF-IDF (Term Frequency-Inverse Document Frequency) [5]. However, this technique needs a large quantity of documents to be successful in finding appropriate words. Second, by using Google Web API[3], we tried to score words by co-occurrence relations to tags that had been added when the video was registered. But in this method, the scores of words that appear in general documents were higher than those of words that were strongly related to the scene. These results show that it is very difficult to perform appropriate screening of scene tags by automatic processing and that manual processing by humans is necessary for it to be successful. So we developed a tag screening system that enables online users to select appropriate scene tags from data created automatically from online video annotation (Figure 1). We can guess that the quality of tags selected by humans is high. But the more human cost we include in this process, the more we lose the advantages of using online video annotation. Mechanisms that enable users to select tags efficiently are required in order to reduce human costs. This system is used by one or more users. Users watch a video, and when a time code to which tags have been added comes, the video stops temporarily and the users select tags that are appropriate to the scene. We performed experiments on screening tags using this system. The number of subjects for each video was two or three. We defined the human cost for creating tags by using this system as the time that each user spent in selecting. We calculated this value from the automatically measured time. The average time was 314 seconds, which is 1/5 of the time spent for creating tags using the

[3] Google Code: http://code.google.com/

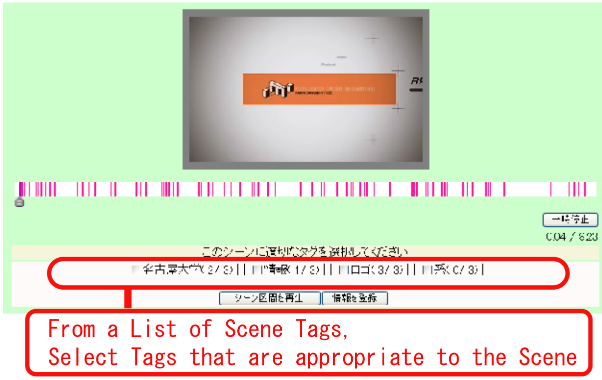


Fig. 1. Online tag screening system

Table 1. Comparison of human costs for creating tags

Tag Creation Method	Human Costs (Second)
Tagging Offline	1480
Extracting Automatically	0
Screening Online	314

annotation tool described in section 2.1. In this experiment, 55 scene tags were created on average for 27 videos. The results show that 36.2 percentages of tags that were created automatically from online video annotation were judged to be appropriate to the scene. A comparison of the three methods in terms of human cost for creating tags is given in Table 1.

3 Video Scene Retrieval

We have developed a tag-based video scene retrieval system based on a new concept. And we performed experiments on video scene retrieval using the scene tags created as described in the previous section.

3.1 Tag-Based Video Scene Retrieval System

Our video scene retrieval system is based on the mechanism of tag-cloud [6] and makes the most of the characteristics of scene tags. Scene tags generated from annotations have an essential problem in that appropriate tags are not necessarily given to all scenes, and their completeness is small. When there are not enough annotations for each video, it is hard to apply usual search techniques such as exact matching using tags. But tags also have a strong point in that a large number of tags can be displayed in a small space on a browser and this can be helpful for efficient retrieval. We developed this system considering these characteristics of scene tags. The process of retrieving video scenes is shown below.

1. Select an arbitrary number of tags and submit them as a query.
2. A list of videos is returned corresponding to the query. And a timeline seek bar that highlights the time ranges of tags used as the query has been added and a list of all scene tags that have been added to the video are displayed with each video.
3. Select tags from a list to correspond to the timeline seek bar.
4. Move the seek bar to view thumbnail images for an arbitrary time code.
5. Play the video from the arbitrary time code.

Fig. 2. Tag-cloud for video scene retrieval

The top page of this video scene retrieval system is shown in Figure 2. A tag-cloud composed of scene tags and tags that were added when the video was registered are displayed. Tags are classified into nouns (including unknown words), verbs, and adjectives in ABC order and A-I-U-E-O order (Japanese order). When a tag is clicked, the word of the tag is added to the text field for searching, so it is not necessary to input text using a keyboard. And users can use an incremental search for tags. Incremental searching is a search that progressively finds a match for the search string as each character is typed. In this system, when a letter is typed, only tags that start with that letter are displayed and the others are hidden. These functions help users to find tags for making queries from a large number of tags. The output of the search is a list of videos that include these tags (Figure 3).

Fig. 3. Results of search

Each video has a seek bar associated with scene tags and thumbnail images arranged along the time axis. The timeline seek bar helps users to view video scenes on a web browser without accessing the video itself. When the user drags the seek bar to an arbitrary time code, the system displays thumbnail images synchronized with the time code of the seek bar. This function helps users to view images of a time code to which tags have not been added. Because the time ranges to which tags have been added are highlighted on the seek bar, the user can understand the content of the video by browsing these tags and thumbnail images without actually watching it. Moreover, when the user clicks an interesting-looking tag, the temporal location of the tag is displayed on the seek bar. These actions are repeated and video scenes that users want to see are found. An example of an image for which video scene retrieval was performed is shown in Figure 4. We are continuing with the development to make retrieval more efficient. And we have been running a public experimental service from February 27, 2007.[4]

3.2 Experiment on Video Scene Retrieval

We performed experiments on our video scene retrieval system. We chose nine scenes as retrieval targets. The questions asked for a "scene where a certain animal was reflected by the parent and child", "scene before the person who was snowboarding crashed into the edge of the course", etc. and did not necessarily include words and phrases given as scene tags, but subjects could guess the scene by getting hints from the scene tags. Moreover, to ensure that the answer to each question was the only time range, a thumbnail image was also given as well as the

[4] Divie: http://video.nagao.nuie.nagoya-u.ac.jp/search/top

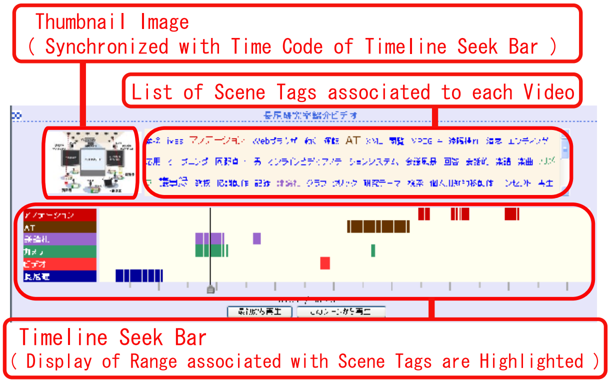


Fig. 4. Interface for video searching with a seek bar and tags

text. Subjects retrieved the answer scene to each question, and the time spent on the answer was measured automatically. The number of subjects was nine. Subjects retrieved scenes using tags created by the three methods described in the previous section. Each subject retrieved three scenes by using tags created by each method (9 scenes in total). Therefore, each scene tag creation method could be compared impartially. We prepared an experimental top page that did not reveal which method was used to create the tags that subjects used for each retrieval. The data that we acquired in this experiment are shown below.

– Scenes decided as answers.
– Time spent on retrieval.
– Queries used for retrieval.
– Viewed scenes.

Because all subjects were able to discover a correct scene for all questions in this experiment, we could not compare the scene tag creation methods from this viewpoint. Therefore, we compared them according to the time taken for the retrieval. The experimental results are shown in Table 2. It can be thought that the difference did not go out by some influences other than unlike tags at the retrieval time because the number of queries submitted increased with the average retrieval time. The retrieval time was longest for tags extracted automatically from online video annotation and shortest for tags created using an annotation tool. This result shows that retrieval time was shorter in the two methods that involved human cost for creating scene tags. Though automatic tag creation from online video annotation was the best method in terms of tag creation cost, its retrieval costs were very high. And because it is essential to shorten the retrieval time for video contents, which will grow in volume in the future, it is necessary to put the human cost for tag creation.

Therefore, we compared tags created using an annotation tool with ones created using the online tag screening system in terms of their cost performance

Table 2. Results of experiments on video scene retrieval

Tag Creation Method	Time (Second)	Number of Queries
Tagging Offline	118.1	132
Extracting Automatically	169.6	202
Screening Online	145.4	156

(cost-effectiveness), that is, the ratio of retrieval cost to creation cost. Cost performance C of the tag was calculated by the following equation, where n is equal to the time spent for retrieval when automatically created tags were used, RT is the time spent for retrieval when the tags created by each method were used, and CT is time spent creating the tags.

$$C = \frac{n - RT}{CT} \times 100 \qquad (1)$$

From this equation, we can calculate how much the retrieval time was reduced by spending 100 seconds creating tags. The results are given in Table 3. These results indicate that, when comparing methods from the viewpoint of cost performance of each tag, the method of creating tags using the online tag screening system was best in this experiment.

Table 3. Cost performance of each tag

Tag Creation Method	Cost Performance of Each Tag
Tagging Offline	3.48
Screening Online	7.71

4 Conclusion and Future Work

4.1 Conclusion

In this paper, by using and screening online video annotation, we could create useful scene tags without high human costs. The results of experiments showed the usefulness of online video annotation and of the screening method. We have developed a tag-based scene retrieval system based on a new concept and proposed an application of online video annotation.

4.2 Future Work

We must verify our method using a large quantity of data accumulated on a public experimental service. Moreover, we must improve the interface and the algorithm of the video scene retrieval system for more efficient retrieval.

References

1. Nagao, K.: Digital Content Annotation and Transcoding. Artech House Publishers (2003)
2. Yamamoto, D., Ohira, S., Nagao, K.: Weblog-style Video Annotation and Syndication. In: Proceedings of the 1st International Conference on the Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS) (2005)
3. Nagao, K., Ohira, S., Yoneoka, M.: Annotation-Based Multimedia Summerization and Translation. In: COLING 2002. Proceedings of the Nineteenth International Conference on Computational Linguistics (2002)
4. Kudo, T., Matsumoto, Y.: Fast Methods for Kernel-Based Text Analysis. In: ACL 2003. Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, pp. 24–31 (2003)
5. Salton, G., Wong, A., Yang, C.S.: A Vector Space Model for Automatic Indexing. Communications of the ACM 18(11), 613–620 (1975)
6. Rivadeneira, A.W., et al.: Getting our head in the clouds: Toward evaluation studies of tagclouds. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 995–998 (2007)

Spatio-temporal Semantic Map for Acquiring and Retargeting Knowledge on Everyday Life Behavior

Yoshifumi Nishida[1], Yoichi Motomura[1], Goro Kawakami[2,1], Naoaki Matsumoto[2,1], and Hiroshi Mizoguchi[2,1]

[1] Digital Human Research Center, National Institute of Advanced Industrial Science and Technology, Aomi 2-41-6, Koto-ku, Tokyo, 135-0064, Japan
[2] Tokyo University of Science, Yamazaki 2641, Noda-shi, Chiba, 278-8510, Japan

Abstract. Ubiquitous sensing technology and statistical modeling technology are making it possible to conduct scientific research on our everyday lives. These technologies enable us to quantitatively observe and record everyday life phenomena and thus acquire reusable knowledge from the large-scale sensory data. This paper proposes a "Spatio-temporal Semantic (STS) Mapping System," which is a general framework for modeling human behavior in an everyday life environment. The STS mapping system consists of a wearable sensor for spatially and temporally measuring human behavior in an everyday setting together with Bayesian network modeling software to acquire and retarget the gathered knowledge on human behavior. We consider this STS mapping system from both the theoretical and practical viewpoints. The theoretical framework describes a behavioral model in terms of a random field or as a point process in spatial statistics. The practical aspect of this paper is concerned with a case study in which the proposed system is used to create a new type of playground equipment design that is safer for children, in order to demonstrate the practical effectiveness of the system. In this case study, we studied children's behavior using a wireless wearable location-electromyography sensor that was developed by the authors, and then a behavioral model was constructed from the measured data. The case study shows that everyday life science can be used to improve product designs by measuring and modeling the way it is used.

1 Introduction

Scientists and engineers have a limited understanding of the dynamics and properties of everyday life despite its familiarity. Although standard models in scientific fields such as quantum theory and cosmology exist to explain and generate most phenomena, nothing yet exists that might represent a standard model of everyday life. Modeling everyday life requires representing it by quantitatively observing it and constructing a model from a large-scale amount of observed data. The recent development of ubiquitous sensing technology, which enables

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 63–75, 2008.

the observation of physical phenomena in total living space, and statistical modeling technology, which enables the construction of a model from the observed data, will open the way for the field of *science and technology of everyday life.*

An everyday life model should be considered a "reusable model" from the viewpoint of practicality as well as science, to represent the apparent phenomena in human behavior, but also to explain the underlying semantic causality among behaviors, the environment, situations, and conditions. Even if the situations and conditions are different from those for which the model was created, the model can be reused under different situations and conditions if they can be abstracted in terms of the same semantic structure. In this paper, a reusable model indicates that we can re-use the causal model to simulate human behavior under such differing situations and conditions.

This paper focuses on everyday life behavior for the following reasons. First, wearable sensors [1] and location sensors [2] are available to help quantitatively and spatio-temporally describe everyday life phenomena. Just as the Global Positioning System (GPS) [3] and the Geographical Information System (GIS) [4] software packages are useful for representing the spatial information for given positions worldwide, these ubiquitous sensing technologies will result in the proliferation of spatio-temporally indexed data sets that can be obtained from everyday life settings. These data sets can be used for assisting with the application of science and technology in studying everyday situations.

Second, statistics customized for use with spatial data, referred to as spatial statistics [5,6], have recently been developed for analyzing spatial data. Although our aim is to model a semantic structure that underlies a spatio-temporal phenomena rather than a spatial or spatio-temporal structure, we can take advantage of the spatial statistics approach as a starting point. However, we need to expand spatial statistics to acquire the reusable semantic knowledge from everyday life data that are spatio-temporally indexed. Fortunately, another statistical modeling technology has become available for acquiring reusable semantic knowledge from a large amount of data. In particular, a Bayesian network method can be used to create a realistic model [7] and therefore, it can be used for bridging the gap between a spatio-temporal data space and a semantic state space.

Finally, the use of science and technology for studying the behavior of people in everyday life is urgently required in our society. By better understanding everyday life behavior, we can better improve the quality of life. For example, as children develop their behavioral capabilities in everyday life, their rate of injury incidence rapidly increases. After a child learns to walk at around one year of age, the primary cause of death is surprisingly not illness, but injury [8]. In 2006, the World Health Organization (WHO) announced their ten-year action plan for child injury prevention [9]. Children behavior science is applicable to preventing childhood injury. Our research group has been working on childhood injury prevention [10].

This paper addresses the problem of creating a semantic model from the behavior of people in everyday life from spatio-temporally indexed data. We propose a "Spatio-temporal Semantic (STS) Mapping System", which is the general

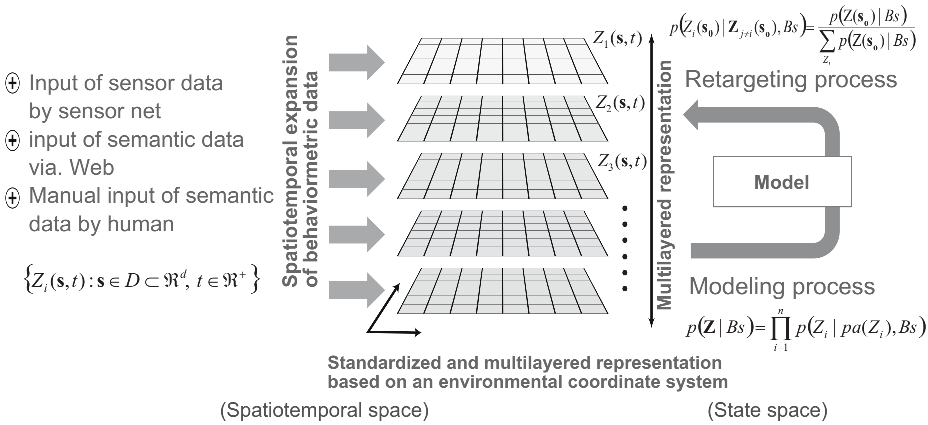


Fig. 1. Concept of spatiotemporal-semantic mapping system

framework for modeling human behavior in an everyday environment. The STS mapping system consists of a wearable sensor for spatially and temporally measuring human behavior, together with Bayesian network modeling software to acquire and retarget the gathered knowledge on human behavior. This paper also presents a case study for applying the proposed system to the development of a new playground equipment design that is safer for children, in order to show the practical effectiveness of the system. In this case study, in situ observations and measurements were made of 47 children playing with or on equipment using a wireless wearable location-electromyography sensor that was developed by the authors, and then a children's behavior model was constructed from the measured data.

2 Spatio-temporal Semantic Mapping System

2.1 Concept of STS Mapping System

Modeling human behavior, in order to develop a reusable causality model based on the behavior, environment, situations and conditions, can be divided into two components: representation of the scenario and the knowledge acquisition process.

We will begin by discussing the representation. Our environment consists of objects spatially distributed in an everyday life space. Humans exhibit a variety of behaviors by interacting with such spatially distributed objects. Therefore, the way a person behaves in their everyday life can be represented in an environmental coordinate system. A geographical information system (GIS) is well known as a representation system for describing a wide variety of information based on an environmental coordinate system. We utilize a similar representation: we standardize and structuralize human behavior in terms of a multilayered information structure by tagging them with environmental coordinates.

Next, we will discuss the knowledge acquisition process. In order to gather information on the behaviors and retarget the acquired knowledge to a new environment, we have to abstractly represent the data by expressing them in terms of a semantic state space, and then find stable structures in this space. Using a statistical modeling method can help us do this. The Bayesian network paradigm is well known as a method for developing a graphical model for a semantic state space. We utilize a Bayesian network for the knowledge acquisition.

Figure 1 presents our proposed concept for a Spatio-temporal Semantic (STS) mapping system. The STS mapping system consists of the following components: 1) a spatio-temporal extension of the behaviormetric sensor integrated with a non-location behaviormetric sensor and a location sensor, 2) a standardized and multilayered representation of information based on an environmental coordinate system, and 3) a statistical modeling process for knowledge acquisition from the represented information, and a retargeting process.

We can express the sensory data in a standardized and multilayered way, extract the knowledge using the modeling process, and then apply the extracted knowledge to a new target by using this STS mapping system.

2.2 Formulation of Behavior Phenomena in STS Mapping System

In this section we will describe the formulation of behavior phenomena in the STS mapping system. We view the observed behavior attributes as realizations from a kind of random field referred to as a spatial point process in which random variables have a kind of structure in a state space. We express the spatial point process by

$$\mathbf{Z} = \left\{ Z_i(\mathbf{s}, t) : \mathbf{s} \in D \subset \Re^d, t \in \Re^+ \right\}, \tag{1}$$

where $Z_i(\mathbf{s}, t)$ denotes the i-th attribute at location $\mathbf{s}$ at time t, D denotes a region of interest, and typically $d = 2$ or 3 and $\mathbf{s} = [x, y]'$ or $[x, y, z]'$ if we are dealing with two- or three-dimensional space.

For example, the data obtained by an ultrasonic location sensor, which is a kind of GPS, is composed of three-dimensional spatial data and temporal data. So it can be expressed by

$$\mathbf{E} = \left\{ E_i(\mathbf{s}, t) : \mathbf{s} \in D \subset \Re^d, t \in \Re^+ \right\}, \tag{2}$$

where $E_i(\mathbf{s}, t)$ denotes the occurrence of an event at location $\mathbf{s}$ at time t. Then E_i normally equals a constant 1 for all $\mathbf{s}$ and t. On the other hand, the data obtained by non-location sensors, such as a wearable electromyography (EMG) sensor, can be expressed by

$$\mathbf{Z} = \left\{ Z_i(t) : t \in \Re^+ \right\}. \tag{3}$$

By integrating the system with a location sensor expressed by Eq. (2) with a non-location sensor expressed by Eq. (3) and simultaneously measuring both data, we can obtain spatiotemporal attribute data expressed by Eq. (1). Thus we can expand the non-location sensor to a spatiotemporal attribute sensor by combining it with the location sensor. We call this integration a *spatio-temporal extension*. An example of a spatio-temporal extension of the wearable (EMG) sensor is described in detail in the next section.

There is another type of data expressed by

$$\mathbf{Z} = \left\{ Z_i(s) : \mathbf{s} \in D \subset \Re^d \right\} . \tag{4}$$

For example, this type of data can be obtained by using a three-dimensional scanner (e.g., a laser range finder equipped with stereo vision) and an environmental map (e.g., a map created by an autonomous robot system using a Simultaneous Localization And Mapping (SLAM) method [11]). If the data is stationary in terms of time, we can view it as data expressed by Eq. (1). For example, the shape data of a building can be seen as stationary in terms of time.

We can utilize a Bayesian network to create a model of causality among the observed attributes expressed by Eq. (1). The Bayesian network model is useful for not only developing the model by combining the observed data and external knowledge, but also for inferring and predicting behavior with novel targets. We can create a cross tabulation table by normalizing and quantizing the data set (Eq. 1). We can construct a Bayesian network model from this cross tabulation, utilizing several software packages for this purpose (e.g., BAYONET [12]). The Bayesian network model constructed from Eq. (1) can be expressed as a joint distribution in a state space $\mathbf{Z}$ by

$$p\left(\mathbf{Z}|Bs\right) = \prod_{i=1}^{n} p\left(Z_i|pa(Z_i), Bs\right), \tag{5}$$

where Bs denotes a probabilistic structure for the Bayesian network, $pa(Z_i)$ denotes a parent of Z_i, and n is the number of attributes. We can infer and predict desired attributes by using Eq. (5). For example, we can predict the desired attributes for one layer Z_d, given the others at location $\mathbf{s_0}$, using

$$p\left(Z_d(\mathbf{s_0})|\mathbf{Z}_{i\neq d}(\mathbf{s_o}), Bs\right) = \frac{p\left(\mathbf{Z}_i(\mathbf{s_o})|Bs\right)}{\sum\limits_{Z_{,}=} p\left(\mathbf{Z}_i(\mathbf{s_o})|Bs\right)} . \tag{6}$$

2.3 Advantages of STS Mapping System

The advantages of the STS mapping system are as follows: 1) both an expert and a layperson can see the relations among the behavior and environment, because most people are familiar with the representations based on an environmental coordinate system. 2) It is easy to integrate the sensory data and other data when these data are standardized by representations based on an environmental coordinate system. 3) It is possible to extract knowledge by using a modeling process and then applying this knowledge to a new target. 4) It is easy to intuitively confirm the results of the modeling and the retargeting by visualizing the results in an environmental coordinate system.

3 Application of STS Mapping System: Modeling Children's Behavior and Using Model to Design a New Product

3.1 Overview of Implemented STS Mapping System

We implemented the proposed STS mapping system to model children's behavior while playing on playground equipment. The realized system consisted of a wireless wearable location-EMG (L-EMG) sensor for conducting in situ observations and measurements of children playing, a system for representing the measured data based on an environmental coordinate system, and a Bayesian network for modeling and retargeting. Figure 2 shows the implemented system and the process flow. The details are described below.

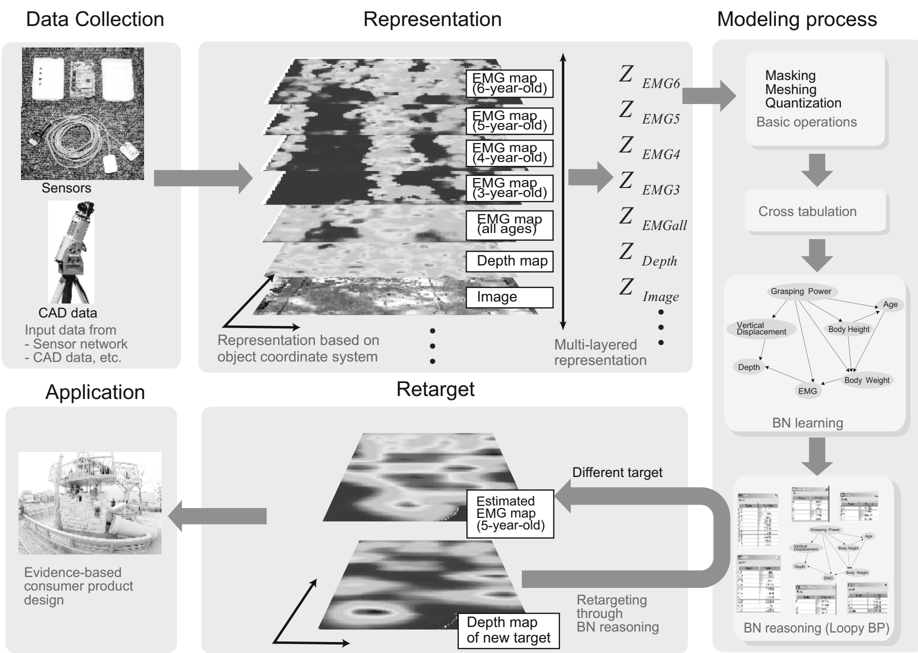


Fig. 2. Spatio-temporal semantic mapping system

3.2 Step 1: Spatio-temporal Extension of Child Behavior Measurement with Wearable Location-EMG

Development of Wearable EMG Sensor. We have developed a wearable EMG sensor to be used as a behaviormetric sensor for measuring children's physiological state [13,14]. We use the EMG sensor to measure the behavior of children for the following reasons. 1) Robustness: We can robustly obtain sensor signals related to the playing behavior because the EMG sensor can measure muscle activity. 2) Versatility: It is possible to use it to record other behavioral data; for example, it can be used to measure the electrooculogram (EOG) by placing it in a different position. The developed wireless type of EMG sensor has the following advantages. 1) The children's behavior is not disturbed because it was wearable. The measurement data was directly preserved in a personal computer (PC) via wireless communication. 2) As a result, we are able to secure sufficient storage capacity for long durational EMG measurement. 3) We developed an active electrode system that amplified the EMG signal near an electrode. We were able to begin quickly measuring EMG data because the active electrode could be easily attached to the body using an armband.

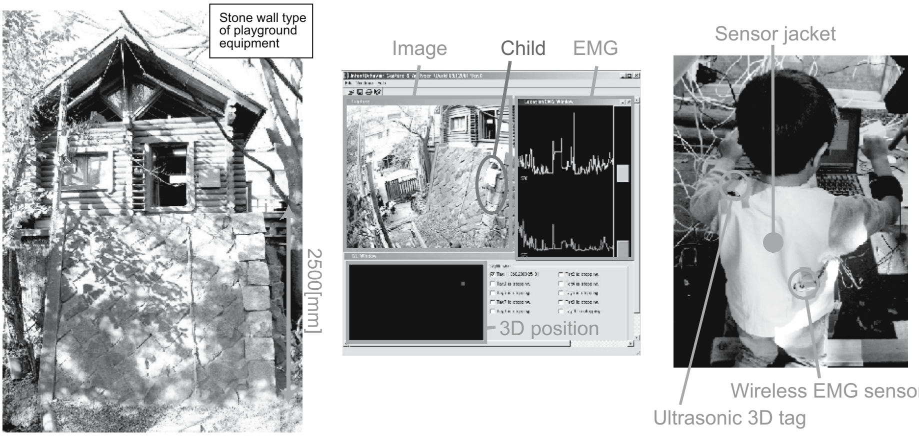
Spatio-temporal Extension of Wearable EMG Sensing. We have combined a wearable EMG sensor and an ultrasonic location sensor that was also developed by the authors [15] for spatio-temporal extension of the data from the wearable EMG sensor. This has enabled us to obtain EMG data that are spatio-temporally indexed. The ultrasonic location system consists of ultrasonic receivers, ultrasonic tags with a transmitter, and a radio controller. By attaching the ultrasonic tag to a child, we can detect and record the three-dimensional position data of the child. The ultrasonic location system can track the positions of the child within an error of 3 cm.

Observing Playing Children by Utilizing Location-EMG Sensor. We collected spatio-temporally indexed EMG data by measuring the children's playing behavior using the developed location-EMG (L-EMG) sensor in cooperation with the Kawawa nursery. Specifically, we measured children's behavior as they were climbing a stone wall, as shown in the picture on the left in Fig. 3. The L-EMG system consists of a section for recording a video image from a USB camera, a section for recording the three-dimensional position data from the ultrasonic location system, and a section for recording the EMG data. Thus, the three-dimensional position data, the video image, and the EMG signal are simultaneously measured using the L-EMG system.

The details concerning the experimental procedures are as follows. First, the electrodes were attached to the flexor digitorum superficialis muscle and the extensor digitorum of the right forearm of 47 toddlers (6 three-year-olds, 17 four-year-olds, 14 five-year-olds, and 10 six-year-olds) in the Kawawa nursery. Second, we prepared a sensor jacket in which the ultrasonic tag and the EMG sensor were embedded. Using the sensor jacket, the sensors could be easily attached to the body. The picture on the right in Fig. 3 shows a child wearing the sensor jacket and the one in the middle shows a snapshot of the data recorded by the L-EMG system software.

3.3 Step 2: Representing Spatio-temporal EMG Data and Spatial Depth Data from Stone Wall Using STS Mapping System

The EMG data measured by the L-EMG system was spatio-temporally indexed. The measured EMG data can be visualized in the stone wall coordinate system by



Loghouse with stone wall for climbing Integrated sensing system

Fig. 3. Stone wall type of playground equipment

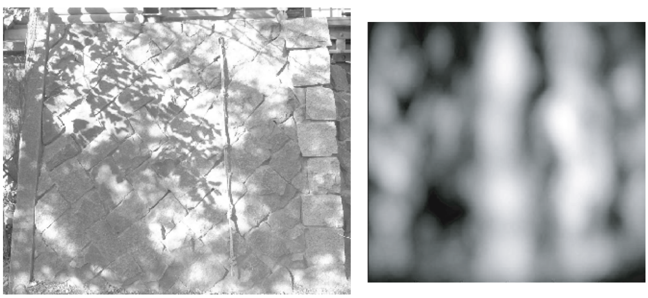


Fig. 4. EMG map that visualizes Location-EMG data

being input into the STS mapping system. Figure 4 shows an example of the measured EMG data visualized with respect to the stone wall coordinate system. This EMG map was made from the EMG data measured for all subjects aged between three and six years old. The parts in red indicate that a significant amount of muscle power was used. This figure helped us to confirm that the upper part of the wall required a significant amount of muscle power, which showed that it was difficult to climb the stone wall. We also measured the three-dimensional shape data of the stone wall using a laser scanner. We were also able to obtain the spatial data on the depth distribution using the measured shape data.

3.4 Step 3: Creating Bayesian Network Model

We conducted a mesh division for the EMG data, the depth map data, and the other data in the stone wall coordinate system. After normalizing and quantizing the mesh data, a cross tabulation table was constructed. The items in the table included the following attributes; age, weight, height, intensity of EMG, maximum grasping power, vertical displacement in playing, and depth of each block in the stone wall. The data sets that we obtained were expressed in terms of

$$Z_1(\mathbf{s},t) = Age(x,y,t), \quad Z_2(\mathbf{s},t) = BodyWeight(x,y,t),$$
$$Z_3(\mathbf{s},t) = BodyHeight(x,y,t), \quad Z_4(\mathbf{s},t) = EMG(x,y,t),$$
$$Z_5(\mathbf{s},t) = MaxPower(x,y,t), \quad Z_6(\mathbf{s},t) = Depth(x,y,t),$$
$$Z_7(\mathbf{s},t) = VerticalDisplacement(x,y,t)$$

A Bayesian network was created from the constructed cross tabulation table. We customized and used BAYONET [12] for this process. BAYONET finds probabilistic semantic structures via a kind of greedy algorithm based on one of several information criteria, such as Akaike's information criterion (AIC) [16], from the given data.

3.5 Step 4: Retargeting Bayesian Network Model: Application of Playing Behavior Model to Construct New Product Design

We used the model to create a new design for safer playground equipment. The model constructed in Step 3 above expresses the causal relation between age, weight, height, intensity of EMG, maximum grasping power, vertical displacement in playing, and the depth of each block in the stone wall. Falling from the higher part of the stone wall can cause more serious injury than falling from the lower part. The difficulty in climbing is strongly related to the risk of falling from the equipment. As stated above, the intensity of the EMG indicates the difficulty in climbing. So, we used the intensity of the EMG as a criterion for the risk of falling. Among the parameters in the product design, the depth and height of the stone wall blocks are controllable. We can infer the intensity by varying the controllable parameters and the target age as inputs to the constructed Bayesian network. More concretely, we can calculate the expectation of the EMG by utilizing Eqs. (7) and (8).

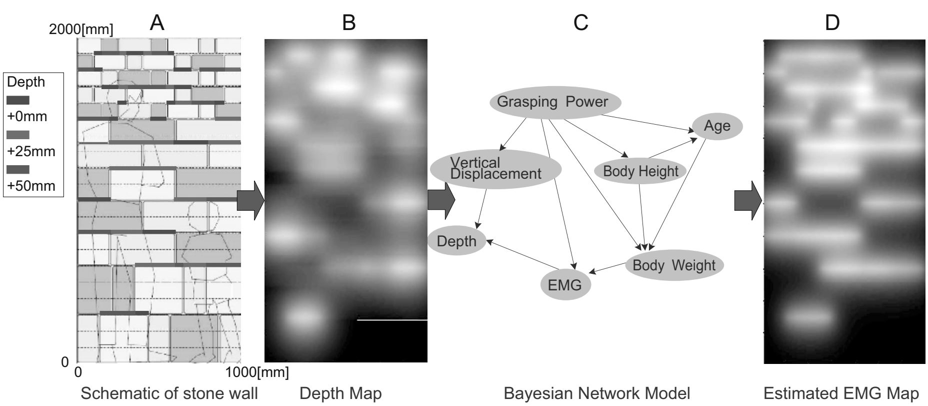


Fig. 5. Estimated EMG map using Bayesian network

$$p\left(Z_4(x,y)|\mathbf{Z}_{i\neq 4,7}(x,y),Bs\right)=\frac{p\left(\mathbf{Z}_i(x,y)|Bs\right)}{\sum\limits_{Z_i=4,7} p\left(\mathbf{Z}_i(x,y)|Bs\right)}, \tag{7}$$

$Z_1 = 3, 4, 5, 6$ *years old*, Z_2 = average body weight,
Z_3 = average body height, Z_5 = average maximum power,
and $Z_6(x, y)$ = depth map data from schematic, and

$$E(Z_4) == \sum_j A_j \times p_j\left(Z_4|\mathbf{Z}_{i\neq 4},Bs\right), \tag{8}$$

where A_j denotes the coefficient for calculating the expectation from the inferred probability distribution p_j. In general, the probability distribution inferred by a Bayesian network is a discrete distribution. Figure 5 shows an example of an estimation of the intensity of the EMG ($E(Z_4(x, y))$). In this figure, section A depicts a new schematic of the stone wall structure, section B depicts the depth map of the new schematic, section C depicts the Bayesian network model, and section D depicts the intensity of the EMG as inferred by the Bayesian network model. We used the customized BAYONET for this inference process. Figure 6 shows the results of the estimation of the EMG map for different ages. In the figure, the EMG intensity is normalized so that the maximum intensity of a three year-old becomes one. The figure shows that children can climb the wall using less strength, as they get older. By repeating the design of a new schematic and estimation of the EMG, we can interactively design a new stone wall type of playground equipment.

For this case study, we created the specifications listed below in cooperation with playground equipment makers.

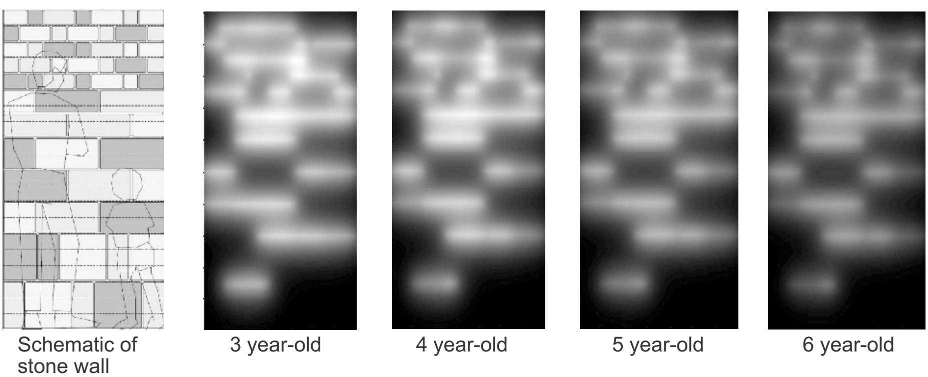


Fig. 6. Estimated EMG map of different ages (the EMG intensity is normalized so that the maximum intensity of a three year-old becomes one)

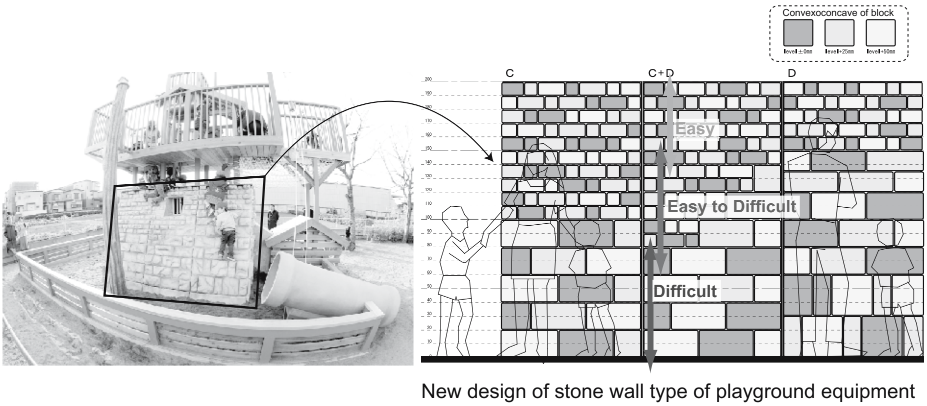


Fig. 7. Application of new type of playground equipment design

1. In order to exclude children that were 1 to 2 years of age, we designed the lower part of the stone wall to be too difficult for them to climb, i.e., so that they cannot exert the required degree of muscular power in this section.
2. To allow those children who can climb the lower section to enjoy the rest of their climb, we designed the middle part of the stone wall with a variety of difficulties.
3. To ensure safety in climbing and to help the children climb securely, we designed the upper part of the stone wall to be relatively easy for them to climb, i.e., not much strength is required.

We made some patterns for a wall that satisfied these specifications by repeating the trial design using the STS mapping system in collaboration with playground equipment designers. Figure 7 shows three examples of stone walls for playground equipment based on the created model and the constructed playground equipment based on the model.

4 Conclusion

This paper has highlighted the possibility of creating a data-driven type of an everyday life behavior model and using the model to improve everyday life. This stems from the recent development of sensing and modeling technologies. A wearable sensing technology and a statistical modeling technology make the application of science to everyday life situations feasible.

We have proposed the concept of a Spatio-temporal Semantic mapping system as one approach to dealing with knowledge acquisition based on real-life behavior data. The STS Mapping system consists of the following components: 1) a spatio-temporal extension of behaviormetric sensing with the integration of a non-location behaviormetric sensor and a location sensor, 2) a standardized and multilayered representation of the measured data based on an environmental coordinate system, and 3) a statistical modeling process for knowledge acquisition from the represented data, and a retargeting process. This paper has described the formulation of behavior phenomena in terms of a spatial point process as dealt with in spatial statistics and a concrete computation method for creating a model using the STS mapping system.

To show the effectiveness of the proposed STS mapping system, this paper reported on an implemented system and a case study using it. In the case study, we conducted in situ observations and measurements of 47 children playing with or on playground equipment by using a wireless wearable location-electromyography sensor and constructed a model of the children's behavior from the measured data. This paper also reported on a new play equipment design that had a climbing section that was suitable for the children's target age group. The new design was created using the constructed model in collaboration with playground equipment designers. The case study showed that everyday life behavior science could possibily be applied to evidence-based product design as well as indicating the effectiveness of the proposed system from a practical standpoint.

Creation of a new design as described in this paper is our first trial towards developing safer playground equipment for children. This year, our research group constructed a system for monitoring children playing in the nursery with the new equipment that has sensors installed in it. We hope that, by collecting such everyday data over a long term, we will be in a better position to clarify the relationships between injuries and children's behavior through the continuous improvement of the model and the equipment. We believe that incremental knowledge development, which means continuous improvement and application of knowledge using real data for feedback, is very important for making knowledge really useful in our society.

References

1. Yang, G. (ed.): Body Sensor Networks. Springer, London (2005)
2. Higbtower, J., Borriello, G.: Location systems for ubiquitous computing. IEEE Computer 34(8), 57–66 (2001)
3. Parkinson, B., Spilker Jr, J. (eds.): Global Positioning System: Theory and Applications, vol.2, American Institute of Aeronautics and Astronautics, Inc. (1996)
4. Heywood, I., Cornelius, S., Carver, S.: Geographical Information Systems, 3rd edn., Pearson Education Limited (2006)
5. Cressie, N.: Statistics for Spatial Data. Revised edition, John Wiley & Sons, Inc., England (1993)
6. Stoyan, D., Kendall, W., Mecke, J.: Stochastic Geometry and its Applications, 2nd edn. John Wiley & Sons, Inc., England (1996)
7. Neapolitan, R. (ed.): Learning Bayesian Networks. Pearson Education, Inc. (2004)
8. Wallis, A., Cody, B., Mickalide, A.: Report to the nation: Trends in unintentional childhood injury mortality, 1987-2000 (2003), http://www.usa.safekids.org/content_documents/nskw03_report.pdf

9. WHO (ed.): Child and Adolescent Injury Prevention —A WHO Plan of Action (2006)
10. Nishida, Y., et al.: Methodology of everyday life computing and application to children injury prevention. In: FOCI 2007. The 1st IEEE Symposium on Foundations of Computational Intelligence, pp. 652–659 (2007)
11. Thrun, S., Burgard, W., Fox, D.: Probablistic Robotics. Oxford Universuty Press, Oxford (2005)
12. Motomura, Y.: Bayonet: Bayesian network on neural network. Foundation of Real-World Intelligence, 28–37 (2001)
13. Nishida, Y., Kawakami, G., Mizoguchi, H.: Everyday grasping behavior measurement with wearable electromyography. In: Sensors2006. The 5th IEEE Conference on Sensors, pp. 988–991 (2006)
14. Kawakami, G., Nishida, Y., Mizoguchi, H.: In situ measurement of playing children by wireless wearable electromyography. In: Sensors2007. The 6th IEEE Conference on Sensors, pp. 993–996 (2007)
15. Nishida, Y., et al.: 3D ultrasonic tagging system for observing human activity. In: IROS 2003. IEEE International Conference on Intelligent Robots and Systems, pp. 785–791 (2003)
16. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: 2nd International Symposium on Information Theory, pp. 267–281 (1973)

Part II

Logic and Engineering of Natural Language Semantics

Overview of Logic and Engineering of Natural Language Semantics (LENLS) 2007

Kei Yoshimoto

Graduate School of International Cultural Studies, Tohoku University
Kawauchi 41, Aoba-ku, Sendai, 980-8576 Japan
kyoshimoto@mail.tains.tohoku.ac.jp

LENLS 2007 was held at World Convention Center Summit (Phoenix Seagaia Resort), Miayzaki, Japan on June 18 and 19, 2007 as one of the international workshops collocated with the 21st Annual Conference of the Japanese Society for Artificial Intelligence (JSAI 2007). This year the workshop was held for the fourth time, with its predecessors taking place annually. Since the time it started in 2004, LENLS has developed both in its scale and in the quality of presentations into an academic meeting unique in Japan and the surrounding area at which participants can exchange innovative ideas about dynamic semantics, its related fields, and the application of the theories.

LENLS 2007 had a special topic 'Dynamic approaches to semantics, syntax, and pragmatics.' In the past few decades, Dynamic Semantics (DSem) and its analogue Discourse Representation Theory (DRT) have been developed to elucidate how a sentence is interpreted in relation to the context and how the sentence meaning updates the context. On the other hand, more and more attention has been paid recently on investigating syntactic theories which enable parsing of sentences in the same manner as humans do. To take an instance, Dynamic Syntax (DSyn) advocated by Ruth Kempson and her colleagues proposes a theoretical framework which builds up partial trees incrementally following the left-to-right sequence of sentence processing. It is important to note that these two significant trends in formal linguistics are complementary to each other: while DSem or DRT has mainly shown little interest in parsing or interpreting each sentence incrementally or compositionally, in DSyn it remains unclear how contextual information is referred to during sentence analysis/interpretation and conversely how the output logical formula contributes to context updating. It was therefore of theoretical importance that in the workshop linguistic researchers shared and exchanged information under the same rubric 'dynamic approaches.' Furthermore, LENLS 2007 offered two invited lectures by Kōiti Hasida and Gerhard Jäger on the application of Game Theory to linguistics, another field of research which is shedding a new light on dynamic aspects of dialogue. It is also noteworthy that at the workshop Yoshiki Mori gave an invited lecture titled 'ATM once more dynamically approached', which is not published in this post-proceedings.

This post-proceedings contains eight papers related to dynamic approaches. While some of them follow the Dynamic Semantics framework, others adopt the continuation approach, Scope Control Theory, and syntactic theories such as

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 79–80, 2008.

Dynamic Syntax and categorial proof net. Elin McCready compares existing formal semantic theories including his own based on modal subordination to evaluate the extent to which evidential phenomena in various languages can be accounted for. Tomoyuki Yamada proposes an extension of the logic of acts of commanding in order to model acts of promising together with those of commanding. For this purpose, changes in deontic status of relevant action alternatives are modeled as deontic updators. Norihiro Ogata has developed a generalized QG-semantics of Quantified Modal Logics (QMLs) to solve the problem of Kripke-incompleteness of QMLs and proposes a dynamic semantics of a quantified modal mu-calculus. There are two contributions from the 'continuation' perspective. Chung-chieh Shan models operational semantic interpretation and applies the notion of *context* and *order* to explanation of inverse-scope quantifiers and their polarity sensitivity. Rui Otake and Kei Yoshimoto exploit a continuation mode in multimodal type logical grammar to give an analysis of the interaction of scrambling and quantifier scope and focus and split-QP constructions. In his paper first presented as an invited lecture, Alastair Butler develops Scope Control Theory he advocates by giving a uniform formalism to subordinating and coordinating pronominal binding relations. One of the syntactic approaches is taken by Hiroaki Nakamura, who applies the categorial proof net to incremental parsing of the topicalized sentence construction in Japanese including sentence-internal topics, cleft-constructions, and multiple topic sentences. Masahiro Kobayashi proposes an algorithm to optimize the application of transition rules in Dynamic Syntax by dividing parsing states into distinct partitions to which different rules are applied.

This proceedings also contains papers with miscellaneous interests. Emar Maier argues for compositional interpretation of expressions with quotation based on a hybrid use-mention account. Mana Kobuchi-Philip, attempting to solve the difficulties with the analysis of the *mo*-phrase in Japanese as a generalized quantifier, presents an analysis as a modifier and applies it to cases in which the particle combines with a PP and IP. Koji Mineshima advances a proof-theoretic analysis of presuppositions involving definite descriptions based on the natural deduction system of ϵ-calculus and construction type theory. Lastly, Kōiti Hasida, Shun Shiramatsu, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno propose the *meaning game*, a formalization of intentional communications based on game theory, and demonstrate that centering theory is derived from a meaning game.

The Organizing Committee of LENLS 2007 was constituted by Elin McCready, Yasuo Nakayama, Norihiro Ogata, Atsushi Shimojima, Satoru Suzuki, Katsuhiko Yabushita, Tomoyuki Yamada, and Kei Yoshimoto. Besides the committee members, the submissions were reviewed by the following reviewers, for whom full credit goes: Daisuke Bekki, Alastair Butler, Yurie Hara, Gerhard Jäger, Masahiro Kobayashi, Yoshiki Mori, Chidori Nakamura, Hiroaki Nakamura, Rick Nouwen, Rui Otake, Chris Potts, Brian Reese, Rolf Schwitter, and Anders Søgaard. Also I would like to acknowledge the organizational and financial assistance of The Japan Society for Artificial Intelligence, especially Professor Akihiro Inokuchi, who was the chief organizer of the international workshops.

Semantic Heterogeneity in Evidentials

Elin McCready

Department of English, Aoyama Gakuin University
mccready@cl.aoyama.ac.jp

Abstract. In this paper I will examine some data relating to evidential systems in various languages and existing formal semantic analyses of evidential phenomena, with an eye to determining the extent to which available systems are capable of accounting for the full range of facts. Data will be drawn largely from [1] and from the formal semantic works I'll discuss.

1 Introduction

Evidentiality is currently a topic of much interest in formal semantics. As far as I know, only four formal accounts are on the market at the present time. The first involves presuppositions([2,3,4]. The second is the speech-act based analysis of the Cuzco Quechua evidential system of [5]. This system is typical in some respects of the evidential systems of languages with 'grammatical' evidentiality, evidentiality which is required to appear in every sentence or clause (though this claim is in fact much too strong, as even in such languages evidentials can often be left out. Aikhenvald takes the easy out by claiming the presence of an evidential morpheme with a 'null realization;' this claim is disputed by many, probably for good reason). Faller has recently extended this system with a complex analysis making use of SDRT discourse relations. The fourth system is that of [6], which involves changes in Gricean quality thresholds, formulated using probability theory. The last system is that of [7], who analyze Japanese evidentials as as special kind of source-indicating modality.

The next section will provide the basics of these three analyses and compare them first in terms of the formalism, second in terms of the predictions they make about e.g. the scope of evidentials, and third with respect to the motivations behind the particular choices made. As we'll see, each is well motivated *for the language it is intended to analyze*. But—perhaps surprisingly—evidential systems vary a great deal across the world's languages, enough that it isn't obvious whether any single system can account for all the facts involving them. Section 3 will give an overview of some facts that look problematic for one or more of the theories outlined in section 2 (omitting facts already discussed in that section in reference to motivations for one theory or another). I will there also discuss what I take to be some possible solutions, and also some other possible formal theories of evidentiality. It should be stressed that the data I present is drawn largely from [1]; the approach I take to the data and the semantic facts in these languages are also drawn from this source, which is usually brief both for

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 81–94, 2008.

reasons of the wide scope of the work and because of its lack of formal semantic orientation, a feature of almost all work on evidentiality. This is a lack that must be overcome before definitive conclusions can be made. Still, in section 4, I will give some tentative conclusions, one of which is that it will not be easy to find a semantic theory of evidentiality that is capable of accounting for the full range of data across languages.

2 Semantic Theories of Evidentials

Here I briefly introduce the four theories of evidentiality. As we will see, the data that motivated each theory differ quite broadly, which results in very different looking theories. We'll also see that a consequence of this difference in base data is that none of the theories do very well in explaining the data that motivates the others.

2.1 Presupposition

The reason for treating evidential content as presuppositional was (in the McCready-Asher case, which I concentrate on here) that, in the case of certain Japanese modals that seem to have evidential content, this content can in fact be bound in a conditional antecedent. The particular modal for which this seems to hold is *nichigainai*. Consider the following examples from [3]. The first translates the famous Roberts modal subordination example in (1).

(1) A wolf might come in. It would eat you first. [8]

(2) ookami-ga kuru kamoshirenai. # ∅ anata-o taberu nichigainai.
wolf-NOM come might ∅ you-ACC eat surely
'A wolf$_i$ might come in. It$_i$ would eat you first.'

As we see here, the Japanese translation is infelictious. [3] argues that the reason for this is that the necessity modal *nichigainai* has evidential content that is not satisfied in this context. In particular, φ in the scope of *nichigainai* must follow by inference from contextually supplied evidence (see the above-cited works for arguments to this effect). However, as it turns out, the evidential content need not be independently satisfied in other contexts, specifically those provided by the addition of discourse particles to the second sentence or by putting the second sentence in a conditional (as consequent).

(3) ookami-ga kuru kamoshirenai. moshi ∅ kitara ∅ anata-o taberu
wolf-NOM come might if ∅ came-COND ∅ you-ACC eat
nichigainai.
surely
'A wolf$_i$ might come in. If (one) did, it$_i$ would eat you.'

What the above example shows is that, given the right context, *nichigainai* can be used felicitously even when the discourse (or situation of utterance) doesn't provide the right evidence to support inference. If it is indeed correct that the content of *nichigainai* is partly evidential (which it does seem to be), then any theory which forces evidential content to always be independently satisfied (in present terms, to scope over other operators) would be too strong to provide the correct analysis. Presupposition seems an ideal tool for explaining the *nichigainai* facts. If content in a conditional antecedent is sufficient to satisfy a given presupposition, the presupposed content does not project, which is precisely what is needed here. But the Japanese pure evidentials can embed quite generally, which indicates that this approach is wrong for them.

However, we also see cases of evidentiality in which the evidential content does, necessarily, project. The particular case considered in the works above was that of *hazu*, another Japanese modal which McCready and Asher argued to have evidential content (a viewpoint also espoused by [9] (127-129)). The example in (4) shows that the evidential meaning cannot be satisfied in conditionals (or also by the presence of a discourse particle).[1]

(4) ookami-ga kuru kamoshirenai. # moshi ∅ kitara ∅ anata-o taberu
wolf-NOM come might if ∅ came-COND ∅ you-ACC eat
hazu da.
surely COP
'A wolf$_i$ might come in. If (one) did, it$_i$ would eat you.'

The previous work mentioned assumed the binding theory of presupposition [10,11], in particular the version used in the theory of discourse structure SDRT ([12]; see [13] for the basics of DRT, on which SDRT is based). Within this framework, the meanings of *nichigainai* and *hazu* were handled by separating them into asserted (modal) and presupposed (evidential) content. Detailed discussion can be found in [4,7]; but simplifying, the account involved assigning the bindable evidential content 'normal' presuppositions, and the unbindable content presuppositions which were required to be externally anchored, meaning that binding in conditionals was impossible.

As it turns out, at least one of the Japanese pure evidentials, the hearsay evidential *soo-da*, cannot be bound either. Consider the following example involving a conditional, in which the antecedent makes the event of communicating the propositional content in the scope of *soo-da* explicit.

(5) # moshi dare-ka-ga [konya Jon-ga kuru] to osiete-kure-tara
if who-∃-Nom [tonight John-Nom come] COMP tell-give-COND
konya Jon-ga kuru soo-da
tonight John-Nom come SOO-DA
'If someone tells me John will come tonight, then John will come tonight (I heard).'

[1] It should be noted that there is a great deal of speaker variation on examples like these. Most people find them all terrible, but some allow them (at least marginally). Some speculations as to why this may be so can be found in [3] and [4].

This means that *soo-da* must be formalized using the same mechanisms used for *hazu*, if an analysis like this is adopted for the pure evidentials.

In the case of the Japanese hearsay evidential *rashii* (similar to *soo-da* in its behavior) and *soo-da*, it is not necessary to make use of a modal component. We need a more direct sort of assertion. What we want is to presuppose an externally anchored event of communication. The same is true for *soo-da*. We also need to ensure that ϕ is not asserted, since the speaker need not subscribe to any beliefs concerning the the proposition (see [7] for details). This can be done by introducing a special sort of hearsay evidence, and presupposing that an externally anchored hearsay event exist.

2.2 Speech Acts

Let's look next at the theory of [5], who did extensive work on the evidential system of Cuzco Quechua. Cuzco Quechua has several enclitic suffixes that mark evidentiality or the nature of the speaker's justification for making the claim. Faller analyzes three suffixes in detail: *-mi*, which indicates that the speaker has direct (usually perceptual) evidence for the claim, *-si*, which indicates that the speaker heard the information expressed in the claim from someone else, and -*chá*, which indicates that the speaker's background knowledge, plus inferencing, leads him to believe the information in the claim true.

Faller uses Vanderveken's (1990) speech act theory for her analysis. Vanderveken's theory assigns speech acts preconditions for successful performance. Faller takes evidentials to introduce additional content into the set of preconditions. Vanderveken's analysis makes use of three elements: propositional content, illocutionary force, and sincerity conditions on successful performance of the SA, e.g. that $Bel(s, p)$ holds in the case of assertions.

In large part, the focus of Faller's analysis of *-mi* and *chá* is on the sincerity conditions for the assertion. Essentially, *-mi* adds an additional sincerity condition to the assertion, that $Bpg(s, p)$. The formula $Bpg(s, p)$ means that the speaker has the best possible grounds for believing p. Faller does not attempt to make this notion completely precise, noting only that it is dependent on the content in the scope of *-mi*: for externally visible events Bpg will ordinarily be sensory evidence, while for reports of people's intentions or attitudes even hearsay evidence will often be enough.

Faller analyzes *-chá* as being simultaneously modal and evidential. As a result, the asserted propositional content p is mapped to $\Diamond p$, as is the corresponding belief object $Bel(s, p)$ in SINC. The condition $Rea(s, Bel(s, \Diamond p))$ is also added to SINC. $Rea(s, Bel(s, \Diamond p))$ indicates that the speaker's belief in the possibility of p follows from his own reasoning/inference. *-Si* is also complex; the propositional content p is not asserted when this hearsay evidential is used (as is also the case in Japanese), as we saw, which means that the propositional content of the utterance cannot be asserted. Faller posits a special speech act PRESENT for this situation, on which the speaker simply presents a proposition without making claims about its truth. Therefore $Bel(s, p)$ is eliminated from SINC, and the condition $\exists s_2[Assert(s_2, p) \wedge s_2 \notin \{h, s\}]$ is added to the set of sincerity conditions.

The reason for using a speech act-based analysis is that the Cuzco Quechua evidentials do not embed semantically. Even if they appear under negation or in a conditional consequent, their content cannot be 'bound' by some content in the antecedent; this shows that they are not standard presuppositions. Faller provides examples with negation for each of the three evidentials:

(6)	Ines-qa	**mana-n/-chá/-s**	qaynunchaw	ñaña-n-ta-chu	watuku-rqa-n
	Ines-Top	not-MI/CHÁ/SI	yesterday	sister-3-Acc-CHU	visit-Pst1-3

'Ines didn't visit her sister yesterday.' (and speaker has evidence for this) NOT 'Ines visited her sister yesterday' (and speaker doesn't have evidence for this)

This solution doesn't allow for embedding of evidentials, which makes sense given that it is designed precisely to account for evidentials in a language which disallows such embeddings. But making this possibility available is necessary for Japanese, as the next subsection will show. So, although Faller's analysis may be right for Quechua, it doesn't extend easily to the Japanese case. This may not be something we should take as a shortcoming of the Faller system: as noted by [15], Japanese is a marginal case with respect to evidentiality in that the evidential report is not obligatory, unlike most languages that are studied intensively in this area (or so the orthodoxy has it, though [16] disputes this conclusion). However, section 3 will show that similar problems arise for more typical evidential systems as well. The same point applies to another possible analysis of evidentials, one in terms of expressive content ([17,18]). Potts explicates the content of expletives and appositives, among other constructions, in terms of conventional implicature: pragmatic aspects of meaning that never embed semantically, and are associated with particular lexical items. Many evidentials seem to be prime examples of such forms. I will discuss this kind of account more below.

2.3 Manipulating Quality Thresholds

[6] analyze evidentials of the Quechua type as altering the quality threshold for assertion. The analysis is set within a larger picture cashing out the Gricean maxims in probabilistic terms. The Maxim of Quality is handled by setting a quality threshold based on probability distributions. The idea is that each proposition is associated with a certain probability, and that the context sets a degree of probability below which one cannot felicitously assert a proposition. For instance, if the context requires a 'degree of belief' of 0.8, but one thinks that φ has only a 50% chance of being true, it will be unassertable.

Within this general picture, evidentials are taken to manipulate the quality threshold itself, changing it to some point in [0,1] in accordance with the reliability that speakers assign to the particular evidence source speakers make use of in their assertion. For example, if I have little faith in hearsay, my utterance of $Hearsay(\varphi)$ might indicate that I assert φ using a quality threshold of 0.2.

This analysis has many interesting points and also raises a number of questions. I cannot address these issues here. For the purposes of this paper, which is intended simply to see whether evidential content is unified by means of checking formal accounts that have been given for different types of evidentials, we can treat it in a way similar to the Faller approach, given that both effectively involve speech acts (since Potts *et al.*'s analysis is stated in Gricean terms). This would indicate that embedding is not possible, as far as I can see, unless other speech acts are involved.

2.4 Probabilistic Modality

The last approach I sketch is that of [7]. These authors treat evidentiality in Japanese, analyzing the evidentials there using a system of probabilistic dynamic logic. The evidentials are of two types: a range of inferential evidentials and two hearsay evidentials. Here I will not provide the basic examples but show instead the data that motivated treating them in the way they did.[2]

The inferential evidentials were analyzed as a kind of probabilistic modal (an informal version of the relevant definition is provided below). The hearsay evidentials, on the other hand, were analyzed (within the same system) as just testing for the existence of a past event of hearsay with content φ, as usual. The inferential evidentials were treated as modals essentially for two reasons, or rather three if one considers basic truth conditions. One is that they enable modal subordination, both with modals in subordinate position, and with other evidentials, as shown by the following two examples respectively.

(7) a.

ookami-ga	kuru	mitai	da
wolf-Nom	come	MITAI	Cop.Pres

‘A wolf will come in, it seems.’

b.

anta-o	taberu	kamoshirenai
you-Acc	eat	might

‘It might eat you.’

(8) a.

ookami-ga	kita	mitai/yoo	da
wolf-Nom	came	MITAI/YOO	Cop.Pres

‘A wolf/Some wolves has/have come, it seems.’[3]

b.

yatsu(ra)-wa	totemo	onaka-o	sukaseteiru	mitai/yoo	da
it(they)-Top	very	stomach-Acc	emptied	MITAI/YOO	Cop.Pres

‘It/they seems/seem to be very hungry.’

One cannot replicate these examples with the hearsay evidentials though, which shows that they lack the right kind of modal content. I should note that it’s

[2] It is interesting to note though that Ladakhi seems to have a similar system (Aikhenvald 2004: 53).

[3] The subject here can be interpreted as either singular or plural, as usual with bare nominals in languages like Japanese.

not clear how Japanese might differ from Quechua here; data on Quechua modal subordination, with or without evidentials, is not currently available.

The second reason is that Japanese evidentials behave quite differently from Quechua ones in terms of scope. They can quite easily take scope under other semantic operators, such as conditionals. This is so for both inferential and hearsay evidentials; here I just give one case with an inferential.

(9) Taro-ga kuru yoo da-ttara osiete kudasai
Taro-Nom come YOO Cop.Pres-COND teach please
'If it looks like Taro will come, please tell me.'

Note that although these evidentials can appear in the scope of other operators, they cannot be *bound*, as already shown for *soo-da* above. Again, this indicates that presupposition is not the right way to think about these facts. Evidentials can also embed under modals and negation, which makes things even worse; space precludes giving examples here, but see [7].

The inferential modals were analyzed using an operator $\triangle^i$, where i indexes an evidence source. Informally this was given the following semantics.

(10) $\triangle^i\phi$ is true given a world w, time s, and probability function μ iff:
a. ϕ was less likely at some time preceding s (before introduction of some piece of evidence i);
b. ϕ is still not completely certain at s (given i);
c. the probability of ϕ never decreased between the time the speaker became aware of the evidence i and s as a result of the same piece of evidence i (i.e., the probability of ϕ given i is upward monotonic).

The hearsay evidential on the other hand was modelled with an operator H_a, which is understood as follows.

(11) $\mathsf{H}_a\varphi$ indicates that a has experienced an event of acquiring hearsay knowledge $\mathsf{E}^h_a\varphi$, at some past time.

Given all this it is easy to see how the data is accounted for. Evidentials simply introduce semantic operators, which can scope over and under other bits of content as usual, unlike what happens if they're tied to speech acts or presuppositions.

3 Problematic Data

Here we examine some data that poses problems for one or more theory. The problematic data falls into several classes. The first class involves cases where evidential content interacts with other elements of the compositional semantics: this includes situations where the evidential content appears to be bound within the sentence, and also cases in which evidential content applies only to one part of the sentence, not to the entire thing, which I will not discuss in this paper. These cases are discussed in 3.1. The second type are cases in which we find interactions between different pieces of evidential content. These can be separated into two subclasses: cases in which only evidential content interacts, and cases in which a piece of evidential content acts on a combination of evidential and (other) propositional content.[4] These are discussed in 3.2.

3.1 Interaction with Other Elements

Here is an instance in which evidential content seems to be bound by a restrictor within the sentence proper. This is from a Northeast Caucasian language called Lezgian.[5] According to Aikhenvald, 'this specifies the content of the report; it is the smart people who say that knowing too much is harmful (Aikhenvald 2004: 32).

(12) Gzaf cir xu-n, aq'ullu insan-r.i
much know ANTIC-MSD smart person-PL(ERG)
luhu-zwa-j-wal, zarar ja-lda
say-IMPF-PART-MAN harm COP-REP
'As smart people say, knowing too much is harmful.'

Faller cannot account for this; for her, evidential content cannot interact at all with 'propositional' content. For the presuppositional accounts it depends on what one takes the logical form of the example to be. With the right logical form (one in which two logical clauses were introduced, the evidential presupposition could be bound. The modal-based account gets this straightforwardly on the assumption that the restrictive clause introduces a separate clause in logical form which has an index of E type, yielding a donkey sentence-like effect.

We have already seen that in Japanese evidential content can take scope under other operators. Although it may have been thought that this property was a result of particular properties of the (slightly marginal) Japanese evidential system, in fact it shows up elsewhere as well, even in cases where evidentials are supposed to be obligatory. Consider the following example from Warlpiri (Aikhenvald 2004:97).

(13) Ngana-nku nganta paka-rnu
who-ERG REP hit-PAST
'Who do they say hit him/her?'

One can answer this as follows (on one reading anyway).

(14) ngana-nku mayi nganta paka-rnu
who-ERG don't.know REP hit-PAST
'I do not know who they reckon hit her.'

[4] It is somewhat hard to phrase this without prejudging the issue of what kind of content evidentials have. Still I take it that my intention is clear here.

[5] The diacritics on the examples are impressionistic. See Aikhenvald for a more accurate version.

So both the negation and the modal verb in the answer are able to scope over the evidential. This is an example that—obviously—the modal approach can get—after all, it was designed to. And, not unexpectedly, neither of the approaches that the McCready-Ogata approach was responding to can handle this example without modifications.

3.2 Interaction between Evidentials

Now let's consider cases in which evidentials interact. I start with this example, from Shipibo-Konibo (Aikhenvald 2004: 55). This sentence can be used 'if one is watching a soccer match on TV and sees that a player suddenly falls to the ground and others come to his help'.

(15) oa-ra taské-bira-ke
DIST:ABS-DIR.EV sprain-INFR-COMPL
'He must have sprained his ankle' (Aikhenvald's gloss)

Actually this gloss does not seem to be quite right as it does not reflect the contribution of the direct evidential on the first word. What it really seems to mean, given the scenario, is that the player must have sprained his ankle *given what I saw*. This means that it has got to have a logical form like this.

(16) $Dir(Must(break - ankle(sp)))$

The presuppositional theory cannot even begin to get this right. It predicts that, if there is more than one piece of evidential content, each piece should be completely independent. The modal-based theory can analyze these examples without much trouble. The speech act theory can get these, quite easily; what it must say is that first the inferential modifies the ordinary assertion, and then the direct evidential modifies the resulting speech act. Again something must be said about evidential scope, but this should not be impossible and in any case must be done on any account.

Another, somewhat similar example. Here we find the hearsay evidential together with the direct evidential. What's interesting here is that the hearsay evidential scopes over both the content of the report and the direct and inferential evidentials. These appear to be interpreted conjunctively, as in (18): $Hearsay(Inf\varphi \wedge Phys.evid\varphi)$. The example is from Tsafaki.

(17) Manuel ano fi-nu-ti-e
Manuel food eat-INF-PHYS.EVID-HEARSAY-DECL
'He said/they say Manuel has eaten (they didn't see him, but they have direct physical evidence, i.e. dirty dishes) (Aikhenvald 2004:84)

Presuppositional stories can't handle this either. However the McCready-Ogata account can get this, for on this account evidentials are just modallike operators. Some assumptions need to be made about scope here though, of course. And, as before, the speech act account can get these too.[6]

[6] This example is problematic for possible conventional implicature based theories, as they predict no interaction between evidentials.

Here is an instance of an example that appears to be understood as a kind of conjunction.[7] Here the idea is that we get an interpretation like 'as I had guessed and can pretty well confirm.'

(18) oh, the: zbe zete-k-u
oh 3sg drum beat-INFR-VIS
'Oh, he WAS playing a drum!'

This is just a conjuction: I had guessed p and at this point I can confirm that p. This amounts to $infr(p) \wedge direct(p)$, and possibly also p (though this last may follow pragmatically from the use of the direct evidential). The presuppositional account may be able to get this right, for the presuppositions will both just scope out; though some additional assumptions about the interaction of tense and evidentials must be made; it's not clear however how the modal account can get this example, though, as it requires 'double access' to p, the proposition in the scope of the evidentials. Only conventional implicature (CI) based accounts can get this easily.

This example points up an interesting feature of the CI account that differentiates it from its speech act-based cousin (precursor). On the Potts CI logic, 'ordinary' content to which CI content applies is passed up without change. It makes use of two kinds of content, at-issue and conventionally implicated content, and that the two types are distinguished in the syntax of the logic, more precisely in the type definitions. At-issue content is distinguished from its CI counterpart with a superscript 'a' on the types; CI content is marked with a 'c'. CI content further has the special characteristic that it can only appear as the output member of objects of type $\langle \sigma, t \rangle$, σ any type. So we have types of the form $\langle \sigma, t^c \rangle$, but no types of the form $\langle \sigma^c, \tau \rangle$ for any σ or τ. This means that CI content a) always gets 'widest scope', i.e. it's scopeless, and b) is always associated with a proposition. So we always get a situation like this one.

(19) $\lambda p_t \ .[Op(p)^c](p) = \langle p, Op(p) \rangle$

Here the two elements of the ordered pair $\langle p, Op(p) \rangle$ are in different dimensions of meaning (as Potts assumes a two-dimensional semantics; see the works cited above for details of the system). Note though that p, despite having been modified by the CI operator Op_c, remains present in logical form; in fact it is unaltered by the composition. In effect the operator has applied to a 'copy' of p. What this means is that the CI logic is able to replicate resources in the sense of [19], for example. And this means that it has no difficulty in handling examples like this one, for after modification by each bit of evidential content, the original proposition remains, available for further modification. As long as each modification is compatible with the next, all will be well.

On the other hand, this example looks bad for the speech act account. Can we really have a speech act that simultaneously has distinct sets of sincerity conditions? The issue is that we aren't getting modification of a modified speech act,

[7] Again the transcription is imperfect.

but two distinct modifications of the original assertion (at least so it seems from the gloss; let me stress again that more investigation is needed). We could think of this along the lines of [20] and take sentences with multiple evidential marking to consititute dot-object speech acts—but even in this work it is not assumed that both aspects of the dot object can be selected simultaneously—doing so would result in a zeugma and consequent infelicity. It looks like the CI analysis performs better here. But the CI analysis is not perfect either, as I show in the next section.

4 Discussion

I close the paper with a brief analysis of the scope of each evidential theory; the range of languages to which it seems reasonable to apply it, and the kind of data it can and cannot account for. Finally, I indicate what seems to be needed for the analysis of (some of) the problematic data discussed in the previous section.

Let's start with the presuppositional story. As far as I can see it is applicable only to cases in which evidential content doesn't interact with any other operators; this includes instances in which it is simply independent of 'standard' operators like negation and so forth (which I will call 'semantic operators'), and cases in which we find multiple evidentials that don't interact. Unfortunately it does not seem likely that it can be extended to handle languages in which evidential content interacts with other evidential content or with other operators. I conclude that this account is the least likely to yield a full story about evidentiality: it simply isn't flexible enough, and the commonalities between evidentials and presuppositions are too few.

How about the speech act account? I will separate it from the CI account in what follows. The speech act account shows promise: it can account for languages in which evidential content doesn't scope under semantic operators, and cases in which evidential content modifies other evidential content. But it cannot account for languages in which evidentials can scope under semantic operators, or for languages in which multiple evidentials can appear without interacting. One might try to say that e.g. modals and conditionals are speech act modifiers; but then what about negation? Surely it cannot modify speech acts. In any case, if one were to say this, the advantage of using speech acts goes away, for then there is no reason evidentials can't scope under semantic operators anymore (since they aren't any longer really semantic operators). I don't think that this account can give a totally general story either.

The CI account on the other hand handles cases in which multiple evidentials do not modify each other, and cases in which they scope out. However it cannot begin to analyze cases in which evidentials scope under semantic operators; this is the defining feature of such an account.

We are left with the modal account. This one does well with cases where evidentials can embed under semantic operators and cases in which they modify each other. The modal account by itself cannot force evidentials to scope out or handle cases in which multiple evidentials don't modify each other. However, notice that these cases are precisely those the CI account can get. It seems therefore that starting with the modal account and augmenting it with conventional implicatures *in some languages* might be just what we need to get a complete picture. The idea would be that there is a parametric difference: in some languages evidential content is simply modal (though concurrently requiring the presence of evidence), and in others this modal meaning is also conventionally implicated, which causes it to take wide scope and further allow 'double-access' readings. In the last bit of this paper I would like to spell this idea out a bit and show what it can do and where it runs into problems. As we'll see the situation isn't really very straightforward, and, in fact, I view the attempt as yielding a negative result; that evidential content, probably, is not susceptible to analysis by a uniform theory.

I'll ignore most of the data from Aikhenvald here. Let's assume for the moment that Quechua and Japanese exemplify two possibilities for evidential semantics, and that those two possibilities are the only ones, something that the Aikhenvald data makes very unlikely. It will be convenient to idealize here though to make clear what the account is supposed to do.

First the Japanese case. Here evidential content need not take widest scope as in Quechua, and further acts much like more usual modal content. So for this case we need not say anything special; the McCready-Ogata logic captures the facts. The interesting case when going from this theory is the Quechua one. For this we must assume that Quechua evidential content is conventionally implicated (on the account I am presently exploring).

Now it's easy to see how to modify the modal semantics for Japanese evidentials to account for the Quechua evidentials—simply assume that there is a semantic parameter set differently in these languages, to the effect that the type of Japanese evidentials is $\langle t^a, t^a \rangle$ while that of Quechua evidentials is $\langle t^a, t^c \rangle$. Then we can translate *-mi* as (given a definition for Bpg = 'best possible grounds' in the McCready-Ogata logic, which should be straightforward to give, at least to the level of explicitness that seems possible at the current time)

$$\lambda p.[Bpg(p)] : \langle t^a, t^c \rangle,$$

-chá as (slightly simplifying the indexing etc for expositional purposes)

$$\lambda p.[\triangle p] : \langle t^a, t^c \rangle$$

and finally *-si*, the hearsay evidential, as

$$\lambda p.[\mathsf{H}(p)] : \langle t^a, t^c \rangle.$$

This will have several effects. The first is to cause all evidential content to be scopeless, effectively accounting for most of what Faller dealt with by using speech acts. The other effects, unfortunately, are much less desirable. Let me point out two.

The first is what happens with 'copying' or reusing content. A feature of the CI logic already discussed above is that we always get a 'copy' of whatever content is in the scope of a CI operator: so proof-theoretically speaking the following situation holds:

$$\lambda x_\sigma.[Op(\sigma)] : \langle \sigma^a, \tau^c \rangle, a : \sigma \vdash Op(\sigma) : \tau^c, a : \sigma^a.$$

This means that, in the evidential case, we will get the propositional content back after the evidential operator applies. In the case of *-mi* this seems fine: an assertion of $Bpg(p)$ acts very much like an assertion of p. But it seems that this is probably so for pragmatic and not semantic reasons. Assertions of S-*cha* or S-*si* do not entail that the content of S is true, as Faller shows; rather, the former entails $Might(\llbracket S \rrbracket)$ and the latter entails only that $Hearsay(\llbracket S \rrbracket)$ (as discussed above). We surely do not want this situation. This is a real problem and one I don't see any principled way to solve if one assumes the CI logic.

The next problem involves questions. Faller shows that evidential content can scope under questions: one can question the hearer's evidence using a question, not just indicate the speaker's evidence for the question (whatever that would mean). So the following two scopings are possible for [Q [S Evid]]:

(20) a. Evid(Q(S))
b. Q(Evid(S))

The CI logic can only get a third (non)scoping (angled brackets indicating the two dimensions of content in this theory):

(21) ⟨Q(S), Evid(S)⟩

This seems wrong. Faller indicates that the speech act analysis can get these cases right for, if evidentials are speech act modifiers, surely they admit modification by questions, which are also (on one theoretical picture) speech act modifiers. I don't see how to overcome this problem either. And there are other difficulties indicated in the section on data above.

The lessons of all this seem to be the following. First, though evidential content seems to have some of the properties of conventionally implicated content, it doesn't share enough of them for an analysis designed to handle CI content to deal with evidentials as well, at least not right 'off the shelf.' Maybe there are modifications that could be made so that all this would work, but I do not see how to produce them at this point. Second, and this should be taken probably as the bigger lesson, evidentials vary widely in their semantic behavior. Much more research is needed to really understand how they work cross-linguistically; the work done so far on Quechua, Japanese, and a few other languages is only a beginning. Hopefully as a broader picture of the empirical facts comes into focus, it will become clearer whether a unified analysis is possible and desirable, and, if so, what form it should take.

References

1. Aikhenvald, A.: Evidentiality. Oxford University Press, Oxford (2004)
2. Izvorski, R.: The present perfect as an epistemic modal. In: Lawson, A., Cho, E. (eds.) Proceedings of SALT 7, CLC Publications (1997)
3. McCready, E.: The Dynamics of Particles. PhD thesis, UTexas-Austin (2005)
4. McCready, E., Asher, N.: Modal subordination in Japanese: Dynamics and evidentiality. In: Eilam, A., Scheffler, T., Tauberer, J. (eds.) Penn Working Papers in Linguistics 12.1, pp. 237–249 (2006)

5. Faller, M.: Semantics and Pragmatics of Evidentials in Cuzco Quechua. PhD thesis, Stanford University (2002)
6. Davis, C., Potts, C., Speas, P.: The pragmatic values of evidential sentences. In: Gibson, M., Friedman, T. (eds.) Proceedings of SALT 17, CLC Publications (to appear)
7. McCready, E., Ogata, N.: Evidentiality, modality, and probability. Linguistics and Philosophy 30(2), 147–206 (2007)
8. Roberts, C.: Modal subordination and pronominal anaphora in discourse. Linguistics and Philosophy 12, 683–721 (1989)
9. Masuoka, T., Takubo, Y.: Kisoo Nihongo Bunpoo [Essential Japanese Grammar]. Kuroshio Shuppan, Tokyo (1989)
10. van der Sandt, R.: Presupposition projection as anaphora resolution. Journal of Semantics 9, 333–377 (1992)
11. Geurts, B.: Presupposition and Pronouns. Elsevier, Oxford (1999)
12. Asher, N., Lascarides, A.: Logics of Conversation. Cambridge University Press, Cambridge (2003)
13. Kamp, H., Reyle, U.: From Discourse to Logic. Kluwer, Dordrecht, Reidel (1993)
14. Vanderveken, D.: Meaning and Speech Acts. Cambridge University Press, Cambridge (1990)
15. Aikhenvald, A.: Evidentiality in typological perspective. In: Aikhenvald, A., Dixon, R. (eds.) Studies in Evidentiality, pp. 1–31. Johns Benjamins, Amsterdam (2003)
16. de Haan, F.: Evidentiality and epistemic modality: Setting boundaries. Southwest Journal of Linguistics 18, 83–101 (1999)
17. Potts, C.: Expressive content as conventional implicature. In: Proceedings of NELS, vol. 33 (2003)
18. Potts, C.: The Logic of Conventional Implicatures. Oxford University Press, Oxford (2005)(revised version of 2003 UCSC dissertation)
19. Girard, J.Y.: Linear logic. Theoretical Computer Science 50, 1–102 (1987)
20. Asher, N., Lascarides, A.: Indirect speech acts. Synthese 128, 183–228 (2001)

Acts of Promising in Dynamified Deontic Logic

Tomoyuki Yamada

Division of Philosophy and Cultural Sciences,
Graduate School of Letters, Hokkaido University
Nishi 7, Kita 10, Kita-ku, Sapporo, Hokkaido, 060-0810, Japan
yamada@LET.hokudai.ac.jp

Abstract. In this paper, the logic of acts of commanding ECL II introduced in Yamada (2007b) will be extended in order to model acts of promising together with acts of commanding. Effects of both kinds of acts are captured in terms, not of changes they bring about on propositional attitudes of their addressees, but of changes they bring about on deontic status of relevant action alternatives; they are modeled as deontic updators. This enables us to see how an act of promising performed by an agent and an act of commanding performed by another agent can jointly bring about a conflict of obligations. Complete axiomatization will be presented, and a comparison with Searle's treatment of acts of promising in his argument for the derivability of "ought" from "is" will be made.

1 Introduction

In the last two decades, systems of dynamic epistemic logic (DEL) have been developed to deal with dynamic changes brought about by various kinds of information transmissions including public announcements as well as private communications in Plaza (1989), Gerbrandy & Groeneveld (1997), Baltag, Moss, & Solecki (1999), and Kooi & van Benthem (2004) among others. Model updating operations are introduced to interpret these communicative acts as what update epistemic states of agents involved, and dynamic epistemic logics are obtained as dynamic extensions of static epistemic logics. In the logics of acts of commanding, ECL of Yamada (2007a) and ECL II of Yamada (2007b) , similar model updating operations are introduced to interpret acts of commanding as updators of deontic aspects of the situations in which agents are involved.[1]

ECL and ECL II are dynamic extensions of multi-agent variants of static monadic deontic logic MDL^+ and $\mathsf{MDL}^+\mathsf{II}$ respectively. In MDL^+, deontic operators are allowed to be indexed by a given finite set of agents in order to distinguish agents to whom commands are given from other agents, and in $\mathsf{MDL}^+\mathsf{II}$, they are allowed to be indexed by the Cartesian product of a given finite set of agents and a given finite set of command issuing authorities in order to deal with (possibly conflicting) obligations generated by commands given to the same agents by different authorities. The purpose of

[1] "ECL" is the abbreviation of "Eliminative Command Logic". ECL is said to be eliminative as the model updating operation introduced in its semantics always eliminates zero or more accessibility links. The model updating operations to be introduced in this paper are also eliminative in this sense as will be seen in Section 3.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 95–108, 2008.

the present paper is to extend $\mathsf{MDL}^+\mathsf{II}$ and $\mathsf{ECL\,II}$ in order to deal with changes brought about by acts of promising along with those brought about by acts of commanding.

Consider the following example:

Example 1. Suppose you receive a letter from your political guru, in which she commands you to join an important political demonstration to be held in Tokyo next month. Unfortunately, it is to be held on the very same day on which an international one-day conference on logic is to be held in São Paulo, and you have already promised your former student who organizes that conference that you would give an invited lecture there. It is possible for you to join the demonstration in Tokyo, but if you choose to do so, you will fail to keep your promise. It is also possible for you to give a lecture at the conference in São Paulo, but if you choose to do so, you will fail to obey your guru's command. No available means of transportaion are fast enough to enable you to join both events on the same day even though the time in São Paulo is 12 hours behind the time in Tokyo. You have to decide which alternative to choose.

In this example, your guru's command is in conflict with your earlier promise. In this paper we will analyze how this conflict has been brought about jointly by your act of promising and your guru's act of commanding by developing a logic in which both kinds of acts are modeled as deontic updators. For this purpose, $\mathsf{ECL\,II}$ will be extended by introducing terms standing for types of acts of promising. In order to do so, however, we also have to reconsider $\mathsf{MDL}^+\mathsf{II}$.

In Section 2, we refine $\mathsf{MDL}^+\mathsf{II}$ by allowing deontic operators to be indexed by a triad of agents. Then in section 3, we develop a dynamic extension of this refined Multi-agent Deontic Logic by introducing modalities that model various acts of commanding and promising. In Section 4, we discuss some interesting notions and interplay between acts of commanding and acts of promising expressible in the extended language. Then in section 5, complete axiomatization will be given to the dynamified logic, and finally in section 6, we compare our treatment of acts of promising with John Searle's treatment in his argument for the derivability of 'ought' from 'is', and make a brief comment on the difference between Searle's treatment of illocutionary acts and Austin's.

2 The Static Base Logic $\mathsf{MDL}^+\mathsf{III}$

In $\mathsf{MDL}^+\mathsf{II}$ (and in ECLII), we have a formula of the form $O_{(i,j)}\varphi$. Intuitively, it means that it is obligatory upon an agent i with respect to an authority j to see to it that φ. Let p and q denote the proposition that you will give a lecture at the workshop in São Paulo and the proposition that you will join the demonstration in Tokyo respectively, and let a, b, and c represent you, your former student, and your guru respectively. Then in ECLII, the type of your guru's command is represented by the expression of the form $!_{(a,c)}q$, where a is the commandee and c is the commander, and the following formula is shown to be valid:

$$[!_{(a,c)}q]O_{(a,c)}q \ . \tag{1}$$

Intuitively, this formula means that after c's successful act of commanding a to see to it that q, it is obligatory upon a with respect to c to see to it that q. The validity of this formula guarantees that in the updated model, which represents the situation after the issuance of your guru's command, the following formula holds at the current world:

$$O_{(a,c)}q \quad . \tag{2}$$

Thus the conventional effect of your guru's act of commanding is captured in $\mathsf{MDL}^+\mathsf{II}$ (and in ECLII, too, since ECLII extends $\mathsf{MDL}^+\mathsf{II}$).

Now consider your earlier promise. Let the expression of the form $\mathrm{Prom}_{(i,j)}\varphi$ denote the type of acts of promising in which an agent i promises an agent j that she will see to it that φ. Then the type of your act of promising can be denoted by $\mathrm{Prom}_{(a,b)}p$, and it's not difficult to define semantics which validates the following formula:

$$[\mathrm{Prom}_{(a,b)}p]O_{(a,b)}p \quad . \tag{3}$$

This means that the following formula holds at the current world in the updated model:

$$O_{(a,b)}p \quad . \tag{4}$$

Notice, however, that this is not exactly what we want, since your former student need not be among the command issuing authorities. In order to accommodate obligation generated by acts of promising, we have to reinterpret the second constituent of the indexing pair. If we reinterpret it as the agent whose act generates the obligation, we should have:

$$[\mathrm{Prom}_{(a,b)}p]O_{(a,a)}p \quad . \tag{5}$$

But then, the promisee will be left unmentioned in the formula characterizing the obligation generated. It doesn't seem quite good as the obligations generated by acts of promising are among the kind of obligations sometimes referred to as "special obligations" in the literature; they are owed to some subset of persons (obligees) but not to all the persons, and the promisee is exactly the agent to whom the obligation generated by an act of promising is supposed to be owed.[2] Although there are disputes on whether we have genuinely special obligations, common sense morality seems to understand us as having special obligations, and thus, it would be of considerable theoretical interest to examine how such obligations could be modeled in logical terms.

Notice that we have seen here an interesting difference between acts of commanding and acts of promising. In the development of the logic of acts of commanding, we haven't felt any need for a separate parameter standing for an agent to whom the obligation is owed (an obligee) over and above parameters standing for the commandee (the agent who is obligated) and the commander (the creator of the obligation). In the case of the obligation generated by an act of promising, however, there seems to be an obligee who is, at least in usual cases, distinct not only from the agent who is obligated (the obligor) but also from the creator of the obligation, who is identical with the obligor. Now the question is whether or not the notion of obligee is applicable to the obligations generated by acts of commanding. The negative answer here will imply that the logical form of the obligations generated by acts of commanding is different from that of the

[2] A detailed discussion of special obligation can be found in Jeske (2002).

obligations generated by acts of promising, unless we deny the applicability of the notion of obligee to the obligations generated by acts of promising. The positive answer will require us to explain why we haven't felt any need for a parameter standing for an obligee in developing the logics of acts of commanding. One possible explanation seems to be that the obligee is identical with the creator of the obligation in the case of obligations generated by acts of commanding. As a logic based on the negative answer combined with the abandoning of the notion of obligee can be obtained by simplifying a logic based on the positive answer, we will only develop a logic that incorporates the positive answer in this paper.

Thus, we define:

Definition 1. *Take a countably infinite set* **Aprop** *of proposition letters, and a finite set I of agents, with p ranging over* **Aprop**, *and i, j, k over I. The refined language $\mathcal{L}_{\mathsf{MDL^{+}III}}$ of the multi-agent monadic deontic logic with an alethic operator* $\mathsf{MDL^{+}III}$ *is given by:*

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid \Box\varphi \mid O_{(i,j,k)}\varphi$$

The set of all well formed formulas (sentences) of $\mathcal{L}_{\mathsf{MDL^{+}III}}$ is denoted by $S_{\mathsf{MDL^{+}III}}$ and operators of the form $O_{(i,j,k)}$ are called deontic operators. For each $i, j, k \in I$, we call a sentence (i, j, k)-free if no $O_{(i,j,k)}$'s occur in it. We call sentence alethic if no deontic operators occur in it, and boolean if no modal operators occur in it. For each $i, j, k \in I$, the set of all (i, j, k)-free sentences is denoted by $S_{(i,j,k)\text{-free}}$. The set of all alethic sentences and the set of all boolean sentences are denoted by S_{Aleth} and S_{Boole} respectively.

$\bot$, $\vee$, $\rightarrow$, $\leftrightarrow$, and $\Diamond$ are assumed to be introduced by standard definitions. We also abbreviate $\neg O_{(i,j,k)}\neg\varphi$ as $P_{(i,j,k)}\varphi$, and $O_{(i,j,k)}\neg\varphi$ as $F_{(i,j,k)}\varphi$. Note that **Aprop** $\subset S_{\mathrm{Boole}} \subset S_{\mathrm{Aleth}} \subset S_{(i,j,k)\text{-free}} \subset S_{\mathsf{MDL^{+}III}}$ for each $i, j, k \in I$.

A formula of the form $O_{(i,j,k)}\varphi$ is to be understood as meaning that it is obligatory upon an agent i with respect to an agent j in the name of k to see to it that φ. In order to accommodate this fine grained notion of obligation, we allow deontic accessibility relations to be indexed by $I \times I \times I$. Thus we define:

Definition 2. *By an $\mathcal{L}_{\mathsf{MDL^{+}III}}$-model, we mean a quadruple $M = (W^M, R_A^M, R_D^M, V^M)$ where:*

(i) *W^M is a non-empty set (heuristically, of 'possible worlds'),*
(ii) *$R_A^M \subseteq W^M \times W^M$,*
(iii) *R_D^M is a function that assigns a subset $R_D^M(i, j, k)$ of R_A^M to each triad (i, j, k) of agents $i, j, k \in I$,*
(iv) *V^M is a function that assigns a subset $V^M(p)$ of W^M to each proposition letter $p \in$* **Aprop**.

We usually abbreviate $R_D^M(i, j, k)$ as $R_{(i,j,k)}^M$.

The truth definition for $\mathcal{L}_{\mathsf{MDL^{+}III}}$ can be given in a standard way by associating the alethic modal operator $\Box$ with R_A^M and each deontic operator $O_{(i,j,k)}$ with $R_{(i,j,k)}^M$.

Definition 3. *Let M be an $\mathcal{L}_{\mathsf{MDL\text{-}III}}$-model and w a point in M. If $p \in Aprop$, $\varphi, \psi \in S_{\mathsf{MDL\text{-}III}}$, and $i, j, k \in I$, then:*

(a) $M, w \models_{\mathsf{MDL\text{-}III}} p$ *iff* $w \in V^M(p)$,
(b) $M, w \models_{\mathsf{MDL\text{-}III}} \top$,
(c) $M, w \models_{\mathsf{MDL\text{-}III}} \neg\varphi$ *iff* $M, w \not\models_{\mathsf{MDL\text{-}III}} \varphi$,
(d) $M, w \models_{\mathsf{MDL\text{-}III}} (\varphi \wedge \psi)$ *iff* $M, w \models_{\mathsf{MDL\text{-}III}} \varphi$ *and* $M, w \models_{\mathsf{MDL\text{-}III}} \psi$,
(e) $M, w \models_{\mathsf{MDL\text{-}III}} \Box\varphi$ *iff for every v such that* $(w, v) \in R^M_A$, $M, v \models_{\mathsf{MDL\text{-}III}} \varphi$,
(f) $M, w \models_{\mathsf{MDL\text{-}III}} O_{(i,j,k)}\varphi$ *iff for every v such that* $(w, v) \in R^M_{(i,j,k)}$, $M, v \models_{\mathsf{MDL\text{-}III}} \varphi$.

A formula φ is true in an $\mathcal{L}_{\mathsf{MDL\text{-}III}}$-model M at a point w of M if $M, w \models_{\mathsf{MDL\text{-}III}} \varphi$. We say that a set Σ of formulas of $\mathcal{L}_{\mathsf{MDL\text{-}III}}$ is true in M at w, and write $M, w \models_{\mathsf{MDL\text{-}III}} \Sigma$, if $M, w \models_{\mathsf{MDL\text{-}III}} \psi$ for every $\psi \in \Sigma$. If $\Sigma \cup \{\varphi\}$ is a set of formulas of $\mathcal{L}_{\mathsf{MDL\text{-}III}}$, we say that φ is a semantic consequence of Σ, and write $\Sigma \models_{\mathsf{MDL\text{-}III}} \varphi$, if for every $\mathcal{L}_{\mathsf{MDL\text{-}III}}$-model M and every point w such that $M, w \models_{\mathsf{MDL\text{-}III}} \Sigma$, $M, w \models_{\mathsf{MDL\text{-}III}} \varphi$. We say that a formula φ is valid, and write $\models_{\mathsf{MDL\text{-}III}} \varphi$, if $\emptyset \models_{\mathsf{MDL\text{-}III}} \varphi$.

From a purely formal point of view, $\mathcal{L}_{\mathsf{MDL\text{-}III}}$ is an instance of $\mathcal{L}_{\mathsf{MDL}^-}$, and every $\mathcal{L}_{\mathsf{MDL\text{-}III}}$-model is an $\mathcal{L}_{\mathsf{MDL}^-}$-model, since $I \times I \times I$ is just a finite set. Deontic operators of $\mathcal{L}_{\mathsf{MDL}^-}$ and accessibility relations associated with them are required to be indexed by a finite set, and the deontic operators of $\mathcal{L}_{\mathsf{MDL\text{-}III}}$ and the accessibility relations associated with them satisfy this requirement. This guarantees that the completeness of $\mathsf{MDL}^+\mathsf{III}$ can be derived from the known completeness of MDL^+.[3]

Definition 4. *The proof system for* $\mathsf{MDL}^+\mathsf{III}$ *contains the following axioms and rules:*

(Taut)	*all instantiations of propositional tautologies over the present language*	
($\Box$-Dist)	$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	
(O-Dist)	$O_{(i,j,k)}(\varphi \to \psi) \to (O_{(i,j,k)}\varphi \to O_{(i,j,k)}\psi)$	*for each* $(i, j, k) \in I \times I \times I$
(Mix)	$P_{(i,j,k)}\varphi \to \Diamond\varphi$	*for each* $(i, j, k) \in I \times I \times I$
(MP)	$\dfrac{\varphi \quad \varphi \to \psi}{\psi}$	
($\Box$-Nec)	$\dfrac{\varphi}{\Box\varphi}$	
(O-Nec)	$\dfrac{\varphi}{O_{(i,j,k)}\varphi}$	*for each* $(i, j, k) \in I \times I \times I$.

An $\mathsf{MDL}^+\mathsf{III}$*-proof of a formula φ is a finite sequence of $\mathcal{L}_{\mathsf{MDL\text{-}III}}$-formulas having φ as the last formula such that each formula is either an instance of an axiom, or it can be obtained from formulas that appear earlier in the sequence by applying a rule. If there is a proof of φ, we write $\vdash_{\mathsf{MDL\text{-}III}} \varphi$. If $\Sigma \cup \{\varphi\}$ is a set of $\mathcal{L}_{\mathsf{MDL\text{-}III}}$-formulas, we say that φ is deducible in* $\mathsf{MDL}^+\mathsf{III}$ *from Σ and write $\Sigma \vdash_{\mathsf{MDL\text{-}III}} \varphi$ if $\vdash_{\mathsf{MDL\text{-}III}} \varphi$ or there are formulas $\psi_1, \ldots, \psi_n \in \Sigma$ such that $\vdash_{\mathsf{MDL\text{-}III}} (\psi_1 \wedge \cdots \wedge \psi_n) \to \varphi$.*

[3] The completeness of MDL^+ is proved by building a canonical model in a standard way.

The above rules obviously preserve validity, and all the axioms are easily seen to be valid. Thus this proof system is sound.[4] Moreover, as is said above, the completeness of this proof system is guaranteed by the completeness of MDL^+.

Theorem 1 (Completeness of $\mathsf{MDL}^+\mathsf{III}$). *Let $\Sigma \cup \{\varphi\}$ be a set of $\mathcal{L}_{\mathsf{MDL}^+\mathsf{III}}$-formulas. Then, if $\Sigma \models_{\mathsf{MDL}^+\mathsf{III}} \varphi$ then $\Sigma \vdash_{\mathsf{MDL}^+\mathsf{III}} \varphi$.*

3 The Dynamified Multi-agent Deontic Logic $\mathsf{DMDL}^+\mathsf{III}$

The formulas of $\mathcal{L}_{\mathsf{MDL}^+\mathsf{III}}$ can be used to talk about the situations before and after the issuance of a promise or a command. In the previous example, before the issuance of your promise, it was not obligatory upon you to give a lecture at the workshop in São Paulo, but since the issuance it has become obligatory. Let p, a and b be understood as before, and (L, s) and (M, s) be the model world pairs that represent the situations before and after the issuance respectively. Then we should have:

$$L, s \models_{\mathsf{MDL}^+\mathsf{III}} \neg O_{(a,b,a)} p \tag{6}$$

$$M, s \models_{\mathsf{MDL}^+\mathsf{III}} O_{(a,b,a)} p \quad . \tag{7}$$

Note that we have assumed that in the case of an obligation created by an act of promising, the agent who is obligated is identical with the agent in whose name obligation is created. In the case of an obligation created by an act of commanding, in contrast, they are usually distinct. Let (N, s) represent the situation after the issuance of your guru's command, and let q and c be understood as before. Then we should have:

$$M, s \models_{\mathsf{MDL}^+\mathsf{III}} \neg O_{(a,c,c)} q \tag{8}$$

$$N, s \models_{\mathsf{MDL}^+\mathsf{III}} O_{(a,c,c)} q \quad . \tag{9}$$

In order to have a way of talking about changes of this kind and the acts that bring them about in the object language, we dynamify $\mathsf{MDL}^+\mathsf{III}$. Thus we define:

Definition 5. *Take the same countably infinite set* **Aprop** *of proposition letters and the same finite set I of agents as before, with p ranging over* **Aprop**, *and i, j, k over I. The refined language $\mathcal{L}_{\mathsf{DMDL}^+\mathsf{III}}$ of dynamified multi-agent deontic logic $\mathsf{DMDL}^+\mathsf{III}$ is given by:*

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid \Box\varphi \mid O_{(i,j,k)}\varphi \mid [\pi]\varphi$$
$$\pi ::= \mathrm{Com}_{(i,j)}\varphi \mid \mathrm{Prom}_{(i,j)}\varphi$$

Terms of the form $\mathrm{Com}_{(i,j)}\varphi$ and $\mathrm{Prom}_{(i,j)}\varphi$ are called command type terms and promise type terms respectively, and operators of the form $[\mathrm{Com}_{(i,j)}\varphi]$ and $[\mathrm{Prom}_{(i,j)}\varphi]$ are called command operators and promise operators respectively. The set of all well formed formulas of $\mathcal{L}_{\mathsf{DMDL}^+\mathsf{III}}$ is referred to as $S_{\mathsf{DMDL}^+\mathsf{III}}$.

Here the command type term of the form $!_{(j,i)}\varphi$ of $\mathcal{L}_{\mathsf{ECL\,II}}$ is replaced by the term of the form $\mathrm{Com}_{(i,j)}\varphi$ for mnemonic convenience. Note the inversion of the order of the

[4] Strictly speaking, O-Nec is redundant since it is derivable. It is included here just to record the fact that $\mathsf{MDL}^+\mathsf{III}$ is normal.

constituents of the indexing pair; the term of the form $\mathrm{Com}_{(i,j)}\varphi$ stands for the type of acts of commanding to the effect that j should see to it that φ of which the commander is i and the commandee is j . Note also that $S_{\mathsf{MDL\text{-}III}} \subset S_{\mathsf{DMDL\text{-}III}}$.

The truth definition for this language can be given with reference to $\mathcal{L}_{\mathsf{MDL\text{-}III}}$-models.

Definition 6. *Let M be an $\mathcal{L}_{\mathsf{MDL\text{-}III}}$-model and w a point in M. If $p \in Aprop$, φ, $\psi \in S_{\mathsf{DMDL\text{-}III}}$, and $i, j, k \in I$, then:*

(a) $M, w \models_{\mathsf{DMDL\text{-}III}} p$ *iff* $w \in V^M(p)$,
(b) $M, w \models_{\mathsf{DMDL\text{-}III}} \top$,
(c) $M, w \models_{\mathsf{DMDL\text{-}III}} \neg\varphi$ *iff* $M, w \not\models_{\mathsf{DMDL\text{-}III}} \varphi$,
(d) $M, w \models_{\mathsf{DMDL\text{-}III}} (\varphi \wedge \psi)$ *iff* $M, w \models_{\mathsf{DMDL\text{-}III}} \varphi$ *and* $M, w \models_{\mathsf{DMDL\text{-}III}} \psi$,
(e) $M, w \models_{\mathsf{DMDL\text{-}III}} \Box\varphi$ *iff for every* v *such that* $(w, v) \in R^M_A$, $M, v \models_{\mathsf{DMDL\text{-}III}} \varphi$,
(f) $M, w \models_{\mathsf{DMDL\text{-}III}} O_{(i,j,k)}\varphi$ *iff for every* v *such that* $(w, v) \in R^M_{(i,j,k)}$, $M, v \models_{\mathsf{DMDL\text{-}III}} \varphi$,
(g) $M, w \models_{\mathsf{DMDL\text{-}III}} [\mathrm{Com}_{(i,j)}\chi]\varphi$ *iff* $M_{\mathrm{Com}_{(i,j)}\chi}, w \models_{\mathsf{DMDL\text{-}III}} \varphi$,
(h) $M, w \models_{\mathsf{DMDL\text{-}III}} [\mathrm{Prom}_{(i,j)}\chi]\varphi$ *iff* $M_{\mathrm{Prom}_{(i,j)}\chi}, w \models_{\mathsf{DMDL\text{-}III}} \varphi$,

where

(i) $M_{\mathrm{Com}_{(i,j)}\chi}$ *is the* $\mathcal{L}_{\mathsf{MDL\text{-}III}}$*-model obtained from* M *by replacing* $R^M_D(j, i, i)$ *with* $\{(x, y) \in R^M_D(j, i, i) \mid M, y \models_{\mathsf{DMDL\text{-}III}} \chi\}$, *and*
(ii) $M_{\mathrm{Prom}_{(i,j)}\chi}$ *is the* $\mathcal{L}_{\mathsf{MDL\text{-}III}}$*-model obtained from* M *by replacing* $R^M_D(i, j, i)$ *with* $\{(x, y) \in R^M_D(i, j, i) \mid M, y \models_{\mathsf{DMDL\text{-}III}} \chi\}$.

A formula φ is true in an $\mathcal{L}_{\mathsf{MDL\text{-}III}}$-model M at a point w of M if $M, w \models_{\mathsf{DMDL\text{-}III}} \varphi$. We say that a set Σ of formulas of $\mathcal{L}_{\mathsf{DMDL\text{-}III}}$ is true in M at w, and write $M, w \models_{\mathsf{DMDL\text{-}III}} \Sigma$, if $M, w \models_{\mathsf{DMDL\text{-}III}} \psi$ for every $\psi \in \Sigma$. If $\Sigma \cup \{\varphi\}$ is a set of formulas of $\mathcal{L}_{\mathsf{DMDL\text{-}III}}$, we say that φ is a semantic consequence of Σ, and write $\Sigma \models_{\mathsf{DMDL\text{-}III}} \varphi$, if for every $\mathcal{L}_{\mathsf{DMDL\text{-}III}}$-model M and every point w such that $M, w \models_{\mathsf{DMDL\text{-}III}} \Sigma$, $M, w \models_{\mathsf{DMDL\text{-}III}} \varphi$. We say that a formula φ is valid, and write $\models_{\mathsf{DMDL\text{-}III}} \varphi$, if $\emptyset \models_{\mathsf{DMDL\text{-}III}} \varphi$.

The clause (g) here is a restatement of the clause for the formulas of the form $[!_{(j,i)}\chi]\varphi$ of the truth definition for $\mathcal{L}_{\mathsf{ECL\,II}}$. As the truth of $[\mathrm{Com}_{(i,j)}\chi]\varphi$ at w in M is defined in terms of the truth of φ at w in the updated model $M_{\mathrm{Com}_{(i,j)}\chi}$ in (g), the truth of $[\mathrm{Prom}_{(i,j)}\chi]\varphi$ at w in M is defined in terms of the truth of φ at w in the updated model $M_{\mathrm{Prom}_{(i,j)}\chi}$ in (h).

As we have $\{(x, y) \in R^M_D(j, i, i) \mid M, y \models_{\mathsf{DMDL\text{-}III}} \chi\} \subseteq R^M_D(j, i, i) \subseteq R^M_A$ and $\{(x, y) \in R^M_D(i, j, i) \mid M, y \models_{\mathsf{DMDL\text{-}III}} \chi\} \subseteq R^M_D(i, j, i) \subseteq R^M_A$, the updated models $M_{\mathrm{Com}_{(i,j)}\chi}$ and $M_{\mathrm{Prom}_{(i,j)}\chi}$ are guaranteed to be $\mathcal{L}_{\mathsf{MDL\text{-}III}}$-models. Moreover, as the remaining clauses faithfully reproduce the clauses of the truth definition for $\mathcal{L}_{\mathsf{MDL\text{-}III}}$, we obviously have:

Corollary 1. *Let M be an $\mathcal{L}_{\mathsf{MDL\text{-}III}}$-model and w a point of M. Then for any $\varphi \in S_{\mathsf{MDL\text{-}III}}$, $M, w \models_{\mathsf{DMDL\text{-}III}} \varphi$ iff $M, w \models_{\mathsf{MDL\text{-}III}} \varphi$.*

The following corollary can be proved by induction on the length of the formula:

Corollary 2. *Let ψ and χ be an (j, i, i)-free formula and an (i, j, i)-free formula respectively. Then, for any $\varphi \in S_{\mathsf{DMDL\text{-}III}}$, the following two equivalences hold:*

(i) $M, w \models_{\mathsf{DMDL\text{-}III}} \psi$ *iff* $M_{Com_{(i,j)}\varphi}, w \models_{\mathsf{DMDL\text{-}III}} \psi$
(ii) $M, w \models_{\mathsf{DMDL\text{-}III}} \chi$ *iff* $M_{Prom_{(i,j)}\varphi}, w \models_{\mathsf{DMDL\text{-}III}} \chi$.

One of the things this corollary means is that acts of commanding and acts of promising do not affect so-called brute facts and alethic possibilities in any direct way. But it means more, as we will see below.

DMDL$^+$III inherits the following principle from ECL II:

Proposition 1 (CUGO Principle). *If* $\varphi \in S_{(j,i,i)\text{-free}}$, *then* $\models_{\mathsf{DMDL^+III}} [Com_{(i,j)}\varphi]O_{(j,i,i)}\varphi$.

As is noted in Yamada(2007a), CUGO Principle characterizes, at least partially, the workings of acts of commanding; though not without exceptions, *commands usually generate obligations*.[5] Our semantics also validates the following principle for DMDL$^+$III:

Proposition 2 (PUGO Principle). *If* $\varphi \in S_{(i,j,i)\text{-free}}$, *then* $\models_{\mathsf{DMDL^+III}} [Prom_{(i,j)}\varphi]O_{(i,j,i)}\varphi$.

PUGO Principle means that, though not without exceptions, *promises usually generate obligations*.[6]

Now let's go back to our discussion of Example 1 above. Since p is (a,b,a)-free, PUGO Principle guarantees that we have:

$$L, s \models_{\mathsf{DMDL^+III}} [Prom_{(a,b)}p]O_{(a,b,a)}p \ . \tag{10}$$

This is equivalent to:

$$L_{Prom_{(a,b)}p}, s \models_{\mathsf{DMDL^+III}} O_{(a,b,a)}p \ . \tag{11}$$

Since q is (a,c,c)-free, CUGO Principle guarantees that we have:

$$L_{Prom_{(a,b)}p}, s \models_{\mathsf{DMDL^+III}} [Com(c,a)q]O_{(a,c,c)}q \ . \tag{12}$$

This is equivalent to:

$$(L_{Prom_{(a,b)}p})_{Com(c,a)q}, s \models_{\mathsf{DMDL^+III}} O_{(a,c,c)}q \ . \tag{13}$$

Moreover, since $O_{(a,b,a)}p$ is (a,c,c)-free, Corollary 2 and (11) jointly imply:

$$(L_{Prom_{(a,b)}p})_{Com(c,a)q}, s \models_{\mathsf{DMDL^+III}} O_{(a,b,a)}p \ . \tag{14}$$

Hence we have

$$(L_{Prom_{(a,b)}p})_{Com(c,a)q}, s \models_{\mathsf{DMDL^+III}} (O_{(a,b,a)}p \wedge O_{(a,c,c)}q) \ . \tag{15}$$

The model world pair $((L_{Prom_{(a,b)}p})_{Com(c,a)q}, s)$ here represents the situation you are in after the issuance of your guru's command. In that situation, it is obligatory upon you to see to it that q with respect to your guru in the name of your guru, but it is also obligatory upon you to see to it that p with respect to your former student in your

[5] The restriction on φ here is motivated by the fact that the truth of φ at a point v in M does not guarantee the truth of φ at v in $M_{Com_{(i,j)}\varphi}$ if φ is not (j,i,i)-free. For example, $[Com_{(i,j)}P_{(j,i,i)}q]O_{(j,i,i)}P_{(j,i,i)}q$ is not valid. For more on CUGO Principle, see Yamada(2007a).

[6] The motivation for the restriction on φ here is similar to that for CUGO Principle.

name. Your guru's command added a new obligation without removing your earlier commitment. They are independent from each other as Corollary 2 indicates.

Now, given that we have $L, s \models_{\mathsf{DMDL\text{-}III}} \neg(p \wedge q)$, Corollary 2 again enables us to establish:

$$(L_{Prom_{(a,b)}p})_{Com(c,a)q}, s \models_{\mathsf{DMDL\text{-}III}} (O_{(a,b,a)}p \wedge O_{(a,c,c)}q) \wedge \neg(p \wedge q) \ . \tag{16}$$

Thus we have captured how the contingent conflict of obligations in our example is brought about jointly by your act of promising and your guru's act of commanding.

4 Some Interesting Things Expressible in $\mathcal{L}_{\mathsf{MDL^{+}III}}$ and $\mathcal{L}_{\mathsf{DMDL^{+}III}}$

Our definitions of $\mathcal{L}_{\mathsf{MDL\text{-}III}}$ and of $\mathcal{L}_{\mathsf{DMDL\text{-}III}}$ leave room for interesting possibilities. For example, we may ask if there can be an obligation of the form $O_{(i,j,k)}\varphi$ such that $i \neq j \neq k \neq i$. Suppose, for example, the director of your research center utters the following sentence seriously and sincerely to someone over the phone:

My secretary will call you back right away. (17)

Let a, b, c be the director, her secretary, and the addressee respectively, and let p represent the proposition that b will call c back right away. Does her utterance create an obligation of the form $O_{(b,c,a)}p$?

Notice that no combinations of acts of promising and commanding will generate such an obligation in $\mathsf{DMDL^{+}III}$. Thus, if we are to take the director's utterance as generating an obligation of the form $O_{(b,c,a)}p$, we will have to introduce a new program term of the form, say, $Prom^{*}_{(i,j,k)}\varphi$, and let an instance of it represent the type of her act of promising. Although we will not pursue this possibility further in this paper, we just mention that, if we do so, it will become possible to define usual acts of promising of the form $Prom_{(i,j)}\varphi$ as an abbreviation for $Prom^{*}_{(i,j,i)}\varphi$.

Another interesting possibility is an obligation of the form $O_{(i,i,i)}\varphi$. Such an obligation will be generated by an act of commanding of the type $Com_{(i,i)}\varphi$, which is an act of commanding oneself to see to it that φ, as well as by an act of promising of the type $Prom_{(i,i)}\varphi$, which is an act of promising oneself that (s)he will see to it that φ. If φ is (i, i, i)-free, the following formulas are instances of CUGO Principle and PUGO Principle respectively:

$$[Com_{(i,i)}\varphi]O_{(i,i,i)}\varphi \tag{18}$$

$$[Prom_{(i,i)}\varphi]O_{(i,i,i)}\varphi \ . \tag{19}$$

Note that there is no difference between the effect of an act of commanding oneself to see to it that φ and that of an act of promising oneself to see to it that φ in $\mathsf{DMDL^{+}III}$. We have $M_{Prom_{(i,i)}\varphi} = M_{Com_{(i,i)}\varphi}$.

In ordinary cases where different agents are involved, however, we can capture an interesting interplay between acts of commanding and acts of promising in terms of the different effects they have. Consider the contingent obligational dilemma above again. Suppose you decide to obey your guru and write to her that you will join the demonstration in Tokyo. What effects will your letter have?

One obvious effect will be a change in your guru's epistemic states. She now knows that you have received and understood her command. But there is another interesting effect. Let p, q, a, b, and c be understood as in our earlier discussion of this example. We now have:

$$((M_{Prom_{(a,b)}p})_{Com_{(c,a)}q})_{Prom_{(a,c)}q}, s \models_{\mathsf{DMDL^+III}} O_{(a,c,a)}q \ . \quad (20)$$

Thus we have:

$$((M_{Prom_{(a,b)}p})_{Com_{(c,a)}q})_{Prom_{(a,c)}q}, s \models_{\mathsf{DMDL^+III}} O_{(a,c,c)}q \wedge O_{(a,c,a)}q \ . \quad (21)$$

Now it is obligatory upon you to see to it that q not only in her name but also in your own name. You have explicitly committed yourself.

One tempting step here is to take an obligation of the form $O_{(i,j,k)}\varphi$ as representing commitment of the agent k if $k = i$. As we have seen, however, our semantics validates (18). In the extreme case where the commander and the commandee is identical, an act of commanding can generate an obligation of this form. Whether or not this result shows that commitment should not be considered as special kind of obligation but as something distinct from obligation seems to be a very interesting problem. Although we will not discuss this problem in this paper, we note that it will not be very difficult to develop a propositional modal logic which has operators standing for commitments in addition to operators for obligations as far as we keep them independent from each other.

5 The Proof System for DMDL$^+$III

The proof system for DMDL$^+$III can be obtained by adding so-called reduction axioms and necessitation rules for each command operator and each promise operator to the proof system of MDL$^+$III as follows:

Definition 7. *The proof system for* DMDL$^+$III *contains all the axioms and rules of the proof system for* MDL$^+$III, *and in addition, the following axioms and rules:*

(C1) $[Com_{(i,j)}\varphi]p \leftrightarrow p$

(C2) $[Com_{(i,j)}\varphi]\top \leftrightarrow \top$

(C3) $[Com_{(i,j)}\varphi]\neg\psi \leftrightarrow \neg[Com_{(i,j)}\varphi]\psi$

(C4) $[Com_{(i,j)}\varphi](\psi \wedge \chi) \leftrightarrow [Com_{(i,j)}\varphi]\psi \wedge [Com_{(i,j)}\varphi]\chi$

(C5) $[Com_{(i,j)}\varphi]\Box\psi \leftrightarrow \Box[Com_{(i,j)}\varphi]\psi$

(C6) $[Com_{(i,j)}\varphi]O_{(l,m,n)}\psi \leftrightarrow O_{(l,m,n)}[Com_{(i,j)}\varphi]\psi$ *if* $(l,m,n) \neq (j,i,i)$

(C7) $[Com_{(i,j)}\varphi]O_{(j,i,i)}\psi \leftrightarrow O_{(j,i,i)}(\varphi \rightarrow [Com_{(i,j)}\varphi]\psi)$

(C8) $[Com_{(i,j)}\varphi][Prom_{(l,m)}\psi]\chi \leftrightarrow [Prom_{(l,m)}\psi][Com_{(i,j)}\varphi]\chi$ *if* $(l,m,l) \neq (j,i,i)$

(C9) $[Com_{(i,j)}\varphi][Prom_{(l,m)}\psi]\chi \leftrightarrow [Prom_{(l,m)}\varphi][Com_{(i,j)}\psi]\chi$

if $(l,m,l) = (j,i,i)$, *i.e.* $i = j = l = m$

(P1) $[Prom_{(i,j)}\varphi]p \leftrightarrow p$

(P2) $[Prom_{(i,j)}\varphi]\bot \leftrightarrow \bot$

(P3) $[Prom_{(i,j)}\varphi]\neg\psi \leftrightarrow \neg[Prom_{(i,j)}\varphi]\psi$

(P4) $[Prom_{(i,j)}\varphi](\psi \wedge \chi) \leftrightarrow [Prom_{(i,j)}\varphi]\psi \wedge [Prom_{(i,j)}\varphi]\chi$

(P5) $[Prom_{(i,j)}\varphi]\Box\psi \leftrightarrow \Box[Prom_{(i,j)}\varphi]\psi$

(P6) $[Prom_{(i,j)}\varphi]O_{(l,m,n)}\psi \leftrightarrow O_{l,m,n)}[Prom_{(i,j)}\varphi]\psi$ *if* $(l,m,n) \neq (i,j,i)$

(P7) $[Prom_{(i,j)}\varphi]O_{(i,j,i)}\psi \leftrightarrow O_{(i,j,i)}(\varphi \rightarrow [Prom_{(i,j)}\varphi]\psi)$

([Com]-Nec) $\dfrac{\psi}{[Com_{(i,j)}\varphi]\psi}$

([Prom]-Nec) $\dfrac{\psi}{[Prom_{(i,j)}\varphi]\psi}$

The notion of $\mathsf{DMDL}^{+}\mathsf{III}$*-proof and the notion of logical consequence* $\vdash_{\mathsf{DMDL}^{+}\mathsf{III}}$ *can be defined in the obvious way.*

The above axioms can easily be seen to be valid, and the above rules obviously preserve validity. Thus this proof system is sound.

Moreover, the axioms (C1), (C2), (P1), and (P2) allow us to eliminate command operators and promise operators prefixed to propositional letters and $\top$ respectively, and other axioms enable us to reduce the length of sub-formulas to which command operators and promise operators are prefixed. Thus, these axioms enables us to define a translation function that translates any formula of $\mathcal{L}_{\mathsf{DMDL}^{+}\mathsf{III}}$ into a formula of $\mathcal{L}_{\mathsf{MDL}^{+}\mathsf{III}}$ that is provably equivalent to it. Then the completeness of $\mathsf{DMDL}^{+}\mathsf{III}$ is derived from that of $\mathsf{MDL}^{+}\mathsf{III}$.

Theorem 2 (Completeness of $\mathsf{DMDL}^{+}\mathsf{III}$). *Let* $\Sigma \cup \{\varphi\}$ *be a set of* $\mathcal{L}_{\mathsf{DMDL}^{+}\mathsf{III}}$*-formulas. Then, if* $\Sigma \models_{\mathsf{DMDL}^{+}\mathsf{III}} \varphi$ *then* $\Sigma \vdash_{\mathsf{DMDL}^{+}\mathsf{III}} \varphi$.

6 The Comparison with Searle's Treatment of Acts of Promising

In this section we will compare our treatment of acts of promising with Searle's treatment in his argument for the derivability of "ought" from "is" (1964, and 1969). Searle's argument can be considered as consisting of three parts. In the first part, he derives the statement about an institutional fact (ii) below from the factual premise (i) with the help of the constitutive rule (ia) and the empirical assumption (ib) (1969, pp.177–178.):

(i) Jones uttered the words "I hereby promise to pay you, Smith, five dollars".
(ia) Under certain conditions *C* anyone who utters the words (sentence) "I hereby promise to pay you, Smith, five dollars" promises to pay Smith five dollars.
(ib) Conditions *C* obtain.
(ii) Jones promised to pay Smith five dollars.

In the second part, he derives the "evaluative statement" (iii) below from (ii) and (iia), and then derive (iv) from (iii) and (iiia) (pp.178–180.):

(ii) Jones promised to pay Smith five dollars.
(iia) All promises are acts of placing oneself under (undertaking) an obligation to do the thing promised.

(iii) Jones placed himself under (undertook) an obligation to pay Smith five dollars.
(iiia) All those who place themselves under an obligation are (at the time when they so place themselves) under an obligation.
(iv) Jones is under an obligation to pay Smith five dollars.

And finally, in the third part, he derives (v) from (iv) and (iva) (pp.180-181):

(iv) Jones is under an obligation to pay Smith five dollars.
(iva) If one is under an obligation to do something, then as regards that obligation one ought to do what one is under an obligation to do.
(v) As regards his obligation to pay Smith five dollars, Jones ought to pay Smith five dollars.

As (iii) is evaluative, the first two parts of this argument, if sound, have done the substantial work of deriving the evaluative from the factual.

The first part of this argument derives the institutional fact (ii) from the factual premise (i). According to Searle,

> Every institutional fact is underlain by a (system of) rule(s) of the form "*X* counts as *Y* in context *C*". (1969, pp.51-52.)

In the second part, (iii) and (iv) unfold what is involved in the institutional fact (ii) with the help of (iia) and (iiia).[7]

Broadly speaking, PUGO Principle corresponds to the conjunction of (iia) and (iiia), and can vindicate the second part of this argument. Let j, s, r, and (M,t) represent Jones, Smith, the proposition that Jones will pay Smith five dollars, and the situation before Jones uttered the sentence in question respectively. Then we have:

$$M,t \models_{\mathsf{DMDL^{+}III}} [Prom_{(j,s)}r]O_{(j,s,j)}r \ . \tag{22}$$

This is equivalent to:

$$M_{Prom_{(j,s)}r}, t \models_{\mathsf{DMDL^{+}III}} O_{(j,s,j)}r \ . \tag{23}$$

Since $(M_{Prom_{(j,s)}r}, t)$ represents the situation after Jones made his promise, this confirms (iv).

This is not a fortuitous correspondence, as $\mathsf{DMDL^{+}III}$ is designed to incorporate Austin's notion of illocutionary acts as acts producing conventional effects (Austin, 1955, pp.103-104). Searle's treatment of acts of promising inherits much from Austin's. Searle finds Austin's definition of commissives unexceptionable, and includes the category of commissives in his taxonomy as the class of "those illocutionary acts whose point is to commit the speaker $\cdots$ to some future course of action "(Searle 1979, p.14).

[7] Searle refers to "the constitutive rule that to make a promise is to undertake an obligation" in his argument (1969, p.185). But it is not this "rule", which looks like (iia), but (ia) that seems to instantiate the general form of constitutive rules. We may rewrite it as follows: uttering the sentence "I hereby promise to pay you, Smith, five dollars" counts as promising to pay Smith five dollars in context where conditions *C* hold. The notion of institution and "count-as" relation have come to be the topics of lively discussions recently.

Thus Searle's characterization of commissives refers to conventional effects of commissives.

The category of commissives, however, is the only category of Austin's classification that survives in Searle's taxonomy. As CUGO Principle indicates, $\mathsf{DMDL^{+}III}$ treats acts of commanding also as acts producing conventional effects. But according to Searle, the illocutionary point of the category of directives, to which acts of commanding belong, "consists in the fact that they are attempts (of varying degrees, $\cdots$) by the speaker to get the hearer to do something"(Searle 1979, p.13). It doesn't seem to refer to conventional effects at all. Notice that getting the hearer to do something is a perlocutionary act. Thus Searle characterizes directive illocutionary acts as attempts to perform certain perlocutionary acts. In this respect, $\mathsf{DMDL^{+}III}$ is more Austinian than Searle's standard theory.[8]

7 Conclusion

We have refined $\mathsf{MDL^{+}II}$ into $\mathsf{MDL^{+}III}$, and developed a dynamified multi-agent deontic logic $\mathsf{DMDL^{+}III}$ as a dynamic extension of $\mathsf{MDL^{+}III}$. In $\mathsf{DMDL^{+}III}$, acts of promising as well as acts of commanding are characterized as acts producing conventional effects. By modeling them as deontic updators, we have incorporated Austin's notion of illocutionary acts in the limited domain consisting of acts of promising and acts of commanding. $\mathsf{DMDL^{+}III}$ enables us to analyze how a conflict of obligations can be generated jointly by an act of commanding and an act of promising, as well as what effects an act of promising to do what is commanded can have. Moreover, PUGO principle can be used to vindicate part of Searle's argument for derivability of "ought" from "is". As $\mathsf{DMDL^{+}III}$ is proven to be sound, having such a vindication is of considerable significance.

References

Austin, J.L.: How to Do Things with Words. The William James Lectures. Harvard University. In: Urmson, J.O., Sbisà, M. (eds.) How to Do Things with Words, 2nd edn., Harvard University Press, Cambridge (1955)(1975)

Baltag, A., Moss, L.S., Solecki, S.: The Logic of Public Announcements, Common Knowledge, and Private Suspicions. Technical Report TR534. Department of Computer Science (CSCI), Indiana University (1999)

van Benthem, J., Liu, F.: Dynamic Logic of Preference Upgrade. Journal of Applied Non-Classical Logics 17, 157–182 (2007)

Gerbrandy, J., Groeneveld, W.: Reasoning about Information Change. Journal of Logic, Language, and Information 6, 147–169 (1997)

Jeske, D.: Special Obligations. In: Stanford Encyclopedia of Philosophy (2002), http://plato.stanford.edu/entries/special-obligations/

[8] For more on Austin's distinction between illocutionary acts and perlocutionary acts, see Yamada (2002), Yamada (2007c), or Sbisà (2005). In Yamada (2007c), the deontic update of Yamada (2007b) is combined with the preference upgrade of van Benthem & Liu (2007), and used to differentiate illocutionary acts of commanding from perlocutionary acts that affect preferences of addressees.

Kooi, B., van Benthem, J.: Reduction Axioms for Epistemic Actions. In: Schmidt, R., et al. (eds): Preliminary Proceedings of AiML-,: Advances in Modal Logic. Technical Report Series, UMCS-04-9-1, Department of Computer Science, University of Manchester pp. 197-211 (2004)

Plaza, J.A.: Logics of public communications. In: Emrich, M.L., et al. (eds.) Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems, pp. 201–216 (1989)

Sbisà, M.: How to Read Austin. A lecture read at the 9th International Pragmatics Conference, Riva del Garda, July 10–15 (2005)

Searle, J.: How to Derive “Ought” from “Is”. The Philosophical Review (January 1964)

Searle, J.: Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press, Cambridge (1969)

Searle, J.: Expression and Meaning. Cambridge University Press, Cambridge (1979)

Yamada, T.: An Ascription-Based Theory of Illocutionary Acts. In: Vanderveken, D., Kubo, S. (eds.) Essays in Speech Act Theory. Pragmatics & Beyond, New Series, vol. 77, pp. 151–174. John Benjamins, Amsterdam/Philadelphia (2002)

Yamada, T.: Acts of Commanding and Changing Obligations. In: Inoue, K., Satoh, K., Toni, F. (eds.) CLIMA 2006. LNCS (LNAI), vol. 4371, pp. 1–19. Springer, Heidelberg (2007a)

Yamada, T.: Logical Dynamics of Commands and Obligations. In: Washio, T., et al. (eds.) JSAI 2006. LNCS (LNAI), vol. 4384, pp. 133–146. Springer, Heidelberg (2007b)

Yamada, T.: Logical Dynamics of Some Speech Acts that Affect Obligations and Preferences. In: van Benthem, J., Ju, S., Veltman, F. (eds.) A Meeting of Minds. Proceedings of the Workshop on Logic, Rationality and Interaction, Beijing. Texts in Computer Science, vol. 8, pp. 275–289. College Publications, London (2007c)

Dynamic Semantics of Quantified Modal Mu-Calculi and Its Applications to Modelling Public Referents, Speaker's Referents, and Semantic Referents

Norihiro Ogata

Department of Language and Information, Osaka University
1-8 Machikaneyama, Toyonaka, Osaka, Japan
ogata@lang.osaka-u.ac.jp

Abstract. A generalized QG-semantics of Quantified Modal Logics (QMLs) is proposed by exploiting Goldblatt & Marefs [30] Quantified General Frame semantics of QMLs to solve the Kripke-imcompleteness problem with some QMLs. It is extended by adding formulas of modal mu-calculi to model speaker's referents and public referents. Furthermore, dynamic semantics of a quantified modal mu-calculus is formalized based on the generalized QG-semantics.

1 Introduction

This paper will propose *dynamic semantics* of *quantified modal μ-calculi* and its applications to modeling *Public Referents* and *Speaker's Referents* in natural language discourse.

Modal μ-calculi are proposed to model logics of infinite iterations of modal operators as in [1,2,3]. In particular, Alberucci [4] applies modal μ-calculi to modeling *common knowledge* as

(1) $\mathsf{C}_{A,B}\varphi \equiv \nu X.(K_A(X) \wedge K_B(X) \wedge K_A\varphi \wedge K_B\varphi)$

where $\nu X.\varphi[X]$ represents the greatest fixed point of the denotation of $\varphi[X]$ and $K_A\varphi$ means that agent A knows that φ.

I will extend modal μ-calculi to quantified modal μ-calculi to model *public referents*, which is firstly proposed in this paper, and *Speaker's referents*, which is proposed by Kripke [5], in discourse. *Public referents* are *public* in the sense that all the participants can refer them freely, whereas *Speaker's referents* cannot be referred by the other than the speaker him/herself.

See the following examples of dialogue:

Example 1 ([6]).

(2) H: A magistrate from the Gotham village has confessed batting young girls.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 109–122, 2008.

N: They say *he* suspected them of sorcery. Do you know if more magistrates confessed?
H: I don't know.
N: Do you know who *he* is?
H: No idea, he preferred to remain anonymous.

Apparently, H and N agree to be talking about one and the same magistrate, perhaps one they read about in the newspapers. However it is not at all clear from example 1 which individual is concerned, or which definite must be used to refer to it. But clearly we can understand H and N's utterances, without knowing which definite term, if any, is to replace the indefinite term "a magistrate of the Gotham village."

Example 2 ([7] p. 113). In the following example by Parsons [7], the pronoun *he* is used for denoting what A and C do not know what it is:

(3) A: The man in the doorway over there looks pretty silly.
C: But there is no man in the doorway over there.
A: (Looks again.) Oh! I thought there was; I was wrong.
B: Does *he* looks anything like your department chairman?
A: *Who*?
B: The man in the doorway over there.
A: There isn't any man there; I was mistaken about that.
B: Well, *he* doesn't exist, but *he*'s there, isn't *he*?
A: (Exasperated.) Look, I was talking about a guy who exists; that is I thought I was, but I was wrong, I wasn't talking about anybody. I can't tell you what *"he"* looks like because there is no *"he"* to look like anything!

Example 3. This example is quoted from a scene in the film *The Beautiful Mind*, which is a story of the famous mathematician, John Nash and his wife, Alicia. John has been annoyed by a kind of mental sick, which causes delusions, i.e., he sometimes sees delusional people. *William Parcher* is one of them. But John believes that William Parcher is real. Then the following conversation occurs:

(4)
John: You have to find *William Parcher*.
Alicia: Stop. ... There is *no William Parcher*.
John: Of course there is. I've been working for him. ...
Alicia: There is *no William Parcher*. *It*'s in your mind.

Therefore, the following property must not be held:

$$\mathsf{C}_{A,B}\exists x.\varphi \rightarrow \exists x.\varphi$$

$$\mathsf{C}_{A,B}\forall x.\varphi \leftrightarrow \forall x.\mathsf{C}_{A,B}\varphi$$

whereas (1) allows these properties. This implies that a usual logics of common knowledge is not appropriate for modeling *public referents*. Instead of (1), I will propose the following formalization of *public knowledge*:

(5) $\nu X.(\Box_A X \wedge \Box_B X \wedge \exists x\varphi)$

where $\Box_x$ is K4 and does not satisfy T, i.e., $\Box_x\varphi \rightarrow \varphi$. In formula (6), x represents a *public referent* and X the public situation. As in example 3, public referents sometimes do not have their semantic referents and as in example 3, public referents sometimes do not correspond to speakers' intended referents. Therefore, the following formula may be true in dialogues:

(6) $(\nu X.(\Box_A X \wedge \Box_B X \wedge \exists x\varphi)) \wedge \Box_A(x = c) \wedge \Box_B(\neg Ex \wedge \Box_A Ex)$

where Ex means that x actually exists, c is the speaker's referent of A but it is not believed by agent B and x is dynamically bound by subformulas $\Box_A(x = c)$ and $\Box_B(\neg Ex \wedge \Box_A Ex)$. Even X must be dynamically bound by subsequent subformulas as in

(7) $(\nu X.(\Box_A X \wedge \Box_B X \wedge \exists x\varphi)) \wedge \Box_A(X \wedge \psi(x))$

which must be logically equivalent to the following formula under dynamic interpretation of operator ν:

(8) $\nu X.(\Box_A X \wedge \Box_B X \wedge \exists x\varphi \wedge \Box_A(X \wedge \psi(x))$

To sum up,

(9) a. common knowledge logics such as [8,9] including modal μ-calculi [4] are too strong to treat *public referent*, since sometimes they do not have their semantic referents.
b. dynamic interpretation of quantified modal μ-calculi can treat update of public information or weaker common knowledge and public referents, while it can preserve some kind of noncoincidence between public referents and speakers' referents.

Before treating Quantified μ-calculi, we must see the current rough situation of QMLs (Quantified Modal Logics) which are the basis of Quantified μ-calculi. Traditionally, QMLs are classified into the three basic systems (see [10,11]):

- **Q1** ... all the domains are fixed and all the individual constants are rigid designators, i.e., $\forall w, u(u \neq w) \in W.\mathcal{D}(w) = \mathcal{D}(u) = D$, $I_c(c)(w) = I_c(c)(u)$ and $w, g \models \forall x\varphi \Leftrightarrow \forall a \in D.w, g[a/x] \models \varphi$, where $g \in D^{Var}$. **Q1** is axiomatized by rules and axioms such as UI, ∀1, ∀2, UG, BF, I, rISub, rI¬, i.e.:
  - UI: $\forall x\varphi \rightarrow \varphi[t/x]$
  - ∀1: $\forall x(\varphi_1 \rightarrow \varphi_2) \rightarrow \forall x\varphi_1 \rightarrow \forall x\varphi_2$
  - ∀2: $\varphi \rightarrow \forall x\varphi$ (x is not a free variable in φ)
  - BF: $\forall x\Box\varphi \rightarrow \Box\forall x\varphi$
  - I: $t = t$
  - rISub: $t_1 = t_2 \rightarrow (\varphi[t_1/x] \rightarrow \varphi[t_2/x])$
  - rI¬: $\neg t_1 = t_2 \rightarrow \Box\neg t_1 = t_2$
  - UG: $\dfrac{\varphi}{\forall x\varphi}$

– **Q1R** (Schurz [11] calls **Q2**)... all the domains may be varied and all the individual constants are rigid designators and quantification is interpreted as free quantification, i.e., $I_c(c)(w) = I_c(c)(u)$ and $w, g \models \forall x\varphi \Leftrightarrow \forall a \in \mathcal{D}(w).w, g[a/x] \models \varphi$, where $g \in D^{Var}$. **Q1R** is axiomatized by rules and axioms of **Q1** minus UI and BF but plus fUI and fUG or other rules (see [10,11,12]), which sometimes leads the systems to Free Logics, to avoid the truth-value gap, i.e.:
  - fUI: $\forall x\varphi \rightarrow (Et \rightarrow \varphi[t/x])$
  - fUG: $\dfrac{Ex \rightarrow \varphi}{\forall x\varphi}$
– **Q3** ... all the domains may be varied and all the individual constants are non-rigid and have counterparts: e.g.,

$$w, g \models \Box\varphi \Leftrightarrow \forall u : wRu \& h \sim_{w,u} g \Rightarrow u, h \models \varphi$$

where $range(h)$ is replaced by counterparts of $range(g)$ at u. **Q3** does not include rI¬ nor others (See [13] for the complete axiomatization of **Q3**). Its semantics requires some drastic changes of standard Quantified Kripke semantics, anyway: e.g., the functor semantics [14], the (general) metaframe semantics [15,16,17], the (general) counterpart-frame semantics [18], world-line semantics [11], and so on.
– **Q2** ... all the domains may be varied and consisted of individual concepts. One of the notorious properties of **Q2** is that $\Box\exists x\varphi \rightarrow \exists x\Box\varphi$ is validated in the usual **Q2** systems. Furthermore, according to Thomason [19], all systems of **Q2** are not recursively axiomatizable. However, Kracht & Kutz [20], Aloni [21] and others try to reformulate **Q2** as axiomatizable and complete systems.

However, recent surveys or handbooks such as [22,23,24,25,17] as well as the classic introductions [26,27] by Hughes and Cresswell rather classify QMLs with respect to choices or combinations of constant/varying domains, domains of quantification, individual concepts, counterparts, rigidity/non-rigidity, and so on. Although there have been many articles on the completeness of some systems of QMLs, these results makes us to feel QMLs still unmatured as an easy analysis tool. Furthermore, QMLs are notorious in the sense of the *Kripke-incompleteness*.[1] I will not treat these difficult intrinsic technical or philosophical problems here. Therefore, the road map of this paper is as follows:

1. Introducing a general static semantics of QMLs by Goldblatt & Mares [30], since it is a quite nice and simple general semantics of QMLs.
2. Expanding the Goldblatt-Mares semantics to a semantics which can deal with individual concepts and counterparts and furthermore to a semantics of Quantified μ-calculi
3. Lifting the expanded Goldblatt-Mares semantics to dynamic semantics.

[1] Every **Q1L** or **Q1RL** such that for any $\mathbf{L} \supseteq \mathbf{S4}$, $\mathbf{S5} \not\subseteq \mathbf{L}$ and $\mathbf{L} \not\subseteq \mathbf{S4.3}$ is incomplete with respect to standard Kripke semantics [28,14,29].

2 Preliminary Definitions

Before an introduction of the Goldblatt-Mares semantics, I introduce the definitions of preliminary concepts required to understand the Goldblatt-Mares semantics, as follows:

Definition 1. *Let $\mathscr{P}$ be the covariant functor [31] over category* Set*, i.e.,*

- $\mathscr{P}(X)$ *is the powerset of* X *for any set* X*;*
- $\mathscr{P}(f)(X) = \{f(x)|x \in X\}$ *for any function* $f : X \to Y$ *for some set* Y*;*

Let $\mathscr{P}_+(X) = \mathscr{P}(X) \setminus \emptyset$.

Definition 2. *Let* $x \in Var$ *an individual variable,* $X \in PVar$ *a propositional variable,* $c \in Con$ *an individual constant including agents symbols* $A_1, \ldots, A_n \in Ag$*, and* $R \in Rel(n)$ *an* n*-ary relation symbol* ($Rel = \bigcup_{n\in\omega} Rel(n)$)*, and* $t \in Var \cup Con$*. Then formula* $\varphi \in \mathscr{L}$ *is defined by the following BNF grammar:*

$$\varphi ::= R(t_1, \ldots, t_n)|\varphi_1 \wedge \varphi_2|\exists x\varphi|\neg\varphi|\Box_A\varphi.$$

Formula $\phi \in \mathscr{L}^+$ *is the smallest set of formula including* $\mathscr{L}$ *and* $PVar$ *such that:*

- $t_1 = t_2$, $E(t_1) \in \mathscr{L}_+$,
- *If* $\phi_1, \phi_2 \in \mathscr{L}^+$*, then* $\phi_1 \wedge \phi_2, \neg\phi_1, \Box_A\phi \in \mathscr{L}_+$*, and*
- *If* $X \in PVar$ *occurs within an even scope of negation in* $\phi \in \mathscr{L}_+$*, then* $\nu X.\phi \in \mathscr{L}_+$*, where an even scope of negation in* ϕ *w.r.t.* X *means that* X *is a subformula of* ϕ *and* ϕ *has an even number of negation as its parts,*

where $E(t)$ *means that the denotation of* t *actually exists.*

3 Goldblatt and Mares' General Semantics of QMLs

Now I introduce the Goldblatt-Mares semantics, a. k. a. the *Quantified general frame semantics (QG-frame semantics)*, which is proposed by Goldblatt and Mares [30] as an alternative to Kripke semantics of QMLs, as follows. The basic idea is a lifting a usual interpretation $|\varphi| \in Prop \in \mathscr{P}\mathscr{P}(W)$ for each $\varphi \in \mathscr{L}$ to $|\varphi| \in PropFun \subseteq (D^{Var} \to Prop)$.

Definition 3 ([30]). *Let* W *be a non-empty set of possible worlds,* D *a non-empty set of individuals,* R_A *a binary relation between* W *for each* $A \in Ag$*,* R_A *a binary relation between* W *such that* $\langle W, (R_A)_{A\in Ag}\rangle$ *is a Kripke frame,* $[\Box_A] = \lambda p \in Prop.\{w \in W|\forall u : wR_Au \Rightarrow u \in p\}$*,* $Prop \in \mathscr{P}(\mathscr{P}(W))$ *such that:*

- $\emptyset, W \in Prop$
- *If* $p, q \in Prop$*, then* $p \cap q, W \setminus p, [\Box_A]p \in Prop$*,*

where $\mathsf{MA} = \langle Prop, \cap, W \setminus \cdot, ([\Box_A]_{A\in Ag}), W, \emptyset\rangle$ *is a modal algebra.*

Let PropFun be a set of functions from D^{Var} to $\mathscr{P}(W)$. Then

$$\mathcal{Q} = \langle W, (R_A)_{A \in Ag}, D, PropFun \rangle$$

is called a quantified general frame and $|\cdot|^{\mathcal{M}} : \mathscr{L} \to PropFun$ *is an interpretation w.r.t. a QG-model* $\mathcal{M} = \langle \mathcal{Q}, V \rangle$*, which is defined by recursion on* φ*, as follows:*

- $[R(t_1, \ldots, t_n)]^{\mathcal{M}} = (\lambda g \in D^{Var}.V^1(R)(||t_1||^{\mathcal{M}} g, \ldots, ||t_n||^{\mathcal{M}} g)) \in PropFun$
- $[\varphi_1 \wedge \varphi_2]^{\mathcal{M}} = (\lambda g \in D^{Var}.[\varphi_1]g \cap [\varphi_2]g)$
- $[\Box_A \varphi]^{\mathcal{M}} = [\Box_A]([\varphi]^{\mathcal{M}})$
- $[\exists x \varphi] = (\lambda g \in D^{Var}.\bigsqcup_{a \in D}[\varphi](g[a/x]))$,

where $\bigsqcup S = \bigcap\{X \in Prop | \bigcup S \subseteq X\}$, $V = \langle V^1, V^2 \rangle$, $V^1(Rel(n)) \in (D^n \to PropFun)$*, for each* $c \in Con$, $V^2(c) \in D$*, and*

$$||t||^{\mathcal{M}} g = \begin{cases} g(t) & \text{if } t \in Var \\ V^2(t) & \text{if } t \in Con \end{cases}$$

A triple $s = \langle \mathcal{M}^s, w^s, g^s \rangle$ *is called a(n evaluation) state if* $\mathcal{M}^s$ *is a QG-model,* $w^s \in W$*, and* $g^s \in D^{Var}$ *(a variable assignment).*

$\varphi \in \mathscr{L}$ *is true at s, written* $s \models \varphi$*, if and only if* $w^s \in [\varphi]g^s$*.*

This semantics only treats the constant domain of QMLs, i.e., if $\mathcal{D} \in (W \to \mathscr{P}_+(D))$ then for all $w, u (\neq w) \in W$, $\mathcal{D}(w) = \mathcal{D}(u) = D$. The next definition expands this semantics to deal with relative domains, domains consisting of individual concepts, and domains with a counterpart relation.

Definition 4. *Let* $\mathcal{Q}' = \langle W, (R_A)_{A \in Ag}, D, Prop, PropFun' \rangle$ *be a QG-frame minus PropFun plus* $PropFun' \subseteq (D^W \to Prop)$, $\mathcal{D} \in (W \to \mathscr{P}_+(D^W))$ *such that* $\mathcal{D}(w)(u) = \{bu \in D | b \in \mathcal{D}(w)\}$ *and* $\mathcal{D}(w)(u)$ *is defined if* $w = u$, $V = \langle V^1, V^2 \rangle$ *such that:*

- $V^1(w)(R) \subseteq \{(b_1 w, \ldots, b_n w) \in \mathcal{D}(w)(w) | (b_1, \ldots, b_n) \in \mathcal{D}(w)^n\}$ *for each* $R \in Rel(n)$ *and* $w \in W$,
- $V^2(c)(w) \in \mathcal{D}(w)(w)$ *for each* $c \in Con$,

$\mathcal{C}_{w,u} \subseteq \mathcal{D}(w) \times \mathcal{D}(u)$ *is a counterpart relation.*

$$\mathcal{C}_{w,u}[g] = \{h \in D^W \quad | \forall x \in Var. \exists b.b = h(u)(x) \& (g(w)(x), b) \in \mathcal{C}_{w,u}\},$$

for all $w, u \in W$ *and* $g \in D^W$ *such that* $g(w)(x) \in \mathcal{D}(w)$ *(called a generalized variable assignment).*

$\mathcal{D}$ *decides the set of individuals which live in each possible world and* $\mathcal{C}$ *decides a counterpart, i.e., if* $(a, b) \in \mathcal{C}_{w,u}$ *then a counterpart in u of an individual a who lives in w is b.*

Then

$$\mathcal{M} = \langle \mathcal{Q}, \mathcal{C}, V \rangle$$

is called a generalized QG-model. The interpretation $|\cdot|^{\mathcal{M}} : \mathscr{L} \to PropFun'$ *based on a generalized QG-model* $\mathcal{M}$ *is defined by recursion on* φ*, as follows:*

- $[R(t_1, \ldots, t_n)]^{\mathcal{M}} =$
  $(\lambda g \in D^W \quad .\{w \in W | (||t_1||^{\mathcal{M}}(g, w), \ldots, ||t_n||^{\mathcal{M}}(g, w)) \in V^1(w)(R)\})$

- $[\varphi_1 \wedge \varphi_2]^{\mathcal{M}} = (\lambda g \in D^W . [\varphi_1]^{\mathcal{M}} g \cap [\varphi_2]^{\mathcal{M}} g)$
- $[\Box_A \varphi]^{\mathcal{M}} = \lambda g \in D^W . \{u \in W | w R_A u \forall h \in \mathcal{C}_{w,u}[g] \Rightarrow ([\varphi]^{\mathcal{M}}(h,u))\}$
- $[\exists x \varphi] = (\lambda g \in D^W . \bigsqcup_{w \in W, b \in \mathcal{D}(w)(w)} [\varphi]^{\mathcal{M}}(g[w : b/x]))$,

where

$$g[w : b/x](u)(x')(u) = \begin{cases} b & \text{if } w = u \text{ and } x' = x \\ g(u)(x')(u) & \text{otherwise} \end{cases}$$

and $||t||^{\mathcal{M}} \in D^W \times W$ *such that*

$$||t||^{\mathcal{M}}(g, w) = \begin{cases} g(w)(t)(w) & \text{if } t \in Var \\ V^2(w)(t) & \text{if } t \in Con \end{cases}$$

A triple $s = \langle \mathcal{M}^s, w^s, g^s \rangle$ *is called a(n evaluation) state if* $\mathcal{M}^s$ *is a QG-model,* $w^s \in W$*, and* $g^s \in D^W$ *(a variable assignment).*

$\varphi \in \mathscr{L}$ *is true at* s*, written* $s \models \varphi$*, if and only if* $w^s \in [\varphi](g^s, w^s)$*.*

Constant domain models, varying-domain models, individual concept models, and Lewisian counterpart models (see Lewis [32]) are treated as follows, respectively:

- if for all $w, u (\neq w) \in W$, $\mathcal{D}(w) = \mathcal{D}(u)$, $\mathcal{C}_{w,u}[g] = \{g\}$, $g(w) = g(u)$, $V^i(w) = V^i(u)$ $(i \in \{1,2\})$, and for all $w \in W$, each $b \in \mathcal{D}(w)$ is a constant function such that $b = \lambda w \in W.\mathbf{b}$ for some $\mathbf{b} \in D$, then $\mathcal{M}$ is called *constant-domained*.
- if for all $w, u (\neq w) \in W$, $\mathcal{D}(w) \neq \mathcal{D}(u)$, $\mathcal{C}_{w,u}[g] = \{g\}$, $V^i(w) = V^i(u)$ $(i \in \{1,2\})$ and for all $w \in W$, each $b \in \mathcal{D}(w)$ is a constant function such that $b = \lambda w \in W.\mathbf{b}$ for some $\mathbf{b} \in D$, then $\mathcal{M}$ is called *varying-domained*.
- if for all $w, u (\neq w) \in W$, $\mathcal{D}(w) \neq \mathcal{D}(u)$, $\mathcal{C}_{w,u}[g] = \{g\}$, $g(w) \neq g(u)$, $V^i(w) \neq V^i(u)$ $(i \in \{1,2\})$ and for all $w \in W$, each $b \in \mathcal{D}(w)$ is a function called an individual concept (e.g., $b(w)$ is called a stage of b at w), then $\mathcal{M}$ is called *individual-concept-domained*.[2]
- if for all $w, u (\neq w) \in W$, $\mathcal{D}(w) \cap \mathcal{D}(u) = \emptyset$, $g \notin \mathcal{C}_{w,u}[g]$, $g(w) \neq g(u)$, $V^i(w) \neq V^i(u)$ $(i \in \{1,2\})$ ($\{h(u)(x)(u) | h \in \mathcal{C}_{w,u}[g]\}$ are called the counterparts of $g(w)(x)(w)$ at u), then $\mathcal{M}$ is called *counterpart-domained*.

4 A Generalized QG-Semantics of $\mathscr{L}^+$

If we expand a generalized QG-model $\mathcal{M}$ for $\mathscr{L}^+$, we must define $|X|^{\mathcal{M}}$, $|E(t)|^{\mathcal{M}}$, $|t_1 = t_2|^{\mathcal{M}}$, and $|\nu X(\varphi)|^{\mathcal{M}}$. The definitions of $|E(t)|^{\mathcal{M}}$, $|t_1 = t_2|^{\mathcal{M}}$ are routine, though the restrictions on $|E(t)|^{\mathcal{M}}$ and $|t_1 = t_2|^{\mathcal{M}}$ reflex the philosophy of QMLs: e.g., the actualist semantics requires that $V^1(w)(E) = \mathcal{D}(w)(w)$, whereas the possibilist semantics requires that $V^1(w)(E) = D$, for each $w \in W$; if $V^1(w)(=) \neq V^1(u)(=)$ then the equality means a contingent equality, otherwise

[2] However, this semantics treats individual-concept domained interpretations of individual terms not as individual concepts but as their *stages*.

the equality means the necessary equality which validates $a = b \to \Box(a = b)$, for each $w, u(\neq w) \in W$. However, later, in the definition of the dynamic semantics of $E(x)$, to treat the examples of section 1, we should treat the case that the value of x is undefined but not false, i.e., $\Box_A \neg E(x)$ for each agent A. To solve this problem, I expand each domain $\mathcal{D}(w)$ for each world w to $\mathcal{D}_+(w)$ such that:

$$\mathcal{D}_+(w) = \mathcal{D}(w)(w) \cup \{\lambda w.\bot_n | \bot_n \in \bot_{\mathcal{D}(w)}\}$$

where

$$\bot_{\mathcal{D}} = \{\bot_n | n \in \mathbb{N}\}$$

$$\bot_{\mathcal{D}(w)} \subseteq \bot_{\mathcal{D}}$$

$\bot_n$ is an undefined object for each $n \in \mathbb{N}$ and f assigns to each individual variable such an undefined object which is distinguished by the assigned individual variable.

Furthermore, I introduce variable assginments which assign undefined objects to each variables such as:

$$\{f | f(x)(w) \in \bot_{\mathcal{D}(w)} \text{ for all } x \in Var\}.$$

f does not avoid the accidental identity between the undefined objects which are assigned to different variables. f is updated by $\exists x$ in the dynamic semantics.

Similarly, I introduce function f' such that for each constant c, $f'(c)(w) \in \bot_{\mathcal{D}(w)}$.

$t_1 = t_2$, $E(t)$, $\exists x\varphi$, X and $\nu X(\varphi)$ require an expansion of generalized QG-models, as follows:

Definition 5. *Let $\mathcal{M} = \langle W, (R_A)_{A \in Ag}, D, Prop, PropFun' \rangle, \mathcal{D}, \mathcal{D}_+, \mathcal{C}, V^1, V^2, f' \rangle$ be a generalized QG-model, $X \in PVar$, and $G \in PropFun'^{PVar}$ an assignment to propositional variables.*

Then $\mathcal{M}_+ = \langle \mathcal{M}, f, G \rangle$ is called a generalized QG_ν-model.

The interpretation $|\cdot|^{\mathcal{M}_+} : \mathscr{L}_+ \to PropFun'$, called a QG_ν-semantics, is defined by recursion on $\varphi \in \mathscr{L}_+$, as follows (only the cases of $t_1 = t_2$, $E(t)$, X and $\nu X(\varphi)$):

- $[t_1 = t_2]^{\mathcal{M}_+} = (\lambda(g, f) \in D^W \times \bot_{\mathcal{D}}^{Var} .\{w \in W : ||t_1||^{\mathcal{M}_+}(g, w, f, f') = ||t_n||^{\mathcal{M}_+}(g, w, f, f')\})$
- $[E(t)]^{\mathcal{M}_+} = (\lambda(g, f) \in D^W \times \bot_{\mathcal{D}}^{Var} .\{w \in W : ||t||(g, w, f, f') \in \mathcal{D}_+(w)(w)\})$
- $[\exists x\varphi] = (\lambda(g, f) \in D^W \times \bot_{\mathcal{D}}^{Var} .\bigsqcup_{w \in W, b \in \mathcal{D}_+(w)(w)} [\varphi]^{\mathcal{M}}(g, f)),$
- $[X]^{\mathcal{M}_+} = G(X)$
- $[\nu X(\varphi)]^{\mathcal{M}_+} =$
  $(\lambda(g, f) \in D^W \times \bot_{\mathcal{D}}^{Var} .\bigcup\{P(g, f) \in Prop | Y(g, f) \subseteq [\varphi]^{\mathcal{M}_+[P/X]}(g, f)\}),$
  where $\mathcal{M}_+[P/X] = \langle \mathcal{M}, G[P/X] \rangle$.

where $||t||^{\mathcal{M}_+}$ *such that*

$$||t||^{\mathcal{M}_+}(g,w,f,f') = \begin{cases} g(w)(t)(w) & \textit{if } t \in Var \textit{ and } g(w)(t)(w) \in \mathcal{D}(w)(w) \\ f(t)(w) & \textit{if } t \in Var \textit{ and } g(w)(t)(w) \notin \mathcal{D}(w)(w) \\ V^2(w)(t) & \textit{if } t \in Con \textit{ and } V^2(w)(t) \in \mathcal{D}(w)(w) \\ f'(t)(w) & \textit{if } t \in Con \textit{ and } V^2(w)(t) \notin \mathcal{D}(w)(w) \end{cases}$$

$[E(t)]^{\mathcal{M}_+} = \{w \in W | f(t)(w) \in \mathcal{D}_+(w)(w) \text{ or } f'(t)(w) \in \mathcal{D}_+(w)(w)\}$ if $||t||(g,w,f,f') \notin \mathcal{D}(w)(w)$ for all $w \in W$. This implies that the interpretation of $E(t)$ *weakly* fails w.r.t. the interpretation world, i.e., it has an interpretation which consists of only undefined objects, but it is not undefined. This fact realizes that $\Box_A \neg E(t)$ for all names or pronouns t represents that agent A that the denotation of x is not actual from her viewpoint. Therefore, if t refers to agent B's referent (a *speaker's referent*), i.e., $\Box_B E(t)$, the occurrence of t in $\Box_A \neg E(t)$ can refer to the speaker's referent and deny its existence simultaneously. A *public referent* can be represented as an occurrence of t in formula of form $\nu X(\Box_A(X \wedge \varphi[t]) \wedge \Box_B(X \wedge \psi[t]))$. On the other hand, a semantic referent of term t is the value of t in an interpretation.

Lemma 1. *Let* $gfp(F)$ *is the greatest fixed point of a map* F. *Then*

$$[\nu X(\varphi)]^{\mathcal{M}_+}(g,f) = gfp(\lambda Y.[\varphi]^{\mathcal{M}_+[Y/X]}(g,f)).$$

Proof. By the Knaster-Tarski theorem, if $\lambda Y.[\varphi]^{\mathcal{M}_+[Y/X]}(g,f))$ is monotone, $\lambda Y.[\varphi]^{\mathcal{M}_+[Y/X]}(g,f))$ has its gfp, i.e., $Y_1 \subseteq Y_2$ implies $\lambda Y.[\varphi]^{\mathcal{M}_+[Y/X]}(g,f))(Y_1) \subseteq \lambda Y.[\varphi]^{\mathcal{M}_+[Y/X]}(g,f))(Y_2)$. This is assured by the definition of $\mathscr{L}_+$, since all $X \in PVar$ in each $\varphi \in \mathscr{L}_+$ occurs in an even scope of negation in φ. Furthermore, $[\nu X(\varphi)]^{\mathcal{M}_+}(g,f) = (\lambda(g,f) \in D^W \times \perp_{\mathcal{D}}^{Var} .\bigcup\{P(g,f) \in Prop | Y(g,f) \subseteq [\varphi]^{\mathcal{M}_+[P/X]}(g,f)\} = \bigcup\{P(g,f) \in Prop | Y(g,f) \subseteq [\varphi]^{\mathcal{M}_+[P/X]}(g,f)\}$ is the greatest fixed point of $\lambda Y.[\varphi]^{\mathcal{M}_+[Y/X]}(g,f))$ by the Knaster-Tarski theorem. □

Theorem 1. *A* QG_ν*-model validates* $\mathscr{L}^+$ *as a modal* μ*-calculus.*

Proof. Directly from Lemma 1. □

5 A Dynamicization of QG_ν-Semantics of $\mathscr{L}^+$

To realize a dynamic update of propositional variable assignments, I exploit the *CPS (continuation passing style) transformation* (see [33]) into our dynamic semantics, as follows:

Definition 6. *Let* $p, q \in Prop$. *Then a* λ*-notation* P *on functions on Prop and elements of Porp is translated into CPS* λ*-notation* $\underline{P}$ *by recursion on* P, *as follows:*

$$\underline{p} = \lambda k.k(p)$$

$$\underline{p \cap q} = \lambda h.\underline{p}(\lambda m.\underline{q}(\lambda n.h(m \cap n)))$$

$$\underline{\cap q} = \lambda f.\lambda h.f(\lambda m.\underline{q}(\lambda n.h(m \cap n)))$$

$$\underline{p \cap q} = \underline{\cap q}(\underline{p})$$

$\underline{G}$ is a propositional variable assignment such that $G : PVar \to \underline{Prop}$, where $\underline{Prop} = \{\underline{p}|p \in Prop\}$.

$$\underline{\mathcal{M}_+} = \langle \mathcal{M}, \underline{G} \rangle$$

For example, $\underline{(p \cap q) \cap p'}$ has a sound β-reduction, as follows:

$$\begin{array}{ll}
 & \underline{(p \cap q) \cap p'} \\
\to_\beta & (\underline{\cap p'}(\underline{\cap q}(\underline{p}))) \\
\to_\beta & (\underline{\cap p'}(\lambda f.\lambda h.f(\lambda m.\underline{q}(\lambda n.h(m \cap n))))(\underline{p})) \\
\to_\beta & (\underline{\cap p'}(\lambda h.\underline{p}(\lambda m.\underline{q}(\lambda n.h(m \cap n))))) \\
\to_\beta & (\underline{\cap p'}(\lambda h.(\lambda k.k(p)(\lambda m.\underline{q}(\lambda n.h(m \cap n)))))) \\
\to_\beta & (\underline{\cap p'}(\lambda h.(\lambda m.\underline{q}(\lambda n.h(m \cap n)))p)) \\
\to_\beta & (\underline{\cap p'}(\lambda h.(\underline{q}(\lambda n.h(p \cap n))))) \\
\to_\beta & (\underline{\cap p'}(\lambda h.(\lambda k.k(q)(\lambda n.h(p \cap n))))) \\
\to_\beta & (\underline{\cap p'}(\lambda h.(\lambda n.h(p \cap n)q))) \\
\to_\beta & (\underline{\cap p'}(\lambda h.h(p \cap q))) \\
\Rightarrow_{def} & (\underline{\cap p'}(\underline{(p \cap q)})) \\
\to_\beta & (\lambda f.\lambda h.f(\lambda m.\underline{p'}(\lambda n.h(m \cap n)))(\underline{(p \cap q)})) \\
\to_\beta & (\lambda h.\underline{(p \cap q)}(\lambda m.\underline{p'}(\lambda n.h(m \cap n)))) \\
\to^*_\beta & (\lambda h.h((p \cap q) \cap p') \\
\Rightarrow_{def} & \underline{((p \cap q) \cap p')}
\end{array}$$

Given a QG_ν-model $\mathcal{M}_+$, $w \in W$, $g \in D^W$, let S be a non-empty set of tuples $s = (\mathcal{M}_+, w, g)$, called a state. By lifting $[\varphi]^{\mathcal{M}_+} \in PropFun'$ to $[\![\varphi]\!] : S \to \mathscr{P}(S)$, a dynamic semantics of $\mathscr{L}^+$ is defined, as follows:

Definition 7. *Let $s = \langle \underline{\mathcal{M}_+}^s, w^s, g^s, f^s \rangle \in S$ be a state.*

$$||t||s \doteq \begin{cases} g(w^s)(t)(w^s) & \textit{if } t \in Var \textit{ and } g(w)(t)(w) \in \mathcal{D}(w^s)(w^s) \\ f^s(t)(w^s) & \textit{if } t \in Var \textit{ and } g(w^s)(t)(w^s) \notin \mathcal{D}(w^s)(w^s) \\ V^2(w^s)(t) & \textit{if } t \in Con \end{cases}$$

$$s[\varphi/X] \doteq \langle \langle \mathcal{M}^s, \underline{G^s}[\underline{\underline{G^s}(X) \cap [\varphi]^{\mathcal{M}_+} g^s w^s}\rangle, w^s, g^s, f^s \rangle,$$

$$\underline{\mathcal{M}_+}[\varphi/X] \doteq \langle \mathcal{M}^s, \underline{G^s}[\underline{\underline{G^s}(X) \cap [\varphi]^{\mathcal{M}_+} g^s w^s}\rangle,$$

$$s[\varphi/x] \doteq \{\langle \underline{\mathcal{M}^s_+}, w^s, g^{s'}, f^s \rangle | s' \in [\![\varphi]\!]s\},$$

$$s[f/f^s] \doteq \langle \underline{\mathcal{M}^s_+}, w^s, g^s, f \rangle$$

and

$$s[g/g^s]\dot{=}\langle\underline{\mathcal{M}^s_+}, w^s, g, f^s\rangle$$

Then dynamic interpretation $[\![\cdot]\!]$ *of* $\mathscr{L}^+$ *is defined by recursion on* φ*: Let* X *be activated;*

- $[\![t_1 = t_2]\!] = (\lambda s.\textbf{if } w^s \in [t_1 = t_2](g^s, f^s) \textbf{ then } \{s[t1 = t_2/X]\} \textbf{ else } \emptyset,$
- $[\![E(t)]\!] = (\lambda s.\textbf{if} w^s \in [E(t)](g^s, f^s) \textbf{ then } \{s[E(t)/X]\}; \textbf{else } \emptyset,$
- $[\![R(t_1, \ldots, t_n)]\!] =$ $(\lambda s.\textbf{if } w^s \in [R(t_1, \ldots, t_n)](g^s, f^s) \textbf{ then } \{s[R(t_1, \ldots, t_n)/X]\} \textbf{ else } \emptyset)$
- $[\![\Box_A\varphi]\!] = (\lambda s.\textbf{if for\_all } u \in R_A(s) \textbf{ for\_all } h \in \mathcal{C}_{w\ ,u}[g^s]\ [\![\varphi]\!]\langle\underline{\mathcal{M}^s_+}, u, h\rangle \neq \emptyset \textbf{ then } \{s[\Box_A\varphi/X][\Box_A\varphi/x])\}$, *where* $R_A(s) = \{u \in W|w_s R_A u\}$
- $[\![\exists x\varphi]\!] = (\lambda s. \bigcup_{b\in\mathcal{D}_+(w\ )(w\ )} [\![\varphi]\!](s')$ *where* $s' = \begin{cases} (\underline{\mathcal{M}_+}[\exists x\varphi/X], w^s, g^s[w : b/x], f^s) & \text{if } b \in \mathcal{D}(w)(w) \\ (\underline{\mathcal{M}_+}[\exists x\varphi/X], w^s, g^s, f^s[w : b/x]) & \text{otherwise} \end{cases}$
- $[\![X]\!] = (\lambda s.\textbf{if } w^s \in \underline{G^s}(X)(\lambda p.p) \textbf{ then } \{s[\underline{G^s}(X)(\lambda p.p)/X]\} \textbf{ else } \emptyset)$
- $[\![\varphi_1 \wedge \varphi_2]\!] = (\lambda s. \bigcup_{s'\in[\![\varphi_1]\!](s[\varphi_1/X])} [\![\varphi_2]\!](s'[\varphi_2/X]))$
- $[\![\neg\varphi]\!] = (\lambda s.\textbf{if } w^s \notin [\varphi] \textbf{ then } \{s[\neg\varphi/X]\} \textbf{ else } \emptyset)$
- $[\![\nu X(\varphi)]\!] = (\lambda s.\textbf{if } w^s \in gfp(\lambda p.|\varphi|^{\mathcal{M}_+[p/X]}(g^s, f^s))$ $\textbf{then } \{s[\underline{gfp(\lambda p.[\varphi]^{\mathcal{M}^{[\ \ ]}_+}(g^s, f^s))}/X][\varphi/x]\} \textbf{ else } \emptyset)$ *(X is reactivated)*

In the above definition of $[\![X]\!]$, $\underline{G^s}(X)(\lambda p.p)$ means a kind of halting the 'continuation' at only this point temporally. For a more detailed explanation, see [33].

Furthermore, if the occurence of $x \in Var$ in formula $\exists x\varphi[x]$ is *undefined* at state s in the sense that $g^s(w^s)(t)(w^s) \notin \mathcal{D}(w^s, w^s)$, then the undefined object assignment f is instead updated. This means that the undefined object in formulas of form $\exists x\varphi$ is also updated as well as the case of that x is defined by a variable assignment g, and therefore, the interpretation of sequences of formulas which they include even $\exists$ makes it possible to keep talking about an undefined object without a denotational conflict.

This dynamic semantics of language $\mathscr{L}$ satisfies the requirements for dealing the public referents and speakers' referents and update of shared beliefs, which described in (9).

Definition 8 (Weak dynamic deduction). *Let* $\varphi_1, \ldots, \varphi_n$ *be a sequence of* $\mathscr{L}^+$*-formulas and* φ *an* $\mathscr{L}^+$*-formula.*

$$\varphi_1, \ldots, \varphi_n \models_S \varphi \Leftrightarrow \forall s, t \in S.t \in [\![\varphi_n]\!] \circ \cdots \circ [\![\varphi_1]\!](s) \Rightarrow t \in [\![\varphi]\!](s) \neq \emptyset.$$

Theorem 2. *Let* S *be a set of states and* νX *activates* X *which is a subformula of* ψ*. Then,*

$$\nu X(\varphi) \wedge \psi \models_S \nu X(\varphi) \wedge \psi[\nu X(\varphi)/X]).$$

Proof. (Sketch)

$$s' \in [\![\nu X(\varphi) \wedge \psi[X]]\!](s) \quad \text{(hyp.)}$$

$$\Leftrightarrow s' \in \bigcup_{s'' \in [\![\nu X(\varphi)]\!](s)} [\![\psi[X]]\!](s'') \quad \text{(def.)}$$

$$\Leftrightarrow s' \in \bigcup_{t \in [\![\varphi]\!]\langle\langle \mathcal{M}, \underline{G'}\rangle, w^{s}, g^{s}\rangle} [\![\psi[X]]\!](\langle\langle \mathcal{M}, \underline{G'}\rangle, w^s, g^t\rangle) \quad \text{(def.)}$$

$$\text{where } \underline{G'} = \underline{G^s}[\underline{gfp(\lambda p.|\varphi|^{\mathcal{M}^{[\ ]}_{+}} g^s w^s)}/X]$$

$$\Rightarrow s' \in \bigcup_{t \in [\![\varphi]\!]\langle\langle \mathcal{M}, \underline{G'}\rangle, w^{s}, g^{s}\rangle} [\![\psi[\nu X(\varphi)/X]]\!](\langle\langle \mathcal{M}, \underline{G'}\rangle, w^s, g^t\rangle)$$

$$\because \underline{G'}(X)(\lambda p.p) = \underline{gfp(\lambda p.|\varphi|^{\mathcal{M}^{[\ ]}_{+}} g^s w^s)}(\lambda p.p)$$

$$gfp(\lambda p.|\varphi|^{\mathcal{M}^{[\ ]}_{+}} g^s w^s) = |\nu X(\varphi)|^{\langle \mathcal{M}, G'\rangle} g^s w^s \qquad \square$$

This theorem is enough strong to treat the examples in section 1, since the all formulas except X and $\nu X(\varphi)$ are not 'updates' but simply 'tests' in the sense of dynamic semantics.

6 Concluding Remarks

As we have seen, I have shown a generalized QG-semantics of QMLs and its extension by adding formulas of modal mu-calculi to solve the problem of Kripke-incompleteness of QMLs and proposed a dynamic semantics of a quantified modal mu-calculus based on the generalized QG-semantics.

The dynamic semantics consists of the *powerset monad* $\langle \mathscr{P}, \{\cdot\}, \bigcup \rangle$ (see Moggi [34]). This means that it is possible that we can provide more abstract dynamic semantics using category theory, monads, and λ_c-calculi [34]. This is a future work. Furthermore, I have connected the update of G (a propositional-variable-assignment) in the dynamic semantics with *continuation*, which can also be treated as the *continuation monad* in category theory and has its λ_c-calculi. This technique can lead us to stronger theorem:

$$\nu X(\varphi) \wedge \psi \models_S \nu X(\varphi \wedge \psi).$$

than theorem 2 Therefore, exploiting computational category theory is promising for more elegant and powerful formalizations of dynamic semantics. This is also a future work.

References

1. Ambler, S., Kwiatkowska, M.Z., Measor, N.: Duality and the completeness of the modal μ-calculus. Theoretical Computer Science 151, 3–27 (1995)
2. Bonsangue, M.M., Kwiatkowska, M.Z.: Re-interpreting the modal μ-calculus. In: Ponse, A., de Rijke, M., Venema, Y. (eds.) Modal Logic and Process Algebra: A Bisimulation Perspective, CSLI, Stanford, pp. 65–83 (1995)
3. Hartonas, C.: Duality for modal μ-logics. Theoretical Computer Science 202, 193–222 (1998)

4. Alberucci, L.: The Modal μ-Calculus and the Logic of Common Knowledge. PhD thesis, Universität Bern (2002)
5. Kripke, S.A.: Speaker's reference and semantic reference. Midwest Studies in Philosophy 2, 255–276 (1977)
6. Dekker, P., van Rooy, R.: Intentional identity and information exchange. In: Gamkrelidze, R.C.T. (ed.) Proceedings of the Second Tbilisi Symposium on Language, Logic and Computation, Tbilisi State University, Tbilisi (1998)
7. Parsons, T.: Nonexistent Objects. Yale University Press, New Heaven (1980)
8. Kaneko, M.: Common knowledge logic and game logic. The Journal of Symbolic Logic 64, 685–700 (1999)
9. Kaneko, M., et al.: A map of common knowledge logics. Studia Logica 71, 57–86 (2002)
10. Garson, J.W.: Quantification in modal logic. In: Gabbay, D., Guenthner, F. (eds.) Handbook of Philosophical Logic, vol. II, pp. 249–307. D. Reidel Publishing Company, Dordrecht (1984)
11. Schurz, G.: Alethic modal logic and semantics. In: Jacquette, D. (ed.) A Companion to Philosophical Logic, pp. 442–477. Blackwell, Oxford (2001)
12. Corsi, G.: A unified completeness theorem for quantified modal logics. Journal of Symbolic Logic 67, 1483–1510 (2002)
13. Thomason, R.H.: Some completeness results for modal predicate calculi. In: Lambert, K. (ed.) Philosophical Problems in Logic: Some Recent Develeopments, pp. 56–76. D. Reidel Publishing Company, Dordrecht (1969)
14. Ghilardi, S.: Incompleteness results in Kripke semantics. Journal of Symbolic Logic 56, 517–538 (1991)
15. Skvortsov, D., Shetman, V.: Maximal Kripke-type semantics for modal and superintuitionistic predicate logics. Annals of Pure and Applied Logic 63, 69–101 (1993)
16. Shirasu, H.: Duality in superintuitionistic and modal predicate logics. In: Balbiani, P., et al. (eds.) Advance of Modal Logic, CSLI, Stanford, vol. I, pp. 223–236 (1998)
17. Gabbay, D.M., Shetman, V., Skvortsov, D.: Quantifiation in Nonclassical Logic (Draft: March 13, 2007). Elsevier Science, Amsterdam (2007)
18. Kracht, M., Kutz, O.: The semantics of modal predicate logic i: Counterpart frames. In: Wolter, F., et al. (eds.) Advances in Modal Logics, vol. 3, pp. 299–320. World Scientific Publishing Co. Pte. Ltd., Singapore (2002)
19. Thomason, R.H.: Modal logic and metaphysics. In: Lambert, K. (ed.) The Logical Way of Doing Thing, pp. 119–146. D. Reidel Publishing Company, Dordrecht (1969)
20. Kracht, M., Kutz, O.: The semantics of modal predicate logic ii: Modal individuals revisited. In: Kahle, R. (ed.) Intensionality, pp. 60–96. Association for Symbolic Logic, San Diego (2005)
21. Aloni, M.: Individual concepts in modal predicate logic. Journal of Philosophical Logic 34, 1–64 (2005)
22. Fitting, M., Mendelsohn, R.L.: First-Order Modal Logic. Kluwer Academic Publishers, Dordrecht (1999)
23. Kracht, M., Kutz, O.: Logically possible worlds and counterpart semantics of modal logic. In: Jacquette, D., et al. (eds.) Philosophy of Logic, pp. 943–996. North Holland, Amsterdam (2006)
24. Garson, J.W.: Modal Logic for Philosophers. Cambridge University Press, Cambridge (2006)

25. Braüner, T., Ghilardi, S.: First-order modal logic. In: Blackburn, P., van Benthem, J., Wolter, F. (eds.) Handbook of Modal Logic, pp. 549–620. Elsevier, Amsterdam (2007)
26. Hughes, G.E., Cresswell, M.J.: An Introduction to Modal Logic. Methuen, London (1968)
27. Hughes, G.E., Cresswell, M.J.: A New Introduction to Modal Logic. Routledge, London (1996)
28. Ghilardi, S.: Presheaf semantics and independence results for some non-classical first-order logics. Archive for Mathematical Logic 29, 125–136 (1989)
29. Cresswell, M.J.: Incompleteness and the Barcan formula. Journal of Philosophical Logic 24, 379–403 (1995)
30. Goldblatt, R., Mares, E.D.: A general semantics for quantified modal logic. In: Governatori, G., Hodkinson, I., Venema, Y. (eds.) Advances in Modal Logics, vol. 6, pp. 227–246. Colledge Publications, London (2006)
31. Mac Lane, S.: Categories for the Working Mathematician. Springer, Berlin (1998)
32. Lewis, D.: Counterpart theory and quantified modal logic. Journal of Philosophy 65, 113–126 (1968) (counterpart)
33. Sørensen, M.H., Urzyczyn, P.: Lectures on the Curry-Howard Isomorphism. Springer, Berlin (2006)
34. Moggi, E.: Notions of computation and monads. Information and Computation 93, 55–92 (1991)

Inverse Scope as Metalinguistic Quotation in Operational Semantics

Chung-chieh Shan

Rutgers University
ccshan@rutgers.edu

Abstract. We model semantic interpretation *operationally*: constituents interact as their combination in discourse evolves from state to state. The states are recursive data structures and evolve from step to step by context-sensitive rewriting. These notions of *context* and *order* let us explain inverse-scope quantifiers and their polarity sensitivity as metalinguistic quotation of the wider scope.

1 Introduction

An utterance is both an action and a product [1]. On one hand, when we use an utterance, it acts on the world: its parts affect us and our conversation. For example, a referring expression reminds us of its referent and adjusts the discourse context so that a pronoun later may refer to it. On the other hand, linguistics describes the syntax and semantics of a sentence as a mathematical object: a product, perhaps a tree or a function. But trees and functions do not remind or adjust; at their most active, they merely contain or map. How then does the meaning of an utterance determine its effects on the discourse and its participants? In particular, how do utterances affect their *context*, and in the *order* that they do?

We approach this "mind-body problem" the way many programming-language semanticists approach an analogous issue. On one hand, a program such as

(1) $$2 \to n;\ n \cdot 3$$

expresses a sequence of actions: store the number 2 in the memory location n, then retrieve the contents of n and multiply it by 3. On the other hand, computer science describes the syntax and semantics of the program as a tree or a function. To view the same program as both a dynamic action and a static product, programming-language theory uses *operational semantics* alongside denotational semantics. A denotational semantics assigns a denotation to each expression—compositionally, by specifying the denotation of a complex expression in terms of the denotations of its parts. In contrast, an operational semantics defines a *transition relation* on states—for each state, what it can become after one step of computation [2]. Technical conditions relate denotational and operational semantics. For example, *adequacy* requires that two utterances with the same denotation must have the same operational outcome.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 123–134, 2008.

To analyze the example (1), we could define a state as an ordered pair of an integer n and a current program, and the transition relation $\rightsquigarrow$ as a set that includes the three transitions

$$(2) \qquad \langle 0,\ (2 \to n;\ n \cdot 3)\rangle \rightsquigarrow \langle 2,\ n \cdot 3\rangle \rightsquigarrow \langle 2,\ 2 \cdot 3\rangle \rightsquigarrow \langle 2,\ 6\rangle.$$

These transitions together say that the program (1) starting with the memory location n initialized to 0 yields the final result 6 with the contents of n changed to 2. Unlike in denotational semantics, we specify the outcome of a program without recursively specifying what each part of the program denotes. Rather, we specify rewriting rules such as

$$(3) \qquad \langle x,\ (y \to n;\ P)\rangle \rightsquigarrow \langle y,\ P\rangle \quad \text{for any numbers } x, y \text{ and any program } P$$

to be elements of the transition relation, so that the first computation step in (2) goes through. In other words, we include

$$(4) \quad \bigl\{\, \bigl\langle \langle x,\ (y \to n;\ P)\rangle,\ \langle y,\ P\rangle \bigr\rangle \mid x \text{ and } y \text{ are two numbers; } P \text{ is a program} \,\bigr\}$$

in the transition relation as a subset.

Following a longstanding analogy between natural and formal languages, we model natural-language semantics operationally. In Section 2, we review a standard operational semantics for a λ-calculus with *delimited control*. In Section 3, we use this operational semantics to explain quantification and polarity sensitivity. We focus on these applications because they go beyond phenomena such as anaphora and presupposition, where dynamic semantics has long been fruitful. In particular, metalinguistic quotation (or *multistage programming*) extends our account to inverse scope. We conclude in Section 4.

2 Context and Order

We use the λ-calculus to introduce the notions of *context* and *order*, then let an expression interact actively with its context. These computational notions recur in various linguistic guises, especially in analyses of quantification, as Section 3 makes concrete.

2.1 Expressions and Contexts

The set of λ-calculus expressions is recursively defined by the following grammar. In words, an expression is either a variable x, an abstraction $\lambda x.\, E$, or an application EE.

$$(5) \qquad E \to x \qquad\qquad E \to \lambda x.\, E \qquad\qquad E \to EE$$

The derivation below shows that $\lambda x.\, xx$ is an expression, even though no function x can apply to itself in standard set theory.

$$(6) \qquad E \to \lambda x.\, E \to \lambda x.\, EE \to \lambda x.\, Ex \to \lambda x.\, xx$$

In programming practice, constants such as 2 and multiplication $\cdot$ are also expressions. Not shown in the grammar is the convention that we equate expressions that differ only by variable renaming, so $\lambda x.\, \lambda x.\, x = \lambda x.\, \lambda y.\, y = \lambda y.\, \lambda x.\, x \neq \lambda x.\, \lambda y.\, x$.

We specify certain expressions to be *values*, namely those generated by the following grammar. In words, a value is any variable or abstraction, but not an application.

(7) $$V \to x \qquad\qquad V \to \lambda x.\, E$$

For example, the identity expression $\lambda x.\, x$ (henceforth I) is a value. In practice, constants such as 2 and multiplication $\cdot$ are also values. Intuitively, a value is an expression that is done computing, so the expression $2 \cdot 3$ is not a value.

An operational semantics is a relation on states. (More precisely, we consider *small-step* operational semantics.) Our states are just closed expressions (that is, expressions with no unbound variables), and the transition relation is in fact a partial function that maps each closed expression to what it becomes after one step of computation. We define the transition relation as follows. A *context* is an expression with a gap, which we write as $[\,]$. For example, $\lambda x.\, x(y[\,])$ is a context. The set of *evaluation contexts* $C[\,]$ is defined by the grammar below, in the style of [3]. The notation $C[\dots]$ means to replace the gap $[\,]$ in the context $C[\,]$ by an expression or another context.

(8) $$C[\,] \to [\,] \qquad C[\,] \to C\big[[\,]E\big] \qquad C[\,] \to C\big[V[\,]\big]$$

The first production above says that the null context $[\,]$—the context with nothing but a gap—is an evaluation context. The second production says that, whenever $C[\,]$ is an evaluation context, replacing its gap by $[\,]E$ (that is, the application of a gap to any expression E) gives another evaluation context. The third production says that, whenever $C[\,]$ is an evaluation context, replacing its gap by $V[\,]$ (that is, the application of any value V to a gap) gives another evaluation context. For example, the derivation below shows that $I\big([\,](II)\big)$ is an evaluation context. (Recall from above that I stands for $\lambda x.\, x$.)

(9) $$C[\,] \to C\big[[\,](II)\big] \to C\big[I\big([\,](II)\big)\big] \to I\big([\,](II)\big)$$

We now define the transition relation $\rightsquigarrow$ to be the set of expression-pairs

(10) $$\{\, \langle C[(\lambda x.\, E)V],\, C[E']\rangle \mid C[\,] \text{ is an evaluation context; } E' \text{ substitutes the value } V \text{ for the variable } x \text{ in the expression } E \,\},$$

or for short,

(11) $$C[(\lambda x.\, E)V] \rightsquigarrow C[E\,\{x \mapsto V\}].$$

In words, a transition is a λ-conversion in an evaluation context where the argument is a value expression. Letting $C[\,] = I\big([\,](II)\big)$, $V = I$, and $E = x$ gives

(12) $$I\big((II)(II)\big) \rightsquigarrow I\big(I(II)\big).$$

In fact, this is the only step that a computation starting at $I\big((II)(II)\big)$ can take. It takes three more steps to reach a value, namely I:

$$I\big(I(II)\big) \rightsquigarrow I(II) \rightsquigarrow II \rightsquigarrow I. \tag{13}$$

In practice, λ-conversions are joined by other transitions such as $1+2\cdot3 \rightsquigarrow 1+6$.

This operational semantics is not the only reasonable one for the λ-calculus. Instead of or in addition to the transition (12), $I\big((II)(II)\big)$ could transition to $(II)(II)$ or $I\big((II)I\big)$. Different operational semantics regulate the *order* in which to run parts of a program differently, by constraining the set of evaluation contexts and the rewriting that takes place inside. Our transition relation in (10)–(11) implements a *call-by-value*, *left-to-right* evaluation order: it only performs λ-conversion when the argument is a value (the V in (10)–(11) rules out a transition to $(II)(II)$) and everything to the left is also a value (the V in (8) rules out a transition to $I\big((II)I\big)$).

2.2 Delimited Control

The simple expression $I\big((II)(II)\big)$ above is not sensitive to the evaluation order: pretty much any reasonable order will bring it to the final value I after a few transitions. However, as we add functionality to the λ-calculus as a programming language, different operational semantics result in different program outcomes. A particularly powerful and linguistically relevant addition is *delimited control* [3, 4], which lets an expression manipulate its context. To illustrate, we add 's delimited-control operators *shift* and *reset* [5, 6, 7] to the λ-calculus.

Before specifying shift and reset formally, let us first examine some example transition sequences among arithmetic expressions. Reset by itself does not perform any computation, so the programs $1 + 2 \cdot 3$ and $1 + \text{reset}(2 \cdot 3)$ yield the same final value:

$$1+2\cdot3 \rightsquigarrow 1+6 \rightsquigarrow 7, \tag{14}$$

$$1+\text{reset}(2\cdot3) \rightsquigarrow 1+\text{reset}\,6 \rightsquigarrow 1+6 \rightsquigarrow 7. \tag{15}$$

Shift means to *remove* the surrounding context—up to the nearest enclosing reset—into a variable. This functionality lets an expression manipulate its context. For example, the variable f below receives the context that multiplies by 2, so the program computes $1 + 3 \cdot 2 \cdot 2 \cdot 5$.

$$\begin{aligned} &1+\text{reset}\big(2\cdot \text{shift}\, f.\, (3\cdot f(f(5)))\big) \\ &\quad\rightsquigarrow 1+\text{reset}\big(3\cdot(\lambda x.\,\text{reset}(2\cdot x))((\lambda x.\,\text{reset}(2\cdot x))(5))\big) \\ &\quad\rightsquigarrow 1+\text{reset}\big(3\cdot(\lambda x.\,\text{reset}(2\cdot x))(\text{reset}(2\cdot 5))\big) \\ &\quad\rightsquigarrow 1+\text{reset}\big(3\cdot(\lambda x.\,\text{reset}(2\cdot x))(\text{reset}\,10)\big) \\ &\quad\rightsquigarrow 1+\text{reset}\big(3\cdot(\lambda x.\,\text{reset}(2\cdot x))10\big) \\ &\quad\rightsquigarrow 1+\text{reset}\big(3\cdot\text{reset}(2\cdot 10)\big) \rightsquigarrow 1+\text{reset}(3\cdot\text{reset}\,20) \\ &\quad\rightsquigarrow 1+\text{reset}(3\cdot 20) \rightsquigarrow 1+\text{reset}\,60 \rightsquigarrow 1+60 \rightsquigarrow 61 \end{aligned} \tag{16}$$

To take another example, the shift expression below does not use the variable f and so discards its surrounding context and supplies the reset with the result 4 right away.

(17) $$1 + \text{reset}\big(2 \cdot 3 \cdot (\text{shift}\, f.\, 4) \cdot 5\big) \rightsquigarrow 1 + \text{reset}\, 4 \rightsquigarrow 1 + 4 \rightsquigarrow 5$$

We call reset a *control delimiter* because it delimits how much context an enclosed shift expression manipulates.

For concreteness, the rest of this subsection formalizes the delimited-control operators shift and reset; it can be skipped if the examples above suffice. We add two productions for expressions.

(18) $$E \to \text{reset}\, E \qquad\qquad E \to \text{shift}\, f.\, E$$

The expression "$\text{shift}\, f.\, E$" binds the variable f in the body E, so for instance the expressions "$\text{shift}\, f.\, f$" and "$\text{shift}\, x.\, x$" are equal because they differ only by variable renaming.

We add one production for evaluation contexts.

(19) $$C[\,] \to C\big[\text{reset}[\,]\big]$$

We call $D[\,]$ a *subcontext* if it is an evaluation context built without this new production. Finally, we add two new kinds of transitions to our transition relation, one for reset and one for shift:

(20) $\{\langle C\big[\text{reset}\, V\big], C\big[V\big]\rangle \mid C[\,]$ is an evaluation context and V is a value$\}$,

(21) $\{\langle C\big[\text{reset}\, D[\text{shift}\, f.\, E]\big], C\big[\text{reset}\, E'\big]\rangle$
$\mid C[\,]$ is an evaluation context; $D[\,]$ is a subcontext;
E' substitutes $\lambda x.\, \text{reset}\, D[x]$ for the variable f
in the expression E, where x is a fresh variable$\}$,

or for short,

(22) $$C\big[\text{reset}\, V\big] \rightsquigarrow C\big[V\big],$$

(23) $$C\big[\text{reset}\, D[\text{shift}\, f.\, E]\big] \rightsquigarrow C\big[\text{reset}\, E\, \{f \mapsto \lambda x.\, \text{reset}\, D[x]\}\big].$$

3 Linguistic Applications

Linguistic theory traditionally views syntax, semantics, and pragmatics as a pipeline, in which semantics maps expressions to denotations. We envision a semantic theory for natural language that is operational in the sense that it specifies transitions among representations rather than mappings to denotations. That is, whereas denotational semantics interprets a syntactic constituent (such as a verb phrase) by mapping it to a separate semantic domain (such as of functions), operational semantics interprets a constituent by rewriting it to other constituents. The rewriting makes sense insofar as the constituents represent real objects. Thus we may specify a fragment of operational semantics as a set of states, a transition relation over the states, and an ideally trivial translation from utterances to states.

3.1 Quantification

We now use delimited control to model quantification. We assume that the sentence

(24) Somebody saw everybody

translates to the program (state)

$$\text{reset}\Big(\big(\overbrace{\text{shift}\, f.\, \exists x.\, fx}^{\textit{somebody}}\big)^{<}\big(\overbrace{\text{saw}}^{\textit{saw}}{}^{>}\big(\overbrace{\text{shift}\, g.\, \forall y.\, gy}^{\textit{everybody}}\big)\big)\Big), \tag{25}$$

where $^<$ and $^>$ indicate backward and forward function application. The occurrences of shift in this translation arise from the lexical entries for the quantifiers *somebody* and *everybody*, whereas the occurrence of reset is freely available at every clause boundary. This program then makes the following transitions to yield the surface-scope reading of (25).

$$\begin{aligned}
&\rightsquigarrow \text{reset}\big(\exists x.\, (\lambda x.\, \text{reset}(x^{<}(\text{saw}^{>}\, \text{shift}\, g.\, \forall y.\, gy)))x\big)\\
&\rightsquigarrow \text{reset}\big(\exists x.\, \text{reset}(x^{<}(\text{saw}^{>}\, \text{shift}\, g.\, \forall y.\, gy))\big)\\
&\rightsquigarrow \text{reset}\big(\exists x.\, \text{reset}(\forall y.\, (\lambda y.\, \text{reset}(x^{<}(\text{saw}^{>} y)))y)\big)\\
&\rightsquigarrow \text{reset}\big(\exists x.\, \text{reset}(\forall y.\, \text{reset}(x^{<}(\text{saw}^{>} y)))\big)\\
&\rightsquigarrow \text{reset}\big(\exists x.\, \text{reset}(\forall y.\, x^{<}(\text{saw}^{>} y))\big)\\
&\rightsquigarrow \text{reset}\big(\exists x.\, \forall y.\, x^{<}(\text{saw}^{>} y)\big) \rightsquigarrow \exists x.\, \forall y.\, x^{<}(\text{saw}^{>} y).
\end{aligned}$$

The last three transitions assume that the final formula and its subformulas are values.

This example illustrates that the context of a shift operator is the scope of a quantifier, and the order in which expressions are evaluated is that in which they take scope. This analogy is appealing because it extends to other apparently noncompositional phenomena [8, 9, 10, 11] and tantalizing because it links meaning to processing.

3.2 Polarity Sensitivity

We now extend the model above to capture basic polarity sensitivity: a polarity item such as *anybody* needs to take scope (right) under a polarity licensor such as existential *nobody*.

(26) Nobody saw anybody.

(27) Nobody saw everybody.

(28) *Somebody saw anybody.

Following [12, 13], we treat a polarity license ℓ as a dynamic resource [14] that is produced by *nobody*, required by *anybody*, and disposed of implicitly. To this end, we translate *nobody* to

$$\text{shift}\, f.\, \neg\exists x.\, fx\ell, \tag{29}$$

translate *anybody* to

$$\text{(30)} \qquad \operatorname{shift} g.\, \lambda\ell.\, \exists y.\, gy\ell,$$

and add license-disposal transitions of the form

$$\text{(31)} \qquad C[V\ell] \rightsquigarrow C[V].$$

In (30), $\lambda\ell$ denotes a function that must take the constant ℓ as argument.

The sentence (26) translates and computes as follows. The penultimate transition disposes of the used license using (31).

$$\text{(32)} \qquad \begin{aligned} &\operatorname{reset}\Big(\big(\overbrace{\operatorname{shift} f.\, \neg\exists x.\, fx\ell}^{\textit{nobody}}\big)^{<}\big(\overbrace{\text{saw}}^{\textit{saw}}{}^{>}\big(\overbrace{\operatorname{shift} g.\, \lambda\ell.\, \exists y.\, gy\ell}^{\textit{anybody}}\big)\big)\Big) \\ &\rightsquigarrow \operatorname{reset}\big(\neg\exists x.\, (\lambda x.\, \operatorname{reset}(x^{<}(\text{saw}^{>} \operatorname{shift} g.\, \lambda\ell.\, \forall y.\, gy\ell)))x\ell\big) \\ &\rightsquigarrow \operatorname{reset}\big(\neg\exists x.\, \operatorname{reset}(x^{<}(\text{saw}^{>} \operatorname{shift} g.\, \lambda\ell.\, \forall y.\, gy\ell))\ell\big) \\ &\rightsquigarrow \operatorname{reset}\big(\neg\exists x.\, \operatorname{reset}(\lambda\ell.\, \forall y.\, (\lambda y.\, \operatorname{reset}(x^{<}(\text{saw}^{>} y)))y\ell)\ell\big) \\ &\rightsquigarrow \operatorname{reset}\big(\neg\exists x.\, (\lambda\ell.\, \forall y.\, (\lambda y.\, \operatorname{reset}(x^{<}(\text{saw}^{>} y)))y\ell)\ell\big) \\ &\rightsquigarrow \operatorname{reset}\big(\neg\exists x.\, \forall y.\, (\lambda y.\, \operatorname{reset}(x^{<}(\text{saw}^{>} y)))y\ell\big) \\ &\rightsquigarrow \operatorname{reset}\big(\neg\exists x.\, \forall y.\, \operatorname{reset}(x^{<}(\text{saw}^{>} y))\ell\big) \\ &\rightsquigarrow \operatorname{reset}\big(\neg\exists x.\, \forall y.\, x^{<}(\text{saw}^{>} y)\ell\big) \\ &\rightsquigarrow \operatorname{reset}\big(\neg\exists x.\, \forall y.\, x^{<}(\text{saw}^{>} y)\big) \rightsquigarrow \neg\exists x.\, \forall y.\, x^{<}(\text{saw}^{>} y) \end{aligned}$$

Similarly for (27), but not for (28): the transitions for (28) get stuck, because $\lambda\ell$ in (30) needs a license and does not get it.

As in Fry's analyses, our translations of *nobody* and *anybody* integrate the scope-taking and polarity-licensing aspects of their meanings. Consequently, *nobody* must take scope *immediately* over one or more occurrences of *anybody* in order to license them. Essentially, (29)–(31) specify a finite-state machine that accepts strings of scope-taking elements in properly licensed order. It is easy to incorporate into the same system other scope-taking elements, such as the positive polarity item *somebody* and its surprising interaction with *everybody* [15, 16, 17]: surface scope is possible in (33) but not the simpler sentence (34).

(33) Nobody introduced everybody to somebody.

(34) Nobody introduced Alice to somebody.

3.3 Quotation

A deterministic transition relation predicts wrongly that quantifier scope can never be ambiguous. In particular, the transition relation above enforces call-by-value, left-to-right evaluation and thus forces quantifiers that do not contain each other to take surface scope with respect to each other. We refine our empirical predictions using the notion of metalinguistic quotation, as expressed in operational semantics by *multistage programming* [18, inter alia].

Briefly, multistage programming makes three new constructs available in programs: *quotation*, *splice*, and *run*. Quoting an expression such as $1+2$, notated $\lceil 1+2 \rceil$, turns it into a static value, a piece of code. If x is a piece of code, then it can be spliced into a quotation, notated $\lfloor x \rfloor$. For example, if x is $\lceil 1+2 \rceil$, then $\lceil \lfloor x \rfloor \times \lfloor x \rfloor \rceil$ is equivalent to $\lceil (1+2) \times (1+2) \rceil$. Finally, $!x$ notates running a piece of code. The transitions below illustrate.

(35)
$$\begin{aligned}&\big(\lambda x.\, !\lceil \lfloor x \rfloor \times \lfloor x \rfloor \rceil\big)\big(\lceil 1+2 \rceil\big)\\ &\quad\rightsquigarrow\ !\lceil \lfloor \lceil 1+2 \rceil \rfloor \times \lfloor \lceil 1+2 \rceil \rfloor \rceil\\ &\quad\rightsquigarrow\ !\lceil (1+2) \times \lfloor \lceil 1+2 \rceil \rfloor \rceil \rightsquigarrow\ !\lceil (1+2) \times (1+2) \rceil\\ &\quad\rightsquigarrow (1+2) \times (1+2) \rightsquigarrow 3 \times (1+2) \rightsquigarrow 3 \times 3 \rightsquigarrow 9\end{aligned}$$

We need to add to our transition relation the general cases of the second, third, and fourth transitions above. We omit these formal definitions, which involve augmenting evaluation contexts too.

We contend that inverse scope is an instance of multistage programming. Specifically, we propose that the sentence (24) has an inverse-scope reading because it translates not just to the program (25) but also to the multistage program

(36)
$$\text{reset}!\Big\lceil \text{reset}\Big(\overbrace{\big(\text{shift}\, f.\, \exists x.\, fx\big)}^{\textit{somebody}}{}^{<}\big(\overbrace{\text{saw}}^{\textit{saw}}{}^{>}\overbrace{\big\lfloor \text{shift}\, g.\, \forall y.\, g\lceil y \rceil \big\rfloor}^{\textit{everybody}}\big)\Big)\Big\rceil.$$

This latter program makes the following transitions, even under left-to-right evaluation.

$$\begin{aligned}&\rightsquigarrow \text{reset}\big(\forall y.\, \big(\lambda y.\, \text{reset}!\big\lceil \text{reset}((\text{shift}\, f.\, \exists x.\, fx)^{<}(\text{saw}^{>}\lfloor y \rfloor))\big\rceil\big)\lceil y \rceil\big)\\ &\rightsquigarrow \text{reset}\big(\forall y.\, \text{reset}!\big\lceil \text{reset}((\text{shift}\, f.\, \exists x.\, fx)^{<}(\text{saw}^{>}\lfloor \lceil y \rceil \rfloor))\big\rceil\big)\\ &\rightsquigarrow \text{reset}\big(\forall y.\, \text{reset}!\big\lceil \text{reset}((\text{shift}\, f.\, \exists x.\, fx)^{<}(\text{saw}^{>}y))\big\rceil\big)\\ &\rightsquigarrow \text{reset}\big(\forall y.\, \text{reset}\,\text{reset}((\text{shift}\, f.\, \exists x.\, fx)^{<}(\text{saw}^{>}y))\big)\\ &\rightsquigarrow \text{reset}\big(\forall y.\, \text{reset}\,\text{reset}(\exists x.\, (\lambda x.\, \text{reset}(x^{<}(\text{saw}^{>}y)))x)\big)\\ &\rightsquigarrow \text{reset}\big(\forall y.\, \text{reset}\,\text{reset}(\exists x.\, \text{reset}(x^{<}(\text{saw}^{>}y)))\big)\\ &\rightsquigarrow \text{reset}\big(\forall y.\, \text{reset}\,\text{reset}(\exists x.\, x^{<}(\text{saw}^{>}y))\big)\\ &\rightsquigarrow \text{reset}\big(\forall y.\, \text{reset}(\exists x.\, x^{<}(\text{saw}^{>}y))\big)\\ &\rightsquigarrow \text{reset}\big(\forall y.\, \exists x.\, x^{<}(\text{saw}^{>}y)\big) \rightsquigarrow \forall y.\, \exists x.\, x^{<}(\text{saw}^{>}y)\end{aligned}$$

The intuition behind this translation is that the scope of *everybody* in the inverse-scope reading of (24) is metalinguistically quoted. The program (36) may be glossed as "Everybody y is such that the sentence *Somebody saw* y is true", except it makes no sense to splice a person into a sentence, but it does make sense to splice the quotation $\lceil y \rceil$ in (36) into a sentence [19].

Several empirical advantages of this account of inverse scope, as opposed to just allowing non-left-to-right evaluation, lie in apparently noncompositional phenomena other than (but closely related to) quantification. In particular, we explain why a polarity licensor cannot take inverse scope over a polarity item [20, 13].

(37) *Anybody saw nobody.

Surface scope is unavailable for (37) simply because *anybody* must take scope (right) under its licensor. All current accounts of polarity sensitivity capture this generalization, including that sketched in Section 3.2. A more enduring puzzle is why inverse scope is also unavailable. Intuitively, our analysis of inverse scope rules out (37) because it would gloss it as "Nobody y is such that the sentence *Anybody saw* y is true" but *Anybody saw* y is not a well-formed sentence.

Formally, we hypothesize that quotation only proceeds by enclosing the translation of clauses in $\text{reset!}\lceil I_t\,\text{reset}\ldots\rceil$, where I_t is an identity function restricted to take proposition arguments only. In other words, the only transition from a program of the form $C[I_t V]$, where $C[\ ]$ is any evaluation context, is to $C[V]$ when V is a proposition (rather than a function such as $\lambda\ell.\ldots$). We hypothesize I_t not because the operational semantics forces us to, but to express the intuition (some might say stipulation) that quotation applies to propositions only.

Replacing $\text{reset!}\lceil\text{reset}\ldots\rceil$ in (36) by $\text{reset!}\lceil I_t\,\text{reset}\ldots\rceil$ does not hamper the transitions there, because $\exists x.\, x^{<}(\text{saw}^{>}y)$ is a proposition. In contrast, even though (37) translates successfully to the program

$$\text{(38)}\quad \text{reset!}\big\lceil I_t\,\text{reset}\big((\overbrace{\text{shift}\,f.\,\lambda\ell.\,\exists x.\,fx\ell}^{\textit{anybody}})^{<}(\overbrace{\text{saw}}^{\textit{saw}}{}^{>}\lfloor\overbrace{\text{shift}\,g.\,\neg\exists y.\,g\lceil y\rceil\ell}^{\textit{nobody}}\rfloor)\big)\big\rceil,$$

it then gets stuck after the following transitions.

$$\begin{aligned}
&\rightsquigarrow \text{reset}\big(\neg\exists y.\,\big(\lambda y.\,\text{reset!}\big\lceil I_t\,\text{reset}((\text{shift}\,f.\,\lambda\ell.\,\exists x.\,fx\ell)^{<}(\text{saw}^{>}\lfloor y\rfloor))\big\rceil\big)\lceil y\rceil\ell\big)\\
&\rightsquigarrow \text{reset}\big(\neg\exists y.\,\big(\text{reset!}\big\lceil I_t\,\text{reset}((\text{shift}\,f.\,\lambda\ell.\,\exists x.\,fx\ell)^{<}(\text{saw}^{>}\lfloor\lceil y\rceil\rfloor))\big\rceil\big)\ell\big)\\
&\rightsquigarrow \text{reset}\big(\neg\exists y.\,\big(\text{reset!}\big\lceil I_t\,\text{reset}((\text{shift}\,f.\,\lambda\ell.\,\exists x.\,fx\ell)^{<}(\text{saw}^{>}y))\big\rceil\big)\ell\big)\\
&\rightsquigarrow \text{reset}\big(\neg\exists y.\,\big(\text{reset}(I_t\,\text{reset}((\text{shift}\,f.\,\lambda\ell.\,\exists x.\,fx\ell)^{<}(\text{saw}^{>}y)))\big)\ell\big)\\
&\rightsquigarrow \text{reset}\big(\neg\exists y.\,\big(\text{reset}(I_t\,\text{reset}(\lambda\ell.\,\exists x.\,(\lambda x.\,\text{reset}(x^{<}(\text{saw}^{>}y)))x\ell))\big)\ell\big)\\
&\rightsquigarrow \text{reset}\big(\neg\exists y.\,\big(\text{reset}(I_t(\lambda\ell.\,\exists x.\,(\lambda x.\,\text{reset}(x^{<}(\text{saw}^{>}y)))x\ell))\big)\ell\big)
\end{aligned}$$

The scope of *nobody*, namely *anybody saw* ___, is a function rather than a proposition, so the intervening I_t blocks licensing. (The I_t would block licensing even if we quote the license ℓ—that is, even if we replace the last ℓ in (38) by $\lceil\ell\rceil$.) In general, a polarity item must be evaluated after its licensor, because a quantifier can take inverse scope only over a proposition.

Our operational semantics of metalinguistic quotation, like Barker and Shan's analyses of polarity sensitivity [8, 21], thus joins a syntactic notion of order to a semantic notion of scope in an account of polarity—as desired [13, 20]. The general strategy is to proliferate clause types: a clause in the analysis above may denote not a proposition but a function from ℓ to propositions. We can generalize this strategy to more clause types in order to account for additional quantifiers and polarity items, such as in English [15], Dutch, Greek, Italian [22], and Hungarian [23].

3.4 Other Linguistic Phenomena

Quantification and polarity sensitivity are just two out of many apparently non-compositional phenomena in natural language, which we term *linguistic side effects* [11]. Two other linguistic side effects are anaphora and interrogation. As the term suggests, each effect finds a natural treatment in operational semantics. For example, it is an old idea to treat anaphora as mediated by a record of referents introduced so far in the discourse [24, 25, 26]. We can express this idea in an operational semantics either by adding the record to the state as a separate component, as sketched in Section 1, or by integrating the record into the evolving program as it rewrites [27]. Another old idea is that wh-words take scope to circumscribe how an asker and an answerer may interact [28]. Our use of delimited control extends to this instance of scope taking.

The payoff of recasting these old ideas in a general operational framework goes beyond conceptual clarity and notational simplicity. Our notions of *context* and *order* apply uniformly to all linguistic phenomena and make borne-out predictions. For example, left-to-right evaluation explains not just the interaction between quantification and polarity sensitivity but also crossover in binding and superiority in questions [8, 9].

4 Conclusion

We have shown how an operational semantics for delimited control and multistage programming in natural language helps explain inverse scope, polarity sensitivity, and their interaction. We are actively investigating the foundations and applications of our metalanguage, where many open issues remain. In particular, there is currently no type system or denotational semantics on the market that soundly combines delimited control and quotation, so we have no way to understand type-logically or statically what it means for a program such as (38) to get stuck, much as we would like to.

Our approach extends dynamics semantics from the intersentential level to the intrasentential level, where side effects are incurred not only by sentences (*A man walks in the park*) but also by other phrases (*nobody*). Thus discourse context is not a sequence of utterances in linear order but a tree of constituents in evaluation order. This view unifies many aspects of human language—syntactic derivation, semantic evaluation, and pragmatic update—as compatible transitions among a single set of states. A tight link between operational and denotational semantics [5] promises to strengthen the connection between the views of language as product and language as action [1].

References

[1] Trueswell, J.C., Tanenhaus, M.K. (eds.): Approaches to Studying World-Situated Language Use: Bridging the Language-as-Product and Language-as-Action Traditions. MIT Press, Cambridge (2005)

[2] Plotkin, G.D.: A structural approach to operational semantics. Technical Report DAIMI FN-19, Department of Computer Science, University of Aarhus, Revised version submitted to Journal of Logic and Algebraic Programming (1981)

[3] Felleisen, M.: The Calculi of λ_v-CS Conversion: A Syntactic Theory of Control and State in Imperative Higher-Order Programming Languages. PhD thesis, Computer Science Department, Indiana University Also as Tech.Rep. 226 (1987)

[4] Felleisen, M.: The theory and practice of first-class prompts. In: POPL 1988. Conference Record of the Annual ACM Symposium on Principles of Programming Languages, pp. 180–190. ACM Press, New York (1988)

[5] Danvy, O., Filinski, A.: A functional abstraction of typed contexts. Technical Report 89/12, DIKU, University of Copenhagen, Denmark (1989), http://www.daimi.au.dk/~danvy/Papers/fatc.ps.gz.

[6] Danvy, O., Filinski, A.: Abstracting control. In: Proceedings of the 1990 ACM Conference on Lisp and Functional Programming, pp. 151–160. ACM Press, New York (1990)

[7] Danvy, O., Filinski, A.: Representing control: A study of the CPS transformation. Mathematical Structures in Computer Science 2, 361–391 (1992)

[8] Barker, C., Shan, C.c.: Types as graphs: Continuations in type logical grammar. Journal of Logic, Language and Information 15, 331–370 (2006)

[9] Shan, C.c., Barker, C.: Explaining crossover and superiority as left-to-right evaluation. Linguistics and Philosophy 29, 91–134 (2006)

[10] Shan, C.c.: Linguistic side effects. In: Barker, C., Jacobson, P. (eds.) Direct Compositionality, pp. 132–163. Oxford University Press, New York (2007)

[11] Shan, C.c.: Linguistic Side Effects. PhD thesis, Harvard University (2005)

[12] Fry, J.: Negative polarity licensing at the syntax-semantics interface. In: Cohen, P.R., Wahlster, W. (eds.) Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics, pp. 144–150. Morgan Kaufmann, San Francisco (1997)

[13] Fry, J.: Proof nets and negative polarity licensing. In: Dalrymple, M. (ed.) Semantics and Syntax in Lexical Functional Grammar: The Resource Logic Approach, pp. 91–116. MIT Press, Cambridge (1999)

[14] Kiselyov, O., Shan, C.c., Sabry, A.: Delimited dynamic binding. In: ICFP 2006. Proceedings of the ACM International Conference on Functional Programming, pp. 26–37. ACM Press, New York (2006)

[15] Shan, C.c.: Polarity sensitivity and evaluation order in type-logical grammar. In: Dumais, S., Marcu, D., Roukos, S. (eds.) Proceedings of the 2004 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Somerset, NJ. Association for Computational Linguistics, vol. 2, pp. 129–132 (2004)

[16] Kroch, A.S.: The Semantics of Scope in English. PhD thesis, Massachusetts Institute of Technology (1974) Reprinted by New York: Garland (1979)

[17] Szabolcsi, A.: Positive polarity—negative polarity. Natural Language and Linguistic Theory 22, 409–452 (2004)

[18] Taha, W., Nielsen, M.F.: Environment classifiers. In: POPL 2003. Conference Record of the Annual ACM Symposium on Principles of Programming Languages, pp. 26–37. ACM Press, New York (2003)
[19] Quine, W.V.O.: Word and Object. MIT Press, Cambridge (1960)
[20] Ladusaw, W.A.: Polarity Sensitivity as Inherent Scope Relations. PhD thesis, Department of Linguistics, University of Massachusetts (1979) Reprinted by New York: Garland (1980)
[21] Shan, C.c.: Delimited continuations in natural language: Quantification and polarity sensitivity. In: Thielecke, H., ed.: CW 2004: Proceedings of the 4th ACM SIGPLAN Continuations Workshop. Number CSR-04-1 in Tech.Rep., School of Computer Science, University of Birmingham, 55–64 (2004)
[22] Bernardi, R.: Reasoning with Polarity in Categorial Type Logic. PhD thesis, Utrecht Institute of Linguistics (OTS), Utrecht University (2002)
[23] Bernardi, R., Szabolcsi, A.: Partially ordered categories: Optionality, scope, and licensing (2006), http://ling.auf.net/lingBuzz/000372
[24] Kamp, H.: A theory of truth and semantic representation. In: Groenendijk, J.A.G., Janssen, T.M.V., Stokhof, M.B.J. (eds.) Formal Methods in the Study of Language: Proceedings of the 3rd Amsterdam Colloquium, Amsterdam, Mathematisch Centrum, pp. 277–322 (1981)
[25] Groenendijk, J., Stokhof, M.: Dynamic predicate logic. Linguistics and Philosophy 14, 39–100 (1991)
[26] Heim, I.: The Semantics of Definite and Indefinite Noun Phrases. PhD thesis, Department of Linguistics, University of Massachusetts (1982)
[27] Felleisen, M., Friedman, D.P.: A syntactic theory of sequential state. Theoretical Computer Science 69, 243–287 (1989)
[28] Karttunen, L.: Syntax and semantics of questions. Linguistics and Philosophy 1, 3–44 (1977)

A Multimodal Type Logical Grammar Analysis of Japanese: Word Order and Quantifier Scope*

Rui Otake and Kei Yoshimoto

Graduate School of International Cultural Studies, Tohoku University,
41 Kawauchi, Aoba-ku, Sendai 980-8576, Japan
otake@mail.tains.tohoku.ac.jp

Abstract. This paper presents an analysis of the interaction of scrambling and quantifier scope in Japanese, based on multimodal type logical grammar [5,6]. In developing the grammar of the language, we will make use of several modes. In particular, we will exploit the continuation mode as used in [2,7]. The concept deals with computational side effects and the evaluation order. After establishing the analysis of simple quantifier cases, we also discuss some morphologically related phenomena such as focus and split-QP construction [8,9], and show how they can be dealt with in our system.

1 Introduction

It is a well-known fact that a sentence like (1a) has two readings: the linear scope reading (1b) and the inverse scope reading (1c), although there is a preference for the former over the latter.

(1) a. Someone praised everyone.
   b. $\exists x[\text{human}(x) \wedge \forall y[\text{human}(y) \rightarrow \text{praise}(x, y)]]$
   c. $\forall y[\text{human}(y) \rightarrow \exists x[\text{human}(x) \wedge \text{praise}(x, y)]]$

In Japanese, however, such ambiguity does not arise from the word-to-word translation of sentence (1a). The only reading available for (2) is the linear scope reading (1b).

(2) Dareka-ga daremo-o hometa
   someone-NOM everyone-ACC praised
   'lit. Someone praised everyone.'

The reading (1c) can be obtained by preposing or scrambling the object noun phrase as in (3). It is also reported that the reading in (1b), *inverse* scope reading (with respect to this new surface word order) is also available (cf. [3]).

* The authors would like to thank Alastair Butler, Yusuke Kinoshita, Oleg Kiselyov, Ken Shan, Willemijn Vermaat, and two anonymous reviewers for valuable comments and discussions.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 135–148, 2008.

(3) Daremo-o dareka-ga hometa
everyone-ACC someone-NOM praised
'lit. Everyone, someone praised.'

These observations suggest that the two languages differ in the strategy to realize the non-canonical scope interpretation. This paper attempts to formalize this idea in terms of multimodal type logical grammar. In particular, we will exploit the notion of continuation as used in [2,7] for the analysis of English, and specify how Japanese grammar is different from English grammar.

The rest of the paper is organized as follows. Section 2 describes scrambling phenomena in Japanese and provides an account within the framework of multimodal type logical grammar. In section 3 we introduce continuation mode into our grammar drawing on the work by Barker and Shan [2], and Shan [7] on English. We present the analysis of the interaction of word order and quantifier scope in Japanese. In section 4, we extend the analysis of the previous sections to the related phenomena, namely focus particle and split-QP construction. Finally, section 5 concludes.

2 Word Order in Japanese

This section aims at the multimodal analysis of scrambling phenomena in Japanese. After the brief introduction of the framework, we will look at some simple data and analyze them by introducing several unary modes. First, case features are encoded by the modal decorations of the form $\boxminus\Diamond$. This captures the optionality of the explicit case marking via the derivability relation $A \vdash \boxminus\Diamond A$. And then scrambling phenomena will be analyzed in terms of unary modes θ and s, which introduce a restricted form of commutativity into our grammar.

2.1 Multimodal Type Logical Grammar

In type logical grammar, all the grammatical/semantic information is encoded in the lexicon.[1] And just by applying logical rules and structural rules to the lexical items, we can derive the proof of (the acceptability of) any given utterance, which can directly be used to specify how to compute the meaning of that utterance (in the given context), due to Curry-Howard correspondence.

The virtue of a multimodal system is that while the base logic **NL** has the rigid notion of constituent structure (both hierarchically and horizontally, because of non-associativity and non-commutativity, respectively), we can fine-tune the structural properties of the grammar by the appropriate set of the modes and the structural rules that regulate the interaction between different modes.

[1] In the framework adopted here, one would argue that some grammatical information is encoded by structural rules. Also, the presence or absence of the Unquote rule as argued in section 3 can be seen as a kind of grammatical information. However, the application of a structural rule is licensed by its associated mode, which is specified by the lexicon (either directly or indirectly, through another structural rule). So there is a sense in which the lexicon is responsible for all the grammatical information.

Figure 1 summarizes the logical rules. Subscripts i and j are metavariables for the indices of binary and unary modes, respectively. Semantically, \E and /E (slash elimination) rules correspond to functional application, and \I and /I (slash introduction) rules correspond to λ-abstraction. The unary connectives $\Diamond$ and $\boxminus$ are used to regulate the structural behavior of the expression marked by them, and they have no significance on the part of semantics.

$$\frac{}{A \vdash A}\,\mathrm{Id} \qquad \frac{\Gamma \vdash A/_iB \quad \Delta \vdash B}{(\Gamma,_i \Delta) \vdash A}\,/_i\mathrm{E} \qquad \frac{(\Gamma,_i B) \vdash A}{\Gamma \vdash A/_iB}\,/_i\mathrm{I}$$

$$\frac{\Delta \vdash B \quad \Gamma \vdash B\backslash_i A}{(\Delta,_i \Gamma) \vdash A}\,\backslash_i\mathrm{E} \qquad \frac{(B,_i \Gamma) \vdash A}{\Gamma \vdash B\backslash_i A}\,\backslash_i\mathrm{I}$$

$$\frac{\Gamma \vdash \boxminus_j A}{\langle\Gamma\rangle^j \vdash A}\,\boxminus_j\mathrm{E} \qquad \frac{\langle\Gamma\rangle^j \vdash A}{\Gamma \vdash \boxminus_j A}\,\boxminus_j\mathrm{I}$$

$$\frac{\Delta \vdash \Diamond_j B \quad \Gamma[\langle B\rangle^j] \vdash A}{\Gamma[\Delta] \vdash A}\,\Diamond_j\mathrm{E} \qquad \frac{\Gamma \vdash A}{\langle\Gamma\rangle^j \vdash \Diamond_j A}\,\Diamond_j\mathrm{I}$$

Fig. 1. Logical Rules

2.2 Case Marking and Scrambling in Japanese

In Japanese, word order is relatively free except that the main predicates (verbs or adjectives) of a clause are strictly fixed to the final position.[2] In particular, noun phrases can appear in any order as long as they are suffixed by a case particle, as illustrated in (4).[3] In such cases, permuting order does not lead to any ambiguity, since the thematic role of a noun phrase with respect to the predicate can be read off from the case marking.

(4) a. Taroo-ga Hanako-o hometa.
Taro-NOM Hanako-ACC praised

b. Hanako-o Taroo-ga hometa.
Hanako-ACC Taro-NOM praised
'Taro praised Hanako.'

[2] There is an exception to this generalization. See (i), an example of Right Dislocation:

(i) Taroo-ga hometa yo, Hanako-o.
Taro-NOM praised ASSERT Hanako-ACC
'Taro praised Hanako.'

This kind of irregularity seems to be confined to the main clause. They are not included in the fragment presented here.

[3] We assume that the meanings of these sentences are identical. Of course, scrambled word order may well have discourse related meaning. However, since we are focusing on the sentential meaning, the above assumption suffices for the current purposes.

Often in colloquial speech, nominal arguments occur without a case particle, in which case its thematic role is determined solely on the basis of the position relative to the predicate. The utterance (5) expresses the same meaning as (4), and cannot mean 'Hanako praised Taro'.[4]

(5) Taroo Hanako hometa (yo).
Taro Hanako praised ASSERT

In view of these observations, a transitive verb like hometa is given the lexical assignment shown in (6):

(6) hometa $\vdash \lambda x \lambda y.\text{praise}(y, x) : \boxminus_a \Diamond_a n \backslash (\boxminus_n \Diamond_n n \backslash s)$

Note that the syntactic type of the (transitive) verb is *curried*, and that the case of the arguments are encoded by the decorations $\boxminus_j \Diamond_j$ where j is either n (for nominative) or a (accusative). In isolation, the case of a bare noun phrase is unspecified. But within a sentence, it must be specified via successive applications of $\Diamond$I and $\boxminus$I rules. Put another way, the optionality of case marking is captured by the theorem $A \vdash \boxminus\Diamond A$. The non-case marked sentence in (5) is derived thus:[5]

$$\dfrac{\dfrac{\dfrac{\text{Taroo} \vdash n}{\langle\text{Taroo}\rangle^n \vdash \Diamond_n n}\Diamond_n \mathrm{I}}{\text{Taroo} \vdash \boxminus_n \Diamond_n n}\boxminus_n \mathrm{I} \quad \dfrac{\dfrac{\dfrac{\text{Hanako} \vdash n}{\langle\text{Hanako}\rangle^a \vdash \Diamond_a n}\Diamond_a \mathrm{I}}{\text{Hanako} \vdash \boxminus_a \Diamond_a n}\boxminus_a \mathrm{I} \quad \text{hometa} \vdash \boxminus_a \Diamond_a n \backslash (\boxminus_n \Diamond_n n \backslash s)}{\text{Hanako hometa} \vdash \boxminus_n \Diamond_n n \backslash s}\backslash \mathrm{E}}{\text{Taroo (Hanako hometa)} \vdash s}\backslash \mathrm{E}$$

To reduce clutter, we will henceforth abbreviate $\boxminus_n \Diamond_n n$ and $\boxminus_a \Diamond_a n$ to n_n and n_a, respectively.

2.3 Scrambling Mode

Given the discussion in 2.2, the function of the case particles is twofold. First, they case mark the attached noun, naturally. The other function is to give the *potential* for the displacement to the resulting noun phrase. To capture this latter function, we introduce two unary modes θ (for θ-position) and s (for scrambling) to our grammar. The structural postulates for these modes are shown in (7).

(7) $$\dfrac{\Gamma[\langle\Delta\rangle^\theta] \vdash C}{\Gamma[\langle\Delta\rangle^s] \vdash C}\text{Incl} \quad \dfrac{\Gamma[\Sigma(\langle\Delta\rangle^s\ \Pi)] \vdash C}{\Gamma[\langle\Delta\rangle^s(\Sigma\ \Pi)] \vdash C}\text{Scramble} \quad \dfrac{\Gamma[\langle\Delta\rangle^s] \vdash C}{\Gamma[\Delta] \vdash C}\mathrm{T}_s$$

Intuitively, θ-mode is licensed by a case particle and used to record the position that the noun phrase would have occupied if it had not been case marked. Incl rule turns θ-mode into s-mode, which can be seen as the potential for the

[4] In the example the sentence final particle yo is added since otherwise the sentence may sound awkward.

[5] The binary structural operator $\cdot,\cdot$ of the default mode will be omitted throughout for the readability. Outermost parentheses are also left out.

movement. Scramble rule says that s-marked noun phrases may move towards the front.[6]

In this system, case-marking indirectly licenses scrambling in that θ-mode originating from a case particle must be converted to s-mode before movement. See footnote 9 for the reason why this roundabout is necessary. T_s rule just discards the s-mode. T_s and Incl rules jointly capture the optionality of the scrambling. Case particles ga and o are accordingly given the lexical assignments in (8). In terms of semantics, they are just the identity function on nouns.

(8) a. $\mathsf{ga} \vdash \lambda x.x : n \backslash \boxminus_\theta n_n$
b. $\mathsf{o} \vdash \lambda x.x : n \backslash \boxminus_\theta n_a$

Given these setup, the scrambling word order is derived thus:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\mathsf{Taroo\ ga} \vdash \boxminus_\theta n_n}{\langle\mathsf{Taroo\ ga}\rangle^\theta \vdash n_n}\boxminus_\theta \mathrm{E}}{\mathsf{Taroo\ ga} \vdash n_n} \quad \dfrac{\dfrac{\mathsf{Hanako\ o} \vdash \boxminus_\theta n_a}{\langle\mathsf{Hanako\ o}\rangle^\theta \vdash n_a}\boxminus_\theta \mathrm{E} \quad \mathsf{hometa} \vdash n_a\backslash(n_n\backslash s)}{\langle\mathsf{Hanako\ o}\rangle^\theta\ \mathsf{hometa} \vdash n_n\backslash s}\backslash \mathrm{E}}{(\mathsf{Taroo\ ga})(\langle\mathsf{Hanako\ o}\rangle^\theta\ \mathsf{hometa}) \vdash s}\backslash \mathrm{E}}{(\mathsf{Taroo\ ga})(\langle\mathsf{Hanako\ o}\rangle^s\ \mathsf{hometa}) \vdash s}\mathrm{Incl}}{\langle\mathsf{Hanako\ o}\rangle^s((\mathsf{Taroo\ ga})\ \mathsf{hometa}) \vdash s}\mathrm{Scramble}}{(\mathsf{Hanako\ o})((\mathsf{Taroo\ ga})\ \mathsf{hometa}) \vdash s}\mathrm{T}_s}{}$$

The derivation for the canonical order is trivial. It is obtained if we simply drop the s-mode by T_s rule at the antepenultimate line, instead of applying Scramble.

To sum up, the Scramble rule gives rise to a limited form of commutativity. An argument can be moved toward the front insofar as it is marked by s-mode which is indirectly qualified by the case particle. Yet the scrambling of a non-case marked noun phrase is disallowed.

3 Continuation Analysis of the Quantifier Scope

In this section, we will first show that quantifier scoping is closely related to the word order, then present an analysis based on the continuation mode [2,7].

3.1 Quantifier Scope in Japanese

Consider the data in (9), repeated from (2) and (3):

---

[6] If utterances like (i), where only one of the arguments is case marked, should also be accepted, then lowering of a case marked noun phrase will be necessary and Scrambling rule must be restated so as to apply in both directions.

(i) Hanako Taroo-ga hometa
Hanako Taro-NOM praised
'Hanako, Taro praised.'

But the status of such sentences is subtle, so we will not pursue this possibility here.

(9) a. Dareka-ga daremo-o hometa
someone-NOM everyone-ACC praised
'lit. Someone praised everyone.'

b. Daremo-o dareka-ga hometa
everyone-ACC someone-NOM praised
'lit. Everyone, someone praised.'

As we saw in the introduction, the non-scrambled sentence (9a) is unambiguously interpreted as the subject taking wide scope, while the scrambled sentence (9b) can have both wide and narrow readings for the object (cf. [3]). In both sentences, at least on one reading, the leftmost quantifier takes the widest scope. Put simply, the two sentences have linear scope reading. Setting aside for a moment the presence of the inverse scope reading for (9b) (which we will discuss in 3.3), it could be argued that the above fact derives from the general tendency to the left-to-right evaluation order. The notion of evaluation order can be neatly modelled by incorporating the continuation mode in a multimodal type logical grammar as demonstrated in [2,7], the details of which we will turn to below.

3.2 Continuation Mode

Basically, the continuation for an element is the *context* in which it occurs. So for example, in the sentence (4a), repeated as (10a), the continuation for Hanako can be depicted informally as in (10b).

(10) a. Taroo-ga Hanako-o hometa.
Taro-NOM Hanako-ACC praised

b. ((Taroo ga)(([ ] o) hometa))

c. $\lambda x.\text{praise}(\text{taro}, x)$

The context (10b) can be seen as the function from the expression of type n (like Hanako) to the expression of type s. Accordingly, the semantic interpretation of (10b) can be given in (10c). Now consider (11), where a quantifier daremo appears in the same context as (10b).

(11) Taroo-ga daremo-o hometa.
Taro-NOM everyone-ACC praised
'Taro praised everyone'.

It is obvious that the meaning in (10c) cannot apply to the interpretation of daremo. Conversely, the latter takes its continuation as its argument as follows (ignoring the restriction for simplicity):

$$\underbrace{(\lambda k \forall x.k(x))}_{\text{daremo}}\ \underbrace{(\lambda x.\text{praise}(\text{taro}, x))}_{(10\text{b})} \rightsquigarrow \underbrace{\forall x.\text{praise}(\text{taro}, x)}_{(11)}$$

It is easy to see that, if we can manipulate the continuation like (10b) as a syntactic unit, then we will have a very simple account of the nonlocal behavior of quantifiers. For this purpose, we introduce a new binary mode, *continuation* mode, following [2,7]. For readability, the symbols ⊚, ⫽ and ⑊ (read *at*, *outside* and *inside*, respectively) are used instead of the structural punctuation $,_c$ and the type constructors $/_c$ and $\backslash_c$. Figure 2 indicates the structural rules that license the legitimate continuation. Figure 3 shows the same rules as well as two other necessary structural rules T and K′ in the Natural Deduction format.

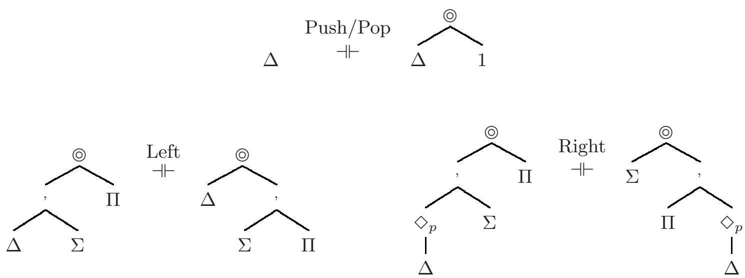


Fig. 2. Structural postulates for continuation mode, represented in tree form

$$\frac{\Gamma[\Delta] \vdash C}{\Gamma[\Delta \circledcirc 1] \vdash C}\text{Push} \qquad \frac{\Gamma[\Delta \circledcirc 1] \vdash C}{\Gamma[\Delta] \vdash C}\text{Pop} \qquad \frac{\Gamma[\langle \Delta\ \Sigma \rangle^p] \vdash C}{\Gamma[(\langle \Delta \rangle^p \langle \Sigma \rangle^p)] \vdash C}\text{K}'$$

$$\frac{\Gamma[\Delta \circledcirc (\Sigma\ \Pi)] \vdash C}{\Gamma[(\Delta\ \Sigma) \circledcirc \Pi] \vdash C}\text{Left} \qquad \frac{\Gamma[(\langle \Delta \rangle^p\ \Sigma) \circledcirc \Pi)] \vdash C}{\Gamma[\Sigma \circledcirc (\Pi\ \langle \Delta \rangle^p)] \vdash C}\text{Right} \qquad \frac{\Gamma[\langle \Delta \rangle^p] \vdash C}{\Gamma[\Delta] \vdash C}\text{T}$$

Fig. 3. Structural postulates for continuation mode, in Natural Deduction format

The connective ⊚ combines two things, by plugging what appears to its left hand side into what appears to its right hand side. The Push/Pop rules concern with the most trivial continuation, the null context. It is expressed by the atomic type 1, which is the right identity for the continuation mode.

Left and Right rules at first sight may give an impression that they are drastically destructing the structural configuration. But they have a very natural interpretation and in fact they do not affect the surface constituent structure. In order to see this, you can think of these trees as representing graphs where the vertical directionality has no significance but only the (counter-clockwise) rotating order count.[7] Then in each of these rules, when you start from the node Π, which stands for the context containing the 1 node (which in turn corresponds to the root of the constituent tree), tracing the same path while skipping the ⊚ node, you arrive at the same node in both sides.

The Left and Right rules can be seen as allowing the ⊚ node recursively going down (or going up) the original constituent tree (of which the root node is temporarily labelled by 1) to establish the legal configuration for a continuation.

[7] We adopt here Shan's version [7] rather than Barker and Shan's [2].

Note that in the Right rule, the ◎ node can go down to (or up from) the *right* (with respect to the original surface syntactic tree) branch, only if the *left* branch is enclosed by the unary structural operator $\langle\cdot\rangle^p$. This new mode p (for *pure* value) indicates that the constituent so marked has already been evaluated. Thus the Right rule enforces the left-to-right evaluation order for the continuation-sensitive elements such as quantifiers. This is the most attractive feature of the continuation mode, since it provides the tool to talk and reason about the evaluation order in the object language of the type logical grammar. Now we redefine the type of sentences as $\Diamond_p s$ rather than just s. Note that this change will not affect the derivability of the examples we considered so far, since in the presence of T rule we have $s \vdash \Diamond_p s$. The rules T and K′ regulate the p mode: T rule says that any expression can be turned into a pure value, and K′ rule says that two pure terms can be concatenated. The lexical entries for quantifiers can now be specified as in (12).

(12) a. dareka $\vdash \lambda k \exists x.\text{human}(x) \wedge k(x) : \Diamond_p s /\!\!/ (n \backslash\!\!\backslash \Diamond_p s)$
b. daremo $\vdash \lambda k \forall x.\text{human}(x) \rightarrow k(x) : \Diamond_p s /\!\!/ (n \backslash\!\!\backslash \Diamond_p s)$

The argument type of the quantifier $(n \backslash\!\!\backslash \Diamond_p s)$ is a type of the continuation that expects an element of type n within it to produce the expression of type $\Diamond_p s$. And the type of quantifier itself expresses the idea that if it is plugged into the continuation of type $(n \backslash\!\!\backslash \Diamond_p s)$, it produces the expression of type $\Diamond_p s$. Figure 4 illustrates the linear scope reading for (9a).[8]

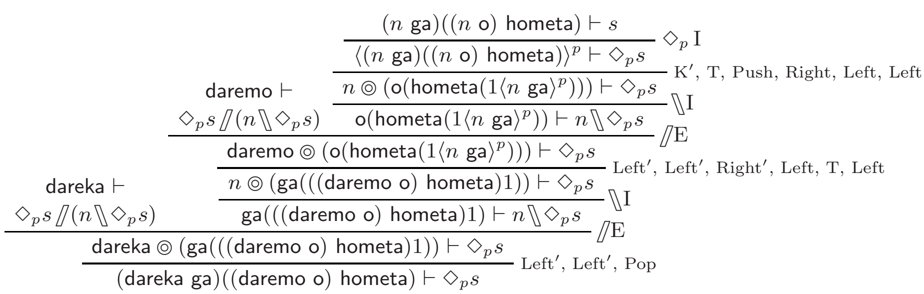


Fig. 4. Linear scope reading for SOV canonical order (9a)

Generally speaking, in the derivations like Figure 4, the element being applied at a lower step (that is, evaluated earlier) is the one that takes precedence over the other element at a higher step (that is, evaluated later). To get the linear scope reading for the scrambled word order (9b), we first derive the scrambled word order and then apply the same sequence of rules as in Figure 4, with the points at which dareka and daremo are introduced exchanged.

[8] Sequence of structural rules are compressed to save space. The labels Left′ and Right′ indicates the bottom-to-top application of the Left and Right rules respectively.

We can explain the impossibility of the inverse scope for (9a) as follows. In order for the object quantifier daremo to take wide scope, we have to plug the expression into the continuation of the form ((dareka ga)((([ ] o) hometa)). However, this configuration cannot be licensed, since we have (dareka ga) which is not a pure value to the left of the gap [ ].

Of course, it is possible to mark (dareka ga) as pure via T rule, but in the absence of the converse of T rule, $\langle\cdot\rangle^p$ punctuation cannot be discharged.

In fact, the presence or absence of the (partial) converse of T is the key to derive the difference of quantifier scoping between Japanese and English discussed in the introduction. Recall that in English, although left-to-right preference for scope interpretation is present, the inverse scope reading is also available. In order to derive this reading, [2,7] introduce Unquote rule, which implements the idea of multistage programming (for the discussion of this see the cited work). This is shown in (13). The modality u is introduced and s is redefined as $\Diamond_u s'$.

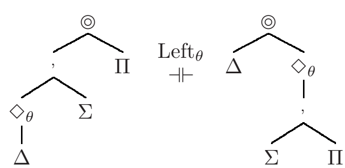
(13) $$\frac{\Gamma[\langle\Delta\rangle^u] \vdash C}{\Gamma[\langle\langle\Delta\rangle^u\rangle^p] \vdash C}\ \text{Unquote}$$

Unquote rule optionally allows that the evaluation of some element with computational side effect to be delayed, through the derivability relation $\Diamond_p s \vdash s$. Although inclusion of such a rule is certainly needed to account for the English data, we cannot follow the same strategy to account for the Japanese data, since Unquote rule would allow the inverse scope reading as a global option. In contrast, the current proposal predicts the fact that Japanese does not allow inverse scope in non-scrambled sentences like (9a), by *not* including the Unquote rule.

3.3 Inverse Scope Reading for Scrambled Sentence

The inverse reading of the scrambled sentence (9b) can be generated if the $\circledcirc$ node can *skip* the θ-modality. This is achieved by the Left$_\theta$ rule shown in (14).[9] Figure 5 illustrates the derivation. It amounts to computing the scope relation with respect to the word order *before* the scrambling of the object quantifier noun phrase takes place.

(14) $$\frac{\Gamma[\langle\Delta\rangle^\theta\ \Sigma] \circledcirc \Pi \vdash C}{\Gamma[\Delta \circledcirc \langle\Sigma\ \Pi\rangle^\theta] \vdash C}\ \text{Left}_\theta$$



[9] If we had allowed the communication between $\circledcirc$-mode and s-mode, the inverse scope reading for the SOV order would be derivable, since it is possible to have

$$(\langle \text{Obj}\rangle^p(\langle \text{Subj}\rangle^s\ \text{Verb})) \vdash \Diamond_p s$$

by first fronting the object, and we would be able to compute the *linear* reading with respect to this structure, and then move the subject at the front. So the distinction between θ-mode and s-mode is necessary.

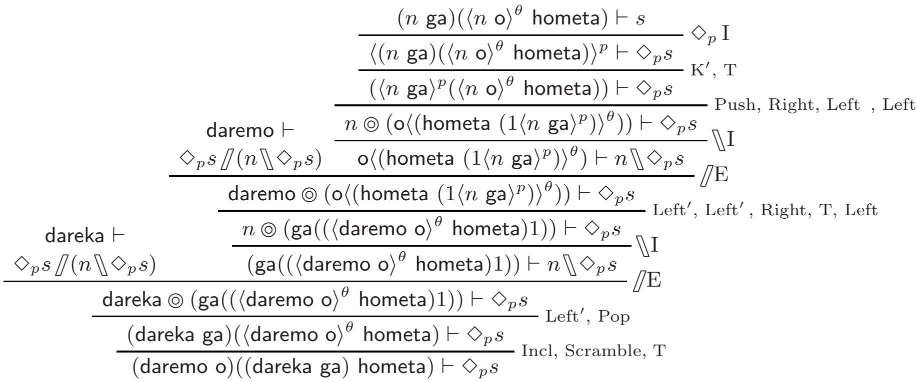


Fig. 5. Inverse scope reading for OSV scrambling word order (9b)

In this section, we have explained the different scope taking possibilities between English and Japanese in terms of the different set of structural rules (and the modes associated with them). With regard to the availability of the Unquote rule in English and its absence in Japanese, it might be speculated that given the strict word order, in deriving inverting scope reading the grammar of English has to take recourse to the Unquote rule, which does not change the word order, but changes the evaluation order. By contrast, the grammar of Japanese, already having a strategy to invert scope, namely Scramble rule, the need for delaying the evaluation simply does not arise. To put it another way, the different degree of flexibility in word order is reflected by the different strategies to the availability of scope reading *reversed* with respect to the thematic roles, assuming that Agent taking scope over Patient is the default case.

4 Some Related Phenomena

In this section, we will look at some related phenomena, namely focus and split-QP construction, that are also properly understood in terms of continuation.

4.1 Focus

Up to now, we have treated the quantifiers like daremo 'everyone' and dareka 'someone' as if they are unanalyzable units. Morphologically speaking, however, they can be decomposed into two parts: daremo can be decomposed into dare 'person' and mo 'every'. Likewise, dareka can be decomposed into dare 'person' and ka 'some'. Here, mo and ka are the elements that are responsible for the quantificational force of these items. On the other hand, dare belongs to the class called *indeterminates* [4], which also have a function similar to *wh*-words in English.

Now, let us investigate the morpheme mo in more detail. When the complement of mo does not contain indeterminates, its meaning will be equivalent to *also*. Consider (15):

(15) a. Taroo-mo Hanako-o hometa.
Taro-also Hanako-ACC praised
'*Taro* also praised Hanako.'

b. Taroo-ga Hanako-mo hometa.
Taro-NOM Hanako-also praised
'Taro also praised *Hanako*.'

c. Taroo-ga Hanako-o home-mo sita.
Taro-NOM Hanako-ACC praise-also did
'Taro also *praised* Hanako.'

As the translation suggests, the complement of mo is construed as a focused element. Continuation semantics is clearly related to the focus denotation [1]. We can specify the lexical assignment of mo as in (16):

(16) $\mathsf{mo}_{\text{also}} \vdash \lambda x \lambda k.k(x) \wedge \exists y.y \neq x \wedge k(y) : \Diamond_i X \backslash (\Diamond_p s /\!\!/ (X \backslash\!\!\backslash \Diamond_p s))$

We show the derivation of (15b), but now in the informal notation:

$$\dfrac{\dfrac{\langle \mathsf{Hanako} \rangle^i \vdash \Diamond_i n \quad \mathsf{mo} \vdash \Diamond_i n \backslash (\Diamond_p s /\!\!/ (n \backslash\!\!\backslash \Diamond_p s))}{\langle \mathsf{Hanako} \rangle^i\ \mathsf{mo} \vdash \Diamond_p s /\!\!/ (n \backslash\!\!\backslash \Diamond_p s)} \backslash \mathrm{E} \quad (\mathsf{Taroo\ ga})([\ ]\ \mathsf{hometa}) \vdash \Diamond_i n \backslash\!\!\backslash \Diamond_p s}{(\mathsf{Taroo\ ga})((\langle \mathsf{Hanako} \rangle^i\ \mathsf{mo})\ \mathsf{hometa}) \vdash \Diamond_p s} /\!\!/\mathrm{E}$$

The derivation above gives us the interpretation (see also (10)):

$$((\lambda x \lambda k.k(x) \wedge \exists y[y \neq x \wedge k(y)])(\text{hanako}))(\lambda x.\text{praise}(\text{taro}, x)) \rightsquigarrow$$
$$(\lambda k.k(\text{hanako}) \wedge \exists y[y \neq \text{hanako} \wedge k(y)])(\lambda x.\text{praise}(\text{taro}, x)) \rightsquigarrow$$
$$\text{praise}(\text{taro}, \text{hanako}) \wedge \exists y[y \neq \text{hanako} \wedge \text{praise}(\text{taro}, y)]$$

In words, this sentence asserts that Taro praised Hanako, and presupposes that there is another person whom Taro praised. For the presuppositional part of the meaning, we blatantly used logical conjunction. But in this way, we also notice the similarity to the meaning of mo in the universal quantifier reading. Recall that (11) Taroo-ga daremo-o hometa has the interpretation:

$$\forall x.\text{praise}(\text{taro}, x)$$

which can be thought of as the conjunction of the form:

$$\text{praise}(\text{taro}, \text{hanako}) \wedge \text{praise}(\text{taro}, \text{jiro}) \wedge \text{praise}(\text{taro}, \text{saburo}) \wedge ...$$

where x in the body of the universal quantifier substituted for constants in the domain of individuals.

4.2 Split-QP and *wh*-Island

There exists a construction where the indeterminate and the quantifier appear separated from each other, but semantically interact. Consider (17).

(17) [Dare-ga kaita hon] mo omosiroi
person.INDET-NOM wrote book every is:interesting
'lit. Every book that a person wrote is interesting.'

Following [9], we will refer to such construction as *Split-QP* (see also [3,4,8], among others). For the lack of space, we will only sketch the analysis in terms of continuation, leaving out the fuller treatment for future study.

We regard mo-phrase in split-QP as forming the complex quantifier phrase. Semantically speaking, the entire mo-phrase is a function that takes its continuation as its argument (or scope):

$$\underbrace{(\lambda k \forall x.k(\iota y.\text{book}(y) \wedge \text{wrote}(x,y)))}_{\textsf{dare ga kaita hon mo}} \underbrace{(\lambda x.\text{interesting}(x))}_{(\textsf{[ ]omosiroi})} \rightsquigarrow$$
$$\underbrace{\forall x.\text{interesting}(\iota y.\text{book}(y) \wedge \text{wrote}(x,y))}_{(17)}$$

By reasoning backwards, dare ga kaita hon mo should have the same syntactic type as daremo, i.e., $\Diamond_p s /\!\!/ (n \backslash\!\!\backslash \Diamond_p s)$. And the lexical assignments of dare and mo can be specified as in (18).

(18) a. $\textsf{mo}_\forall \vdash \lambda p \lambda k \forall x.k(px) : \Diamond_i(n\backslash_? n)\backslash(\Diamond_p s /\!\!/ (n \backslash\!\!\backslash \Diamond_p s))$
b. $\textsf{dare} \vdash \lambda k \lambda x.k(x) : (n\backslash_? X) /\!\!/ (n \backslash\!\!\backslash X)$

The following proof shows that dare ga kaita hon mo has the same type as the quantifier daremo:[10]

$$\cfrac{\cfrac{\cfrac{\textsf{dare} \vdash (n\backslash_? n) /\!\!/ (n \backslash\!\!\backslash n) \quad (((\textsf{[ ] ga}) \textsf{ kaita}) \textsf{ hon} \vdash n \backslash\!\!\backslash n}{((\textsf{dare ga}) \textsf{ kaita}) \textsf{ hon} \vdash n\backslash_? n} \backslash\!\!\backslash E}{\langle((\textsf{dare ga}) \textsf{ kaita}) \textsf{ hon}\rangle^i \vdash \Diamond_i(n\backslash_? n)} \Diamond_i I \quad \begin{array}{c}\textsf{mo} \vdash \\ \Diamond_i(n\backslash_? n)\backslash(\Diamond_p s /\!\!/ (n \backslash\!\!\backslash \Diamond_p s))\end{array}}{\langle((\textsf{dare ga}) \textsf{ kaita}) \textsf{ hon}\rangle^i \textsf{ mo} \vdash \Diamond_p s /\!\!/ (n \backslash\!\!\backslash \Diamond_p s)} \backslash\!\!\backslash E$$

Note that by the lexical assignments in (16) and (18a), we require that the complement of mo be marked by the unary mode i. This is needed for the account of *wh*-island effect: indeterminates like dare cannot be associated with mo or ka, across intervening mo or ka. This is illustrated in (19):

(19) a. [[Dare-ga kaita hon] mo omosiroi to omou gakusei]
person.INDET-NOM wrote book MO interesting COMP think student
mo kita.
MO came

[10] We have made implicit the derivation of the relative clause. We simply assume the head noun is assigned the type $(\Diamond_s \boxminus_s n\backslash\Diamond_p s)\backslash n$ in the lexicon, where $\Diamond_s \boxminus_s n$ corresponds to the *gap* in the relative clause (an alternative way to get the same thing would be to posit a phonologically empty relative pronoun).

b. The student who thinks that for every x, x a person, the book x wrote is interesting came, too.

c. *For every x, x a person, a student who thinks that the book x wrote is also interesting, came.

In (19a), there are two occurrences of mo. Given the presence of the indeterminate dare, one of them has to be construed as the universal quantifier (18a) and the other as the focus particle (16). So in principle there is two possible interpretations as shown in (19b–c).

But in the presence of i mode, and the *lack* of the structural rule that communicates i mode and continuation mode (or the variant of T rule that discards i mode), indeterminate dare can only take the context:

$$[[\ ]\mathsf{ga\ kaita\ hon}]$$

as its argument, but not the larger context:

$$[[[\ ]\mathsf{ga\ kaita\ hon}]\mathsf{mo\ omosiroi\ to\ omou\ gakusei}]$$

so that *wh*-island effect ensues.[11]

Finally, we show how the lexical assignments in (18a–b) relate to the non-split form daremo. Actually, the lexical entry for daremo can be derived from (18a–b), if we posit the phonologically empty element ϵ:

$$\epsilon \vdash \lambda x.x : n \mathbin{\backslash\!\!\backslash} n$$

We can derive the string daremo with type $\Diamond_p s \mathbin{/\!\!/} (n \mathbin{\backslash\!\!\backslash} \Diamond_p s)$:

$$\dfrac{\dfrac{\dfrac{\mathsf{dare} \vdash (n\backslash_? n) \mathbin{/\!\!/} (n \mathbin{\backslash\!\!\backslash} n) \quad \epsilon \vdash n \mathbin{\backslash\!\!\backslash} n}{\mathsf{dare}\ \epsilon \vdash n\backslash_? n} \mathbin{/\!\!/}\mathrm{E}}{\langle \mathsf{dare}\ \epsilon\rangle^i \vdash \Diamond_i (n\backslash_? n)} \Diamond_i \mathrm{I} \qquad \mathsf{mo} \vdash \Diamond_i (n\backslash_? n)\backslash(\Diamond_p s \mathbin{/\!\!/} (n \mathbin{\backslash\!\!\backslash} \Diamond_p s))}{\langle \mathsf{dare}\ \epsilon\rangle^i\ \mathsf{mo} \vdash \Diamond_p s \mathbin{/\!\!/} (n \mathbin{\backslash\!\!\backslash} \Diamond_p s)} \backslash \mathrm{E}$$

However, the derivation above is presented just for explaining the relatedness between the split and non-split quantifiers, and we may have to retain the separate lexical entry for daremo. This is because of the fact that daremo seems to require a case particle, while mo in split-QP does not, the difference remains unexplained in this paper.[12,13]

[11] The unary bracket $\langle\cdot\rangle^i$ can be seen as simulating the *delimited continuation*.

[12] If daremo is used without case particle, it is most likely to be construed as negative polarity item 'nobody', the case not treated in this paper.

[13] This requirement can be encoded in the grammar by introducing another unary mode k and setting the lexical entries for daremo and case particle as in (i).

(i) a. $\mathsf{daremo} \vdash \Diamond_p s \mathbin{/\!\!/} (\boxminus_k \Diamond_k n \mathbin{\backslash\!\!\backslash} \Diamond_p s)$

b. $\mathsf{ga} \vdash \boxminus_k \Diamond_k n \backslash \boxminus_\theta\, n_a$

5 Conclusion

In the multimodal type logical grammar formalism, an analysis of scrambling word order and the limitation on quantifier scope readings in Japanese was presented. We also discussed the difference of scope taking possibilities between English and Japanese as reflecting the different strategy to give a noncanonical scope reading, namely a global optional delaying device (Unquote) and the word order flexibility (Scramble). We also made some comments on the related phenomena in Japanese, namely focus and split-QP construction, however the discussion there was too far from being conclusive, and thus awaits further study.

References

1. Barker, C.: Continuations in Natural Language. In: Thielecke, H. (ed.) CW'04: Proceedings of the 4th ACM SIGPLAN continuations workshop, Tech. Rep. CSR-04-1, School of Computer Science, University of Birmingham, pp. 1–11 (2004)
2. Barker, C., Shan, C.-c.: Types as Graphs: Continuations in Type Logical Grammar. Journal of Logic, Language and Information 15, 331–370 (2006)
3. Hoji, H.: Logical Form Constraints and Configurational Structures in Japanese. Ph.D. dissertation, University of Washington (1985)
4. Kuroda, S.-Y.: Generative Grammatical Studies in the Japanese Language. Ph.D. dissertation, MIT (1965)
5. Moortgat, M.: Categorial Type Logics. In: van Benthem, J.F.A.K., ter Meulen, A.G.B. (eds.) Handbook of Logic and Language, ch. 2, Elsevier Science, Amsterdam (1997)
6. Morrill, G.V.: Type Logical Grammar: Categorial Logic of Signs. Kluwer, Dordrecht (1994)
7. Shan, C.-c.: Linguistic Side Effects. Ph.D. dissertation, Harvard University (2005)
8. Shimoyama, J.: Indeterminate Phrase Quantification in Japanese. Natural Language Semantics 14, 139–173 (2006)
9. Takahashi, D.: Determiner Raising and Scope Shift. Linguistic Inquiry 33, 575–615 (2002)

Coordinating and Subordinating Dependencies*

Alastair Butler

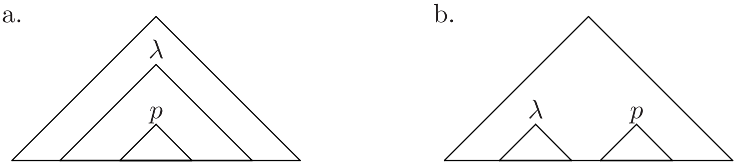
Department of English Language and Literature,
National University of Singapore,
Blk AS5, 7 Arts Link, Singapore, 117570
ellbaj@nus.edu.sg

Abstract. This paper focuses on the differences / similarities between coordinating and (distant) subordinating binding dependencies. We consider data that is suggestive that in natural language these dependencies are implemented with the same underlying mechanism. We look at four (dynamic) systems that capture these dependencies, first with distinct mechanisms, then with single mechanisms.

1 Introduction

This paper focuses on the differences / similarities between the dependency types pictured in (1), where λ is some form of binding operator and p is a bindee that is bound by λ.

(1) a. b.



At first sight, the two dependency types give the impression of being very different. In (1a), the bound element is within the syntactic scope of the binder. This might be referred to as a case of *classical binding*, since it is the sort of relation readily captured with classical static binding systems like predicate logic. This paper will avoid claims about the form of dependencies "dynamic" and "classical static" systems are limited to capture. For this reason, more neutral terminology is required, and so (1a) will be referred to as a *subordinating binding* relation.

In (1b), the binding relation takes place outside and to the right of the syntactic scope of the binder. It is tempting to refer to this as a case of *dynamic binding*, since it is readily captured with systems of dynamic semantics. But there are ways to capture (1b) with classical static semantics (e.g., Cresswell

* I would like to thank the participants of LENLS 2007 and the reviewer for the helpful and challenging remarks I received.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 149–159, 2008.

2002 uses predicate logic). For this reason (1b) will be referred to more neutrally as a *coordinating binding* relation.[1]

One clear difference between (1a) and (1b) might be termed a syntax / semantic difference. While (1a) can be characterised with syntax alone, the binding in (1b) can only arise with some level of semantics (/ pragmatics). Relevant levels might be either a representational level, as with Discourse Representation Theory (DRT), or a "level" of runtime evaluation, as with Dynamic Predicate Logic. It follows that, while syntactic requirements can be placed on relations of the form (1a) (e.g., Chomsky's 1981 'binding' conditions), (1b) is a dependency that defies what can be stated with syntax, and so constraints on (1b) can only be semantic (/ pragmatic). This doesn't mean to say that (1b) is unconstrained. An example of a relevant constraint that applies to (1b) is the requirement of *accessibility* introduced with DRT (see e.g., Kamp and Reyle 1993).

The rest of this paper is organised as follows. Pursuing the theme of differences, section 2 presents Dekker's PLA system. This implements the two dependency types with distinct binding mechanisms. Section 3 looks at some natural language data. The data suggests that, in natural language, both dependency types are captured with the same mechanism, at least when subordinating relations involve deeply embedded relations, as they must when pronouns are involved. Section 4 looks at Vermeulen's DPLE system. This provides a single dynamic way to capture the two dependency types, with primitive coordinating relations and derived subordinating effects. Section 5 looks at a new system designed for this paper: Index Semantics (IS). Built from the parts of PLA that are not predicate logic, IS gives an alternative (static) way of deriving subordinating and coordinating effects with a single mechanism. Section 6 offers yet another system, a reduced form of the Scope Control Theory of Butler (2007), which brings together the insights gained from the other systems.

2 Distinct Mechanisms for Subordinating and Coordinating Dependencies

In this section we look at Dekker's (2002) *Predicate Logic with Anaphora* (PLA). A central design goal of PLA is to provide an extended predicate logic that allows for coordinating binding dependencies but with minimal deviation from a predicate logic core that captures subordinating dependencies. The language is that of predicate logic with atomic formulas taking, in addition to variables, "pronouns" $\{p_1, p_2, ...\}$ as terms.

Definition 1. (PLA semantics).

Suppose a first-order model M with domain of individuals D. Suppose σ is a finite sequence of individuals from D assigned to the positions of σ. Suppose g

[1] Another appropriate name would be *donkey binding* relation, owing to Geach's original examples that involved donkeys.

is a (total) assignment from variables to individuals of D. The PLA semantics can be given as follows:

$\sigma \models_g \exists x\phi$	iff	$\sigma - 1 \models_{g[x/\sigma_1]} \phi$
$\sigma \models_g \phi \wedge \psi$	iff	$\sigma - n(\psi) \models_g \phi$ and $\sigma \models_g \psi$
$\sigma \models_g \neg\phi$	iff	$\sigma'\sigma \not\models_g \phi$ for all $\sigma' \in D^{n(\phi)}$
$\sigma \models_g P(t_1, ..., t_n)$	iff	$([\![t_1]\!]_{g,\sigma}, ..., [\![t_n]\!]_{g,\sigma}) \in M(P)$
$[\![x]\!]_{g,\sigma} = g(x)$,		$[\![p_i]\!]_{g,\sigma} = \sigma_i$

where:
σ_i returns the i-th individual of σ,
$\sigma - m$ is the sequence $e_{m+1}, e_{m+2}, ...$, and
$n(\phi)$ is a count of existentials in ϕ outside the scope of negation: $n(\exists x\phi) = n(\phi) + 1$, $n(\phi \wedge \psi) = n(\phi) + n(\psi)$, $n(\neg\phi) = 0$, $n(P(t_1, ..., t_n)) = 0$.

As an example of PLA in action, consider (2), which can be rendered as (3).

(2) Someone[1] enters. He$_1$ smiles. Someone[2] (else) enters. She$_2$ smiles at him$_1$.

(3) $\exists x \texttt{enters}(x) \wedge \texttt{smiles}(p_0) \wedge \exists x \texttt{enters}(x) \wedge \texttt{smiles at}(p_0, p_1)$

The binding value for an existential quantification comes from the sequence σ. Pronouns likewise take their values from σ. Since σ is an environment parameter, this gives (3) the same interpretation as the predicate logic formula (4).

(4) $\texttt{enters}(x_2) \wedge \texttt{smiles}(x_2) \wedge \texttt{enters}(x_1) \wedge \texttt{smiles at}(x_{1,} x_2)$

We see from (4) that, while remaining a syntactic part of the representing formula, PLA existential quantifiers are deprived of their quantificational effect. We also see how a pronoun's antecedent is not necessarily encoded by use of the same index (compare e.g., the use of indices in (3) with (2)). Instead, a pronoun's antecedent has to be worked out by taking into account the 'existential depth' of the intervening formula. This gives an appropriate way to characterise pronouns. Pronouns don't in principle have fixed antecedents, rather they favour closer (more salient) antecedents.[2] In contrast, variables result in fixed bindings. This seems appropriate for (3), with variables used to capture the clause internal argument links of (2), where there is no option for ambiguity.

3 Some Natural Language

While the division of labour PLA suggests is conceptually appealing, and despite the syntax / semantic differences observed in section 1, when we look at the way natural languages implement coordinating and subordinating dependencies, we find that the two dependency types can be remarkably similar, at least in terms of the lexical items employed and the linking options / constraints followed.

[2] This holds under the assumption that a lower integer is in some sense easier to allocate than a higher integer.

For example, the coordinating dependencies of (5) show the same distinction between reflexive binding and pronominal binding as do the subordinating dependencies of (6) and (7).

(5) a. Someone^1 is happy. You like him$_1$.
b. Someone^1 is happy. *You like himself$_1$.

(6) a. Someone^1 thinks you like him$_1$.
b. *Someone^1 thinks you like himself$_1$.

(7) a. Someone^1 saw someone who likes him$_1$.
b. *Someone^1 saw someone who likes himself$_1$.

Rendering (5a), (6a) and (7a) with PLA gives (8), (9) and (10), respectively.[3]

(8) $\exists x \texttt{happy(}x\texttt{)} \wedge \texttt{you like(}p_0\texttt{)}$

(9) $\exists x ... \textit{thinks} ... \texttt{you like(}x\texttt{)}$

(10) $\exists x \exists y (\texttt{likes(}y, x\texttt{)} \wedge \texttt{saw(}x, y\texttt{)})$

This reveals an inadequacy of the PLA renderings, as the natural language pronouns are not captured in a consistent manner: when in coordinating relations, natural language pronouns are captured as PLA pronouns; when in subordinating relations, natural language pronouns can only be captured as PLA variables.

Another example is that effects of covaluation arise with both subordinating and coordinating relations. Effects of covaluation are most readily seen to arise in discourse with coordinating relations. For example, in (11) we see a straightforward case of a mandatory reflexive. Typically, *him* should not be able to occur in the same environment as *himself* can occur and take the same referent; yet it does in (12).

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

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

It seems that (12) is possible because *him*'s antecedent is not the subject *John himself*, but rather the previous *John* in A's question. That is, the antecedent for *him* in (12) is an occurrence of *John* outside *him*'s local clausal context. This gives an instance of covaluation: that *him* and the subject *John himself* are coreferential is not the consequence of a local clause internal binding link, and so it doesn't fall prey to whatever the typical restrictions on local clause internal binding are.

[3] We will not aim to offer a treatment of embedding relations like *thinks*. To do this adequately would take us outside the PLA language of definition 1. Instead, we use '*...relation name...*' to indicate where the contribution of such an embedding relation should be placed.

What is interesting for our current concern is the observation made in Heim (1993) that covaluation effects can arise under the scope of a quantifier and so with a subordinating relation. What is still required for the covaluation effect to arise is sufficient intervening material to make the dependency cross clause boundaries (that is, the dependency cannot be local to a clause). For example, consider (13a) and (13b).

(13) a. Every man is afraid that only he voted for himself.
b. Every man is afraid that only he voted for him.

(13a) has a binding reading, where every man has the fear 'No one else voted for themselves!'. In contrast, (13b) allows for a covaluation reading where every man has the fear 'No one else voted for me!'. These meanings can be given roughly as (14a) and (14b).

(14) a. Every man is afraid that x voted for x is true only for $x =$ him.
b. Every man is afraid that x voted for him is true only for $x =$ him.

That the same phenomena of covaluation is observable with both subordinating and coordinating relations is very suggestive that natural language employs a single unified mechanism of pronominal binding that is responsible for both dependency types. Whatever explains the coordinating relations in (5) and (12) should bear on what explains the subordinating relations in (6), (7) and (13), and vice versa.

4 Only Dynamic Coordinating Relations

In this section we look at Dynamic Predicate Logic with Exit operators (DPLE) as introduced by Vermeulen (1993, 2000). DPLE gives a supremely elegant example of a dynamic semantics with an in-built storage facility. The most distinctive feature of the system is that it treats scoping of variables by means of stacks. Each variable is assigned its own stack. Roughly, the top of a stack corresponds to the current occurrence of a variable having scoping effect. In addition, the system has `push` and `pop` operations on stacks as basic actions that can be used at any convenient time, allowing for procedures on stacks to be carried out in a very liberal way.

We first introduce some of the fundamental notions on stacks required for a stack based semantics to work.

Definition 2. (stacks).

Let a domain of evaluation D and a set of variables V be given. We introduce the following notation for stacks,

– A (possibly empty) sequence $s \in D^\star$ is called a stack. (Here, $D^\star$ is the set of finite sequences of D.) We use ϵ to denote the empty stack.
– If $s = d_1 ... d_n$ and $s' = d'_1 ... d'_m$, then we write $s \star s'$ for $d_1 ... d_n d'_1 ... d'_m$.

- If $s = d_1...d_n$ $(n \geq 1)$, then we write $\uparrow(s)$ for d_n and $\downarrow(s)$ for d_1. In case $s = \epsilon$, $\uparrow(s)$ and $\downarrow(s)$ are undefined.
- If $s = d_1...d_n$, then $|s| = n$, the length of the sequence s.

stack valued assignments,

- A stack valued assignment, g, is a mapping that assigns a stack of values to each variable: $g : V \longrightarrow D^{\star}$. We use λ as notation for the 'empty assignment': $\lambda(x) = \epsilon$ $(x \in V)$.
- $SASS_D$ is the set of stack valued assignments.

and a relation on stack valued assignments.

- For each $x \in V$ we define $[x] \subseteq SASS_D \times SASS_D$ as follows: $(g, h) \in [x]$ iff $g(x) \star \uparrow(h(x)) = h(x)$ and $g(y) = h(y)$ whenever $y \neq x$.

We are now in a position to give the semantics. Formulas are interpreted in terms of relations on stack valued assignments that change the before-state of a stack valued assignment to an after-state in accordance with definition 3.

Definition 3. (DPLE).

Let a first order model $M = (D, I)$ be given. To each formula ϕ we assign a relation $[\![\phi]\!]_M \subseteq SASS_D \times SASS_D$ as follows:

$$
\begin{array}{lll}
(g,h) \in [\![\mathtt{push}x]\!]_M & \text{iff} & (g,h) \in [x] \\
(g,h) \in [\![\mathtt{pop}x]\!]_M & \text{iff} & (h,g) \in [x] \\
(g,h) \in [\![\phi \wedge \psi]\!]_M & \text{iff} & \exists j : (g,j) \in [\![\phi]\!]_M \text{ and } (j,h) \in [\![\psi]\!]_M \\
(g,h) \in [\![\mathtt{P(}x_1,...,x_n\mathtt{)}]\!]_M & \text{iff} & g = h \text{ and } (\uparrow(g(x_1)),...,\uparrow(g(x_n))) \in I(\mathtt{P})
\end{array}
$$

From definition 3, we see how the after-state of a formula is the before-state of the next conjoined formula. Predicate formulas correspond to tests on the values that are found on the tops of variable stacks. There are two ways of looking at the relation on stack valued assignments $(g, h) \in [x]$: read in one direction (from g to h) it describes the pushing of a random value on the x-stack of the input assignment, which gives the `push` action; read in the other direction (from h to g) it describes the popping of the latest value of the x-stack, which gives the `pop` action.

We can illustrate the uniform approach to subordinating and coordinating relations that DPLE facilitates by considering (15), which can be rendered as in (16).[4]

(15) Someone[1] agrees of someone[2] that he$_2$ likes him$_1$. He$_2$ is happy.

(16) $\mathtt{push}x \wedge \textit{push-...agrees...} \wedge \mathtt{push}y \wedge \mathtt{likes(}y, x\mathtt{)} \wedge \textit{pop-...agrees...} \wedge \mathtt{happy(}y\mathtt{)}$

[4] We include *push-...agrees...* and *pop-...agrees...* merely as a suggestion for how to approach the contribution of the embedding *agrees*.

We can also capture the possibility that accessibility to an antecedent is cut off, as (18) illustrates, which is intended to approximate (17).[5]

(17) Someone[1] doesn't agree of anyone[2] that he_2 likes him_1. *He_2 is upset.

(18) *$\texttt{push}x \wedge \textit{push-neg} \wedge \textit{push-...agrees...} \wedge \texttt{push}y \wedge \texttt{likes}(y,x) \wedge \texttt{pop}y \wedge$
$\textit{pop-...agrees...} \wedge \textit{pop-neg} \wedge \texttt{happy}(y)$

5 Index Semantics

DPLE is a truly dynamic system. Expressions are not only interpreted with respect to contexts (stack value assignments), they have themselves the power to change their context (push or pop values). Changes pass on as outputs and continue to persist unless explicitly changed with subsequent push and pop actions.

With PLA the perspective is different. Context is given as a structured sequence and never changed.[6] Dynamism emerges as an advancing through the given context, or to put this another way, the unveiling of previously inaccessible portions of context. The position reached comes from a division created by conjunction between values in active use and values not in active use. Values in active use are either values that will be used as the values for quantifications in the current conjunct, or as values accessible to pronouns, in which case they will have been used as the values for quantifications in prior conjuncts. Values not in active use are held back to serve as values for the quantifications of subsequent conjuncts.

We might imagine other ways of keeping track on where we are in a structured context. In this section we look at what we will call Index Semantics. The idea is to employ a "dynamic index" to explicitly keep track of the point that has been reached in a linearly structured context. With this change, we can have PLA-like pronouns at work in both coordinating and subordinating contexts.

Definition 4. (Index Semantics).

Suppose a first-order model M with domain of individuals D. Suppose σ is an infinite sequence of individuals from D assigned to the positions of σ. We will use

[5] Realising a dynamic negation in a dynamic framework like DPLE along the lines suggested by (18) is technically a very challenging problem. It is not clear that it is even possible, but see Visser (2002) for a partial answer as to how this could be achieved. The easy answer to negation is to offer the static version (i) that is part of Vermeulen's DPLE, but this defeats the promise of a fully dynamic system based only on coordinating relations.

(i) $(g,h) \in [\![\neg\phi]\!]_M$ iff $g = h$ and $\neg\exists j : (g,j) \in [\![\phi]\!]$

[6] Context can however be extended by the closure that comes with negation. Such extensions are subsequently lost on exit from negation's scope. This gives DRT-like accessibility effects.

$\sigma[k..m]$ for finite sequence $(\sigma_k, ..., \sigma_m)$, $\sigma[k..\omega]$ for the infinite sequence $(\sigma_k, ...)$, and $\sigma[-\omega..k]$ for the infinite sequence $(..., \sigma_k)$. We will write σ for $\sigma[-\omega..\omega]$.[7]

$$
\begin{array}{lll}
\sigma \models_k P(t_1, ..., t_n) & \text{iff} & ([\![t_1]\!]_{\sigma,k}, ..., [\![t_n]\!]_{\sigma,k}) \in M(P) \\
\sigma \models_k \lambda\phi & \text{iff} & \sigma \models_{k+1} \phi \\
\sigma \models_k \neg\phi & \text{iff} & \sigma[-\omega..k]\sigma'[1..n(\phi)]\sigma[k+1..\omega] \not\models_k \phi \quad \text{for all } \sigma'[1..n(\phi)] \\
\sigma \models_k \phi \wedge \psi & \text{iff} & \sigma \models_k \phi \text{ and } \sigma \models_{k+n(\phi)} \psi \\
[\![p_i]\!]_{\sigma,k} = \sigma_{k-i} & &
\end{array}
$$

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

The binding operator $\lambda\phi$ opens a new scope by advancing the index by one value, thus making one more value accessible inside ϕ. With a conjunct $\phi \wedge \psi$, crossing over to ψ from ϕ has the effect of advancing the index by $n(\phi)$ values, that is, by the number of instances of λ visible in ϕ. We can demonstrate this with renderings of (2) as (19), (5a) as (20), (6a) as (21), and (7a) as (22).

(19) $\lambda\texttt{enters}(p_0) \wedge \texttt{smiles}(p_0) \wedge \lambda\texttt{enters}(p_0) \wedge \texttt{smiles at}(p_0, p_1)$

(20) $\lambda\texttt{happy}(p_0) \wedge \texttt{you like}(p_0)$

(21) $\lambda ...\textit{thinks}...\texttt{you like}(p_0)$

(22) $\lambda\lambda(\texttt{likes}(p_0, p_1) \wedge \texttt{saw}(p_1, p_0))$

This captures the uniform role of linking played by pronouns in subordinating and coordinating relations. But this is achieved at the expense of turning all operator-variable dependencies into PLA-like pronominal bindings, including internal clause relations, such as argument links. As we noted when looking at PLA, PLA pronouns seem to capture the ad-hoc nature of pronominal linking, with the actual dependency that holds potentially relying on intervening (formula) material. However, it is not clear that this is appropriate for capturing core clause internal links, which are rigidly grammaticalised and unalterable.[8] What we appear to need is a system that has both the option of pronominal linking and variable name linking in subordinating contexts.

6 Scope Control Theory

In this section we illustrate a system that has both options of pronominal linking and variable name linking in subordinating contexts, and only the option of pronominal linking in coordinating contexts.

---

[7] The reason for employing infinite sequences is that it allows for a more conservative presentation and eliminates the possibility of undefinedness. The system would operate without change over finite sequences with sufficient values.

[8] That syntacticians have proposed ideas like the Minimal Link Condition (Chomsky 1995) suggests that this may in fact be on the right track, but we won't pursue this possibility here.

One way to get such a system would be to import predicate logic variable binding machinery into Index Semantics. While feasible, this would have the disadvantage of making two binding options available simultaneously. In natural language, when there are differing binding options, e.g., reflexive binding vs. pronominal binding, they tend to be in complementary distribution. What we need instead is for quantification to open up a variable binding which is subsequently turned into a pronominal binding. We can accomplish this with some of the dynamic control DPLE allowed over stack valued assignments.

First we define some stack relations and relation iteration.

Definition 5. (relations on stacks and iteration).

We introduce the following relations on stack valued assignments:

- For each $x, y \in V$ where $x \neq y$ we define $\texttt{shift}_{x,y} \subseteq SASS_D \times SASS_D$ as follows: $(g,h) \in \texttt{shift}_{x,y}$ iff $h(x)\star \uparrow(g(x)) = g(x)$, $g(y)\star \uparrow(h(y)) = h(y)$, $\uparrow(g(x)) = \uparrow(h(y))$ and $g(z) = h(z)$ whenever $y \neq z \neq x$.
- For each $x, y \in V$ where $x \neq y$ we define $\texttt{tfihs}_{x,y} \subseteq SASS_D \times SASS_D$ as follows: $(g,h) \in \texttt{tfihs}_{x,y}$ iff $\downarrow(g(x)) \star h(x) = g(x)$, $g(y)\star \uparrow(h(y)) = h(y)$, $\downarrow(g(x)) = \uparrow(h(y))$ and $g(z) = h(z)$ whenever $y \neq z \neq x$.

We will also use an iteration convention for relations on stack valued assignments:[9]

- $R^0 \equiv \top$
- $R^n \equiv R^{n-1}; R$

Having access to stack valued assignments and the iterable relations of definitions 2 and 5, we are in a position to introduce the system of Scope Control Theory (SCT) (cf. Butler 2007). This is given as a system in which formulas are evaluated with respect to a first-order model $M = (D, I)$ and a stack valued assignment $g \in SASS_D$.

Definition 6. (Scope Control Theory).

First, we define term evaluation:

$$[\![v_i]\!]_g = \begin{cases} d & \text{if } \exists h : (h,g) \in [v]^i \textit{ and } d = \uparrow(h(v)) \\ \textit{error} & \text{otherwise} \end{cases}$$

Next, we define formula evaluation:

$$\begin{array}{lll}
g \models_M \exists x\phi & \text{iff} & \exists h : (g,h) \in \texttt{shift}_{e,x} \text{ and } h \models_M \phi \\
g \models_M \rceil x\phi & \text{iff} & \exists h : (g,h) \in \texttt{tfihs}_{x,p} \text{ and } g_n \models_M \phi \\
g \models_M \phi \wedge \psi & \text{iff} & \exists h \left((h,g) \in [e]^{n(\psi)} \text{ and } h \models_M \phi\right) \text{ and} \\
 & & \exists h \left((h,g) \in \texttt{tfihs}_{e,p}^{n(\phi)} \text{ and } h \models_M \psi\right) \\
g \models_M \neg\phi & \text{iff} & \neg\exists h : (g,h) \in [e]^{n(\phi)} \text{ and } h \models_M \phi \\
g \models_M P(t_1, ..., t_n) & \text{iff} & ([\![t_1]\!]_g, ..., [\![t_n]\!]_g) \in I(P)
\end{array}$$

[9] Here, ‘;’ is relational composition: $(x,y) \in R; S$ iff $\exists z : (x,z) \in R$ and $(z,y) \in S$.

A term has form v_n in a predicate formula $\texttt{P(}...v_n...\texttt{)}$, and denotes the $n+1$ element of the v-stack. In practice, indices different from 0 appear rarely. Nevertheless, they can appear. Therefore, we use the convention that the index 0 can be omitted. This could easily be built into SCT by adjusting term evaluation to treat each occurrence of a term v as v_0. This convention has the advantage of making the SCT language a conservative extension of traditional predicate logic notation.

We can now demonstrate SCT with renderings of (2) as (23), (5a) as (24), (6a) as (25) and (7a) as (26).

(23) $\exists x \texttt{enters(}x\texttt{)} \wedge \texttt{smiles(}p\texttt{)} \wedge \exists x \texttt{enters(}x\texttt{)} \wedge \texttt{smiles at(}p, p_1\texttt{)}$

(24) $\exists x \texttt{happy(}x\texttt{)} \wedge \texttt{you like(}p\texttt{)}$

(25) $\exists x ...\mathit{thinks}...\rceil x \texttt{you like(}p\texttt{)}$

(26) $\exists x \exists y (\rceil x \texttt{likes(}y, p\texttt{)} \wedge \texttt{saw(}x, y\texttt{)})$

This illustrates the consistent use of variables named p (PLA-like pronouns) to capture natural language pronouns, and variable binding with variables named x and y to capture clause internal linking. Note how indexing plays a minor role, only needing to be explicitly coded once in (23) for a pronominal link to a lesser salient antecedent. This shows how typical use of SCT will exploit the dimension of variable names much more than the dimension of indices.

In fact, limiting the prospect of indices greater than 0 to p variables, and so the need to encode them, provides a way to single out the exceptional ad-hoc nature of pronominal linking, which is not enforced by grammatical structure alone (linking to alternative accessible antecedents from the discourse context or from superordinating localities is often an option, achieved with a change of index).

Also note how p variables without indices unambiguously link to what is, arguably, the most salient antecedent of the structured discourse context (e.g., in (25) this is the antecedent inherited from the superordinate clause).

7 Conclusion

In this paper, we saw how dynamism (specifically, an option of 'shifting' the variable name of a binding) in the context of subordinating relations, allowed for a uniform take on subordinating and coordinating pronominal binding relations. This result comes without disruption to the classical truth-conditional, propositional based perspective of meaning defended by Cresswell (2002) and Dekker (2002), which allows for an environment to be successively modified, but never passed on as an 'output.' Moreover, we saw how it is possible to preserve the binding option of classical variable name binding, which captures well the more restricted role of clause internal argument linking in natural language, and have a subsequent hand-over to a pronominal machinery in subordinating contexts with sufficient embedding, which captures well the ad-hoc nature of pronominal binding. The same pronominal machinery carries over as the means to establish dependencies across coordinating contexts.

References

[Butler(2007)]Butler, A.: Scope control and grammatical dependencies. Journal of Logic, Language and Information 16, 241–264 (2007)

[Chomsky(1981)]Chomsky, N.: Lectures on Government and Binding. Foris, Dordrecht (1981)

[Chomsky(1995)]Chomsky, N.: The Minimalist Program. MIT Press, Cambridge, Mass (1995)

[Cresswell(2002)]Cresswell, M.J.: Static semantics for dynamic discourse. Linguistics and Philosophy 25, 545–571 (2002)

[Dekker(2002)]Dekker, P.: Meaning and use of indefinite expressions. Journal of Logic, Language and Information 11, 141–194 (2002)

[Heim(1993)]Heim, I.: Anaphora and semantic interpretation: A reinterpretation of Reinhart's approach. Tech. Rep. SfS-Report-07-93, University of Tubingen (1993)

[Kamp and Reyle(1993)]Kamp, H., Reyle, U.: From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Kluwer, Dordrecht (1993)

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

[Vermeulen(2000)]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(2002)]Visser, A.: The donkey and the monoid: Dynamic semantics with control elements. Journal of Logic, Language and Information 11, 107–131 (2002)

Left-Peripheral and Sentence-Internal Topics in Japanese

Hiroaki Nakamura

Department of Liberal Arts, Japan Coast Guard Academy, 5-1, Wakaba-cho, Kure City, Hiroshima 737-8512, Japan
nakamura@jcga.ac.jp

Abstract. In Japanese, information structure is mainly indicated by the use of the topic marker WA. This paper will discuss how the use of this particle carries out the so-called Topic-Comment articulation from the viewpoint of the syntax-semantics interface, especially within a version of categorial grammars. Pragmatic characterizations of the topic marker have quite often been addressed in the literature without clarifying its syntactic properties, while the syntactic analysis of topicalization has been provided just in terms of hierarchical structure and movement in Japanese theoretical linguistics. We attend to a syntactic function of the topic marker, as suggested by the term *kakari-josi* or concord/coherence particle, which requires an expression marked with the particle WA to show a kind of 'concord' with the sentence-final verb. Semantically, we argue that the topic particle induces information packaging as suggested by Vallduví (1992), yielding tripartite information structures in the sense of Hajičová, Partee & Sgall (1998). We also deal with sentences with sentence-internal topics, cleft-constructions and multiple topic sentences in terms of incremental processing within the categorial proof net approach.

Keywords: topic particle WA, topic-comment articulation, information packaging, categorial grammar, syntax-semantics interface.

1 Topic Marking in the Literature

Kuno (1973) argues that topics convey old and predictable information while subjects marked with the nominative marker GA convey new and unpredictable information, but it is well known that a lot of expressions conveying (discourse-)new information can also be marked with WA (see Hinds 1987, Uechi 1992). Recently, following Diesing (1992), it has been argued that subjects of kind/individual-level predicates are marked with WA as default, while those of stage-level predicates should be marked with GA, though the latter may be followed by WA when their referents can be evoked from previous discourse or situations. Assuming that the subjects in generic sentences are required to have existential presuppositions, we can treat all WA-marked phrases as presupposition triggers because the referents of such expressions are not cancellable under negation.

In this paper I want to show that the topic marker WA serves to carry out the topic-comment articulation in two ways; (i) it picks up an expression denoting an

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 160–173, 2008.

entity/entities already familiar from previous discourse as a locus of information update and (ii) it assembles the remaining parts of the sentence into an information unit which updates the information content of the entity/entities. We adopt tripartite structure like (1) for a unified semantic analysis of sentences with topics following Hajičová, et al. 1998, and write the topic-comment segmentation to save space as in (2)(see Nakamura 2006):

(1) TOPIC(x)
Restriction: Predicate1(...x...)
Nuclear Scope: Predicate2(...x...)

(2) TOP x [Px → Qx] (P $\neq \varnothing$)

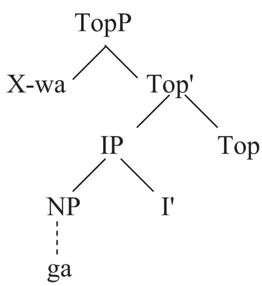
We would also like to present the parsing process of sentences containing topics on a left-to-right, word-by-word basis. Semantic interpretations like (1) or (2) are derived from the resulting parse trees on the basis of the principle of compositionality. To capture the syntactic and semantic properties of the topic marker WA from a dynamic perspective, let us assume first that the topic marker WA has the context change potential (henceforth CCP) as in (3) in the sense of Heim (1983), where CCPs are defined as instructions specifying certain operations of context change. Suppose that her description of the CCP of "if" also holds for the CCP of the topic marker WA.

(3) c + A-*WA* B = c \ (c + A \ c + A + B)

The instruction of WA is to conjoin the current context with the presupposition A (denoted by a WA-marked expression), and the resulting context is updated by the new information B (denoted by the remaining part of the sentence. That is, c (= the current context) admits "A-WA B" only if c admits A and c + A admits B. Different from standard conditionals, WA-marked phrases require their semantic contents to be true because presuppositions expressed by the WA-marked phrases survive when they are negated, questioned, or used as a hypothetical assumption, so if A is false, the value of a whole sentence with a topic must be undefined, not true. Alternatively, we can resort to a metaphor of a 'file' in the file change semantics (Heim 1983). If an expression is marked with the discourse particle WA, a speaker takes it for granted that a hearer already have the card corresponding to its referent in his/her file. This card designates a locus of information update. We will implement such changes in information state in the incremental parsing of topic sentences.

Syntactically, in the generative grammar tradition, the differences between subjects and topics have been captured in terms of configurational positions in the tree structure. Hasegawa (1999:92), for instance, posits the following structure to distinguish the syntactic levels for subjects and topics:

(4)



In the tree structure like (4), it is assumed that topic-marked elements show up only at the left peripheral position of sentences, but this assumption is simply wrong. As we will see later, topicalization can apply to any element in any position though topics occurring sentence-internally tend to be interpreted as conveying contrastive readings. Furthermore, the structure as in (4) cannot offer a mechanism mapping syntactic structures to intended semantic interpretations or information structures.

2 Deriving Topics: A First Approximation

Let us start with comparing the derivations of a topicless sentence (with a nominative-marked subject) and a generic sentence (with a 'topicalized' subject) within the framework of the Combinatory Categorial Grammar (henceforth, CCG). Observe the following pair of sentences:

(5) a. Uma-wa ninjin-o taber-u.
horse-Top carrots-Acc eat-Pres
'Horses eats carrots.'
b. Uma-ga ninjin-o tabete-ir-u.
horse-Top carrots-Acc eat-Prog-Pres
'Horses are eating carrots.'

The subject of the kind/individual-level predicate *taberu* 'eat' in (5a) strongly tends to be marked with the topic marker and, if it is marked with the nominative GA, the sentence forces an exhaustive listing reading. In (5b), the GA-marked subject of the progressive form *tabete-ir-u* 'are eating', a typical stage-level predicate, introduces new discourse referents to the hearer's information state. If the referents have already been present or inferable from the discourse/situation, they can be marked by WA. It should be noted here that it is dubious to directly associate the WA/GA-marking of subjects with the kind/individual-level predicate and stage-level predicate distinction. Consider the following question-answer pair.

(6) a. Nani-ga/*Nani-wa ninjin-o taber-u-no?
What-Nom/ -Top carrots-Acc eat-Pres-Q
'What eats carrots?'
b. Uma-ga/*?Uma-wa tabe-masu.
Horse-Nom/ -Top eat-Pres.
'Horses do.'

Question words and their answers (i.e., foci) can not be marked by the topic marker, at least under normal circumstances[1], whether they occur in kind/individual sentences or in stage-level sentences. Buring (1999), assuming that the Focus value, which is equal to the value of a question ($\cup[[A]]^f = \cup[[Q]]^o$), is a set of propositions (worlds), argues that "the Topic value is basically a 'typed up' Focus value, i.e. a set of

[1] Only the presuppositional/referential question form of dono 'which' + N, e.g., *dono doobutu-ga* 'which animal(s) ...?', can be topicalized, with a obligatory contrastive reading, as in (i).

(i) Dono doobutu-ga/-wa ninjin-o taber-u-no/tabe-teir-u-no?
Which-animal-Nom/-Top carrots-Acc eat-Pres-Q /eat-Prog-Pres-Q
'Which animals eat/are eating carrots?

sets of propositions"(Buring 1999:147). This semantic definition of focus/topic can be said to be similar to the idea of RESTRICTOR/NUCLEAR SCOPE distinction of Hajičová, Partee & Sgull (1998). We do not go into the details of the semantics of topics/foci here, but we extend his definition to cover topics in general and define the meaning of the topic as a relation of sets, as shown in (2). We also consider the contrastive reading of topics in section 5.

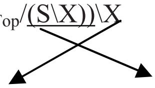
We assume some familiarity with categorial grammars and limit ourselves here to reviewing some rules necessary for our analysis. Combinatory Categorial Grammar (hereafter CCG) is a mildly context sensitive formalism, defining a set of combinatory rules to deal with standard and nonstandard surface constituency (including unbounded dependency) flexibly, while maintaining direct compositionality. Only the three rules of concatenation in CCG are relevant for our purpose here:

(7) a. $X/Y{:}f \quad Y{:}a \quad \Rightarrow \quad X{:}fa$ $\qquad Y{:}a \quad X\backslash Y{:}f \quad \Rightarrow \quad X{:}fa$
b. $X/Y{:}g \quad Y/Z{:}f \quad \Rightarrow_{\mathbf{B}} \quad X/Z{:}gf$ $\qquad Y\backslash Z{:}f \quad X\backslash Y{:}g \quad \Rightarrow_{\mathbf{B}} \quad X\backslash Z{:}gf$
c. $X{:}a \quad \Rightarrow_{\mathbf{T}} \quad T\backslash(T/X){:}\lambda f.fa$ or $T/(T\backslash X){:}\ \lambda f.fa$

Following Steedman (1996, 2000), the 'result leftmost' notation is used here in which a rightward-combining functor over a domain Y into a range X is written as X/Y, while the corresponding leftward-combining functor is written as X\Y. (7a) is the rule of function application to concatenate expressions in the canonical order. An expression of functional category X/Y combines with an adjacent argument of category Y to yield a result of category X and interpretation *fa*, the result of applying *f* to *a*. This rule, for example, combines a transitive verb with an object to yield the verb phrase, and then, combines the verb phrase with a subject to produce the sentence. The rule of function composition (7b) allows a main function of category X/Y to combine with a subordinate function of category Y/Z to yield a function of category X/Z. (7c) is the rule of type-raising, which we need to deal with a wide range of topicalized expressions in subsequent sections. The rule turns an argument category X into a functor over X. For instance, this operation converts a subject NP, which would normally be an argument to a verb phrase of category NP\S, into a function looking forward for a verb phrase, S/(S\NP), to produce a sentence. We also assume the order-preserving associativity (rebracketing) in our grammar.

Let us define the category for the topic marker WA as in (8), which serves as a (lexical) type shifter:

(8) Category for topicalized phrases
X-WA := $S_{Top}/(S\backslash X)$: TOP(x) [P(x) $\rightarrow$ Q(x)]
-WA := $(S_{Top}/\underline{(S\backslash X))}\backslash X$ where X = NP, PP, or S



restrictor (X) nuclear scope (S\X)

The category of the topic particle WA in (8) lifts the expression to be topicalized, such as a common noun, to a functor taking an open proposition, to yield a (matrix) sentence with the topic-comment structure. A topicalized expression of category X corresponds to a restrictor (i.e., the locus of information update) and the remaining part of the sentence of category S\X represents the nuclear scope, which updates or adds information to the context specified by the former. The category (8) for WA finally results in a tripartite information structure including the TOPIC operator. Another function of (8) is to package the remaining part of a sentence with a gap corresponding to the topicalized expression as the concord particle called KAKARI-JOSHI in Japanese traditional grammar. Following the definition, an expression of any category can be marked with the topic marker WA. In terms of semantics, the category for WA indicates that a higher functor category is assigned to a topicalized expression. The meaning denoted by this raised category is completely compatible with the semantics of topics proposed by Buring (1999) or Hajičová, et al. (1998). The category for WA in (8) stands for a kind of generalized quantifier, the meaning of which is defined as a relation of sets. The derivation of sentence (5a) can be shown as in (9):

(9) Uma-wa ninjin-o taber-u

$S_{Top}/(S\backslash NP)$:*horse* NP_{Obj}:*carrots* $(S\backslash NP_{Sub})\backslash NP_{Obj}$:*eat*

TOP x[*horse*(x)→P(x)]

← application

$S\backslash NP_{Sub}$:λx.*eat*(*carrots*)(x)

application →

S_{Top}: TOPx [*horse*(x)→*eat*(*grass*)(x)]

The category which the topic marker WA first combines with is underspecified in definition (8), so any expression can be topicalized, as we will see later, and the remaining part of the sentence can be packaged even if it results in a nonstandard constituent. (9) properly shows how to derive a sentence with a topic compositionally, but it does not reflect the dynamic process of information update suggested by the context change potential of WA in (3). Since Japanese is a head final language in which functors usually apply to arguments backwards, the derivations require all the elements to be present from the beginning though non-incrementality is partially improved by raising categories of topicalized expressions. The processor does not need to know a position and category of a gap (corresponding to a topicalized expression) in the nuclear scope in advance because the category of the topic particle contains the underspecified category-specification for the gap, i.e., X in $S_{Top}/(S\backslash X)$.

Since information update indicated by the particle's CCP should be carried out in a time-linear way, in principle, we want to devise an incremental (left-to-right) parsing and interpretation strategy. In the next section we present the syntactic parsing of sentences with topics, implementing incrementality of information update. We also show how interpretations of sentences with topics (tripartite structures as in (1) and (2)) can be obtained from resulting syntactic structures.

3 Proof Net Approach to the Topic-Comment Articulation

There have been a lot of proof net approaches proposed in theoretical/computational linguistics in recent times. We adopt the categorial proof net analysis advocated by Morrill (2000, 2004). The theory of proof nets are based on Lambek categorial grammar, the derivations of which are usually presented via sequent caluculus for concatenation as in (10). Due to lack of space, we only show some basic assumptions of Type-Logical Grammar here:

(10) a. $A \Rightarrow A$ id. $\qquad \dfrac{\Gamma \Rightarrow A \quad \Delta(A) \Rightarrow B}{\Gamma(\Delta) \Rightarrow B}\text{Cut}$

b. $\dfrac{\Gamma \Rightarrow A \quad \Delta(B) \Rightarrow C}{\Delta(\Gamma, A\backslash B) \Rightarrow C}\backslash L \qquad \dfrac{A, \Gamma \Rightarrow B}{\Gamma \Rightarrow A\backslash B}\backslash R$

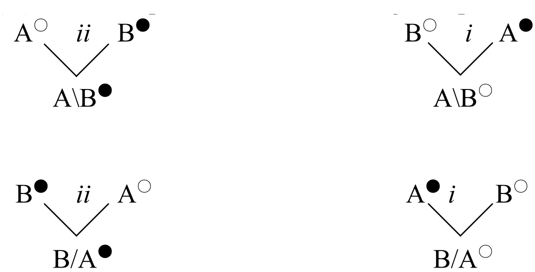
c. $\dfrac{\Gamma \Rightarrow A \quad \Delta(B) \Rightarrow C}{\Delta(B/A, \Gamma) \Rightarrow C}/L \qquad \dfrac{\Gamma, A \Rightarrow B}{\Gamma \Rightarrow B/A}/R$

d. $\dfrac{\Gamma(A,B) \Rightarrow C}{\Gamma(A \cdot B) \Rightarrow C}\cdot L \qquad \dfrac{\Gamma \Rightarrow A \quad \Delta \Rightarrow B}{\Gamma, \Delta \Rightarrow A \cdot B}\cdot R$

Following Roorda (1991), Morrill exploits proofnets for the syntactic (and semantic) analysis of a wide range of discontinuous constructions under the slogan 'syntactic structures as proof nets.' The concatenation rules can be reproduced for the construction of proof nets, as follows. A polar category formula is shown by a Lambek categorial type labelled with input ($^\bullet$) or output ($^\circ$) polarity. A category formula derives a binary ordered tree, a logical link, in which the leaves are labeled with polar atoms. Some examples of logical links relevant for our analysis can be shown in (11) (for the complete definition of categorial proof nets, see Morrill 2004):

(11) a.

b.



The polarities indicate sequent sidedness, input for antecedent and output for succedent. The polarity propagation follows the sidedness of subformulas in the sequent rules as illustrated in (11); in the anteedent (input) rule for A\B the subformula A goes in a succedent (output) and, in the succedent (output) rule, the subformula B goes in an antecedent (input) etc. The labels *i* and *ii* indicate whether a rule with *i* or *ii* is unary or binary. In (11), note that in the output links the order of subformulas is switched; the category with output polarity is adjacent to the first label with input polarity. A frame is a list comprising a unique output polar type tree followed by some input polar type trees. A proof net is the result of connecting by an identity link every leaf in a frame with a complementary leaf. Proof nets must satisfy the correctness criteria in (12) (Morrill 2004):

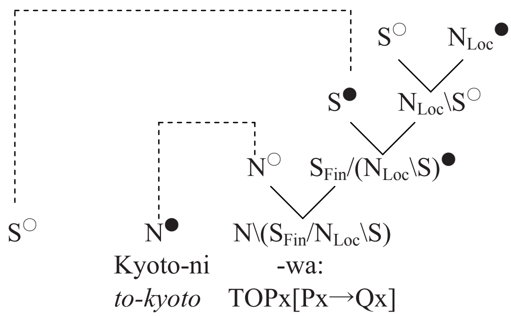
(12) planarity — The identity links are planar in the list of ordering.
acyclicity — Every cycle crosses both edges of some *i*-link.
no subtending — No identity link connects the leftmost and rightmost descendent leaves of an output division node.

Building proof nets are carried out on a word-by-word basis. Let us see how parsing of a sentence with a topic proceeds.

(13) Kyoto-ni-wa takusan-no kankoo-kyaku-ga maitoshi yattekur-u.
Kyoto-Loc-Top a lot of tourists-Nom every_year come-Pres
'To Kyoto, a lot of tourists come every year.'

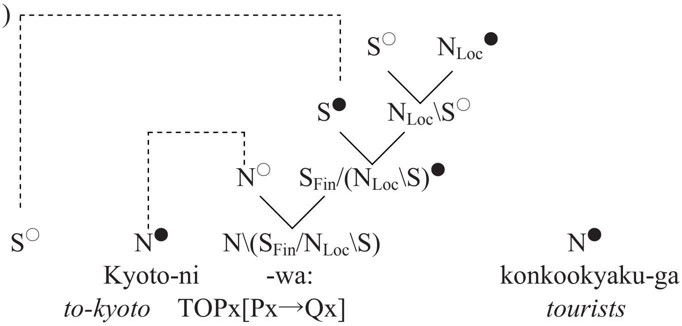
Initially, an S (with output polarity) is expected, and then when the first word, the locative PP in this case, is perceived, there is no identity link in the structure. After the topic particle is processed, we have a partial proof net like (14):

(14)



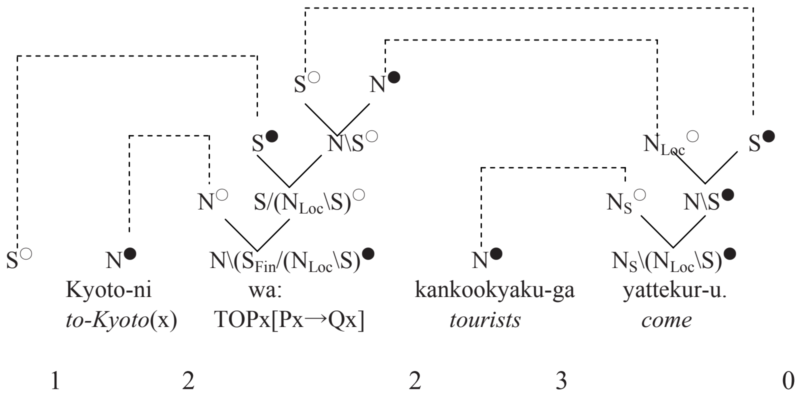
At this point the identity links connect four complementary leaves, and there remain two unmatched valencies (unresolved dependencies) in the net. The number of unresolved dependencies indicates a measure of the course of memory load in optimal incremental processing. After processing of the subject *kankookyaku-ga* 'tourists', we have the intermediate proof net like (15):

(15)



Because no leaf of the topic category can be linked with leaf $N^{\bullet}$ of the subject NP, we have three unmatched literals at this stage. After the verb which the topic is in concord with is processed, the unmatched leaves we have so far can be connected with proper complementary leaves, so we have the complete proof net analysis for sentence (12), as shown in (16). We write the number of unresolved dependencies (overarching identity links), referred to as 'cut', below the proof net.

(16)

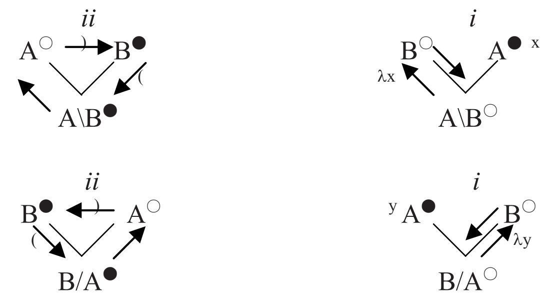


Within the categorial proof net approach, topic-comment structures can be parsed in a time-linear fashion, allowing underspecified dependencies at each intermediate stage of the parsing process, from which we can obtain the complexity profile of a sentence indicated by the numbers of unbounded dependencies (overarching identity links) at each word boundary. Morrill (2004) notes maximal cuts and average cuts. Notice the numbers of cuts written at the bottom of net (16). Here the average cut is 1.6 and the maximal cut is 3. In Section 5 we will examine sentences in which topic-marked elements appear sentence-internally, and compare the complexity profiles of sentences with left-peripheral and sentence-internal topics.

Next, let us see how we can derive the intended interpretation, including the topic meaning defined in (8), from the semantic trip of a proof net, thanks to the syntax-semantics correspondence. Morrill (2004) defines the semantic trip travel instructions as in (17), to get the semantics of sentences from proof nets, among which only four of them are relevant to our semantic analysis:

(17) a.

b.



The semantic trip is the trip starting upwards from the unique output root S$^{\circ}$ and traveling on the net, yielding the associated λ-terms as it proceeds, according to the instructions in (17). The trip bounces with the associated semantic form at input roots and ends when it returns to the origin S$^{\circ}$. From the proof net in (16), we can produce its meaning by the step-by-step interpretation process, which is omitted here due to space limitations. The normalized semantics resulting from the trip on net (16) should be something like (18):

(18) TOP x [*to_Kyoto*(x) → *come*(x)(*tourists*)]

Following the semantic trip travel instructions in (17), this interpretation is automatically derived from the proof net analysis in (16). (18) represents the tripartite structure we got through the topic-comment articulation executed by the topic marker WA, comprising the topic operator, the restrictor (presupposition), and the nuclear scope (update of context). Though the same semantic representation can be derived by the combinatory rules of CCG, the categorial proof nets are completely built incrementally, matching with the cognitive process in communication. The resulting proof nets give complexity profiles of sentences indicated by the number of underspecified dependencies as load on memory in the course of parsing. In Section 5, we will suggest that complexity profiles resulting from parsing of sentences affect the implicatures of sentences, and partially account for contrastive readings of topicalized expressions.

4 Parsing of Cleft Constructions and Sentence-Internal Topics

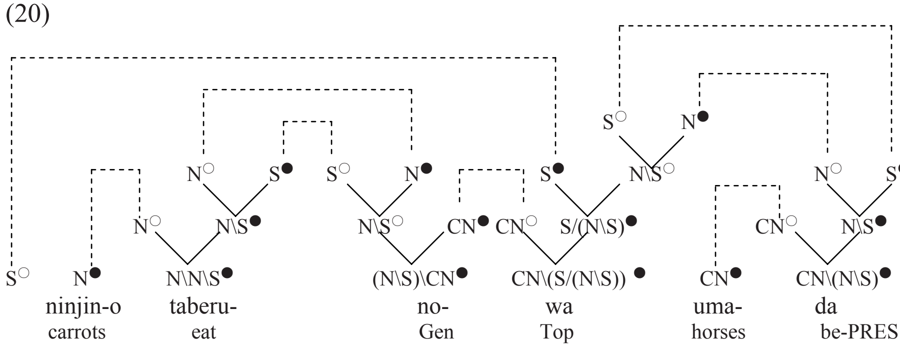
The many recent studies on topicalization have paid much attention only to sentences with the first elements marked with the topic marker, exactly as we have seen so far, often placing emphasis on the differences between subjects and topics. In this section, let us consider more complicated topicalization phenomena, focusing on sentences with higher complexity, including cleft constructions and multiple topic constructions.

In a typical cleft construction, the presupposed element is an open proposition and the nuclear scope comprises a focused expression corresponding to the gap/variable in the former, followed by the copula. A case marker other than the nominative and accusative markers may intervene between a focus and a copula. Observe the sentence in (19). The presupposition and focus parts of (5a) are reversed here and the existence of something which eat(s) carrots is presupposed in (19).

(19) Ninjin-o taber-u-no-wa uma-desu.
carrots-Acc eat-Pres-Nomizalizer-TOP horse-Be-Pres
'It is horses that eat carrots.'

The proof net for (19) is shown in (20)[2]:

(20)



[2] In Japanese, the order of arguments of a verb is relatively free, so the category of a verb needs to be defined to allow a degree of permutation of arguments. Many proposals have been advocated to deal with scrambling phenomena, but we do not go into the details here, assuming that the order of arguments/adjuncts can freely be switched, as in X/(Y/Z) = Y/(X/Z).

The category of the presupposed open proposition is converted to a common noun (CN) by the genitive case marker NO in (20), which should be taken as a nominalization operator mapping property-denoting expressions onto property-correlates in the domain of entities (see Chierchia 1984, Partee 1986). Following Chierchia (1984), we indicate this operator as '$^\cap$', so, for instance, the nominalized open proposition in (20) is translated as $^\cap$*eat*(*carrots*)(*x*). After a predicative expression (of type <e,t>) is shifted to an individual-type expression (of type e), the nominalized open clause is topicalized in the derivation. Our semantic trip proceeds from the output polar formula S° to the input polar leaf $S^\bullet$, then going down to the topic marker WA, so the derived interpretation is TOP x[P(x)→Q(x)] at this stage. It proceeds to the nominalized open proposition marked with WA (i.e., the restrictor) and, then, turns to the focus part. Finally, we get the semantics like (21) (assuming that the meaning of the copula *desu* ' is λxλy.=(x)(y):

(21) TOP x[$^\cap$*eat*(*carrots*)(x) → =(*horse*)(x)]

So far we have dealt with constructions in which expressions marked by WA appear sentence-initially, which is in harmony with the standard process of information update, but the topic particle can also mark sentence-internal expressions, which should be problematic for a simple tree structure analysis like (4). Let us take (22) as an example.

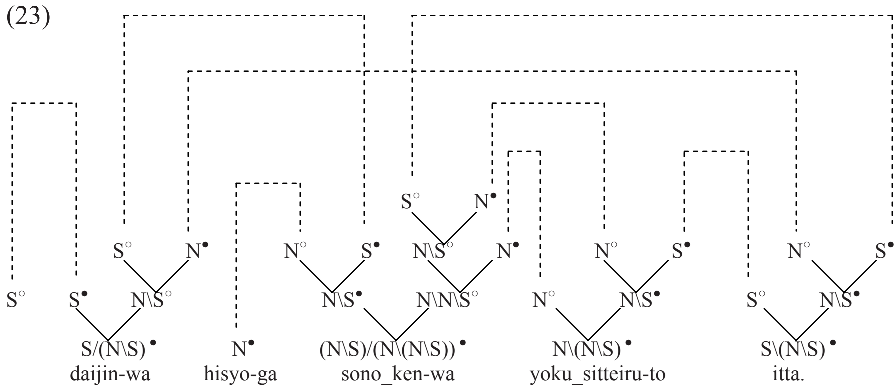
(22) Context: A minister is questioned at the congress about the bribes he alledgedly received. An article in the newspaper reports:

Daijin-wa	hisyo-ga	sono_ken-wa	yoku	sitteiru-to	itta.
minister-TOP	secretary	that issue-TOP	well	know	said

'The minister said that his secretary knows that issue well (better than he).'

Because the familiar information includes the law-maker and that issue (bribery), and the other parts belong to new information in (22), the elements conveying old and new information are intermixed with each other, which appears to make information packaging at surface structure level extremely difficult. The proof net for (21) should be something like (22), where we allow the WA-marked object noun to be lifted to a functor taking a transitive verb as input and returning an intransitive verb phrase as output :

(23)



Notice that the output category S° of the topic in the embedded clause is connected by the identity link with the input category $S^{\bullet}$ of the matrix verb, reflecting the coherence of the topic and the sentence-final verb, which is required by the lexical property of the topic marker as KAKARI-JOSHI or concord/coherence particle. Here we need to modify the category for the topic particle so that expressions other than subjects can be topicalized. The generalized category for WA should be written as X\(S$/(S\N)$ using the $ symbol proposed by Steedman (1996, 2000)[3]. After the semantic trip, we get the intended interpretation from the proof net, as in (23):

(24) TOP x[*minister*(x)→(TOP y[*that_issue*(y)→*said*(*knew_well*(y)(*secretary*))(x)]]

Topic segment Comment segment

(24) shows the result of (partial) information packaging, in which the expressions are grouped into the two segments. The elements conveying old/predictable information are put on the left-hand side and those conveying new/unpredictable information on the right-hand side. Actually, (24) merely represents the layers of topic-comment structures and is not exactly what we have wanted as the topic-comment articulation. An appropriate information structure should be something like TOPxTOPy[(*minister*(x) & *that_issue*(y))→*said*(*knew_well*(y)(*secretary*))(x)]. Therefore, we need an additional device mapping (23) to a complete information structure, but we have to leave this device unexplored here.

5 Contrastive Reading and Complexity Profiling

From the viewpoint of information update, it is natural and preferable for topicalized items to show up left-peripherally in natural languages. Though topics are morphologically marked by the discourse particle in Japanese, there is also a pronounced syntactic tendency for WA-marked expressions (of any category) to be left dislocated, but left-dislocation of topics is not compulsory. It is widely assumed that sentence-initial topics can be thematic (anaphoric) or contrastive, whereas setence-internal (or sentence-final) topics overwhelmingly convey contrastive readings (though it is sometimes difficult to distinguish the anaphoric use and contrastive use of topicalized expressions). Sentence-internal topics appear to be incompatible with the natural process of information update, but they may use a nonstandard order of topic-comment constituency to convey some implicature, referring to alternative elements. Out categorial proof-net approach can account for preferences of time-linear information update process, via incremental processing of sentences with topics.

How then can our approach associate nonstandard ordering of topics with contrastiveness? Though the contrastive use of WA is quite often accompanied by prosodic features such as stress or intonation, let us simply ignore them here. Compare a typical example of a sentence-internal topic in (25a), with sentence (12), repeated here as (25b).

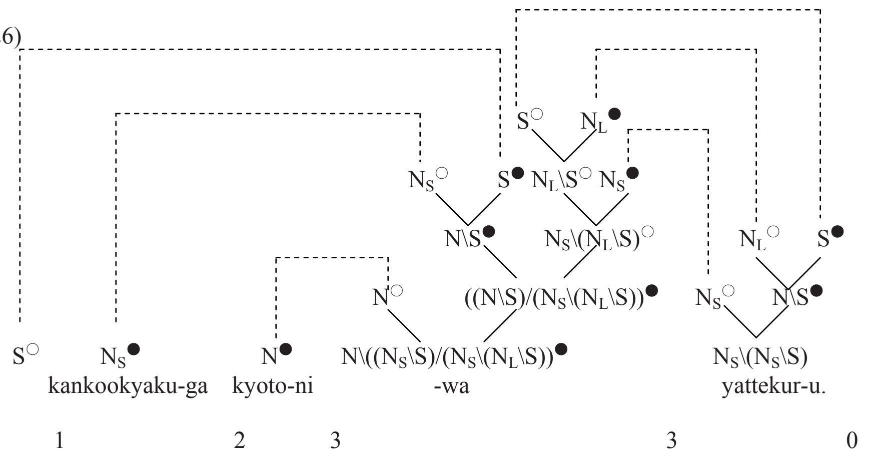
[3] The $ convention: For a category α, {α$} (respectively, {α/$} or {α\$}) denotes the set containing α and all functions (respectively, leftward functions, rightward functions) into a category in {α$}(respectively, {α/$}, {α\$}). (Steedman 2000:42(32))

(25) a. Takusan-no kankoo-kyaku-ga Kyoto-ni-WA maitoshi yatte-kuru.
a lot of tourists-NOM Kyoto-To-Top every_year come-Pres.
'A lot of tourists come to Kyoto every year.'

b. Kyoto-ni-wa takusan-no kankoo-kyaku-ga maitoshi yattekur-u.
Kyoto-Loc-Top a lot of tourists-NOM every_year come-Pres
'To Kyoto, a lot of tourists come every year.'

It should be noticed that even if WA-marked phrases are used contrastively, the speakers assumes the hearers' familiarity with their referents. Topics invoke alternative elements implicitly or explicitly. Sentence (25a) contains the sentence-internal topic, so this WA is not compatible with the natural information flow described by the particle's CCP in (3). Thus, it conveys an implicature, for example, that other cities do not attract so many tourists as Kyoto, etc. (26) is the proof net for a simplified version of (25a), from which we can calculate its complexity as shown below the net.

(26)



The maximal cuts are the same in nets (16) and (26), but the nets differ in the average cuts, i.e., 1.6 in (16) and 1.8 in (26). Extra effort is also necessary to parse sentences with sentence-internal topics, which is type-raised to the category N\((N\S)/(N\(N\S)). As in the canonical topic sentence like (16), the sentence internal WA-marked phrase in (26) should also be understood anaphorically, but a littler higher complexity seems to force contrastive implicature in (26), while the sentence-initial topic in (16) is in accordance with the natural flow of information, presenting the locus of update first, followed by the update of information content. It is natural that the profiling for (26) with higher complexity is partially responsible for the contrastiveness reading. In principle, the two readings of WA share the common property, expressing the speaker's assumption of hearer's familiarity with their denotations.

6 Conclusion

In this study, we assumed that the particle WA combines with expressions of any category to express the speaker's assumption that their referents are already present in the hearer's file or registry (in the sense of Kuno), and the speaker's instruction for hearers to retrieve an existing file card for information update. Syntactically, we noticed an important characteristic of the topic marker as a concord/coherence particle. Following Japanese traditional grammars, expressions marked with the particle are taken to show a kind of concord with sentence final verbs. The functions of the topic marker are twofold: presenting a WA-marked expression as the locus of information update, and packaging the remaining part of the sentence as update of a hearer's information state. The context change potential is encoded as a lexical property of the topic particle.

We showed how to process sentences with various topics incrementally in terms of the categorial proof net approach, which reflects natural flow of information and translate surface strings into intended interpretations, implementing the topic-comment articulation according to the semantic travel on the resulting net. At each stage of processing, the open leaves, unresolved dependencies, indicate the load on working memory for hearers. So, unlike the parsing process of Dynamic Syntax approach (Kempson et. al. 2001, Cann et. al. 2005), which also realizes time-linear processing of an input sentence, the categorial proof net can retain the history of processing in which the complexity in understanding a sentence in question is shown explicitly. The complexity can also be indicated by the two numbers of cuts, a maximal cut and average cut. Morrill suggests that a lot of phenomena like garden paths or scope preferences can be accounted for by the complexity profiles of processed sentences. In Japanese, the topic marked phrases can show up sentence-internally, or sentence-finally, which appear to impose difficulty on the rigid configurational analyses of topics in generative grammars. Our flexible categorial grammar can assign proper categories and interpretations to topicalized expressions to derive a wide range of sentences with topics.

Categorial proof nets resulting from incremental parsing can yield proper unique interpretations for sentences through semantic trips, thanks to the Curry-Howard correspondence. I showed the tripartite information structures for sentences containing topics in various positions can be derived automatically from resulting syntactic proof nets.

References

1. Buring, D.: Topic. In: Bosch, P., van der Sandt, R. (eds.) Focus – Linguistic, Cognitive, and Computational Perspectives, pp. 142–165. Cambridge University Press, Cambridge (1999)
2. Cann, R., Kempson, R., Marten, L.: The Dynamics of Language: An Introduction. Elsevier Inc., San Diego (2005)
3. Carlson, G.: A Unified Analysis of the English Bare Plural. Linguistics and Philosophy 1, 413–458 (1977)
4. Chierchia, G.: Topics in the Syntax and Semantics of Infinitives and Gerunds. Ph. D. Thesis. University of Massachusetts, Amherst (1984)
5. Chierchia, G.: Formal Semantics and the Grammar of Predication. Linguistic Inquiry 16, 417–443 (1985)
6. Diesing, M.: Indefinites. MIT Press, Cambridge (1992)

7. Hajičová, E., Partee, B.H., Sgall, P.: Topic-Focus Articulation, Tripartite Structures, and Semantic Content. Kluwer Academic Publishers, Dordrecht (1998)
8. Hasegawa, N.: Seisei-Nihongogaku Nyumon, Taishukan, Tokyo (1999)
9. Heim, I.: On the Projection Problem for Presupposition. In: Barlow, M., Flickinger, D., Wescoat, M. (eds.) The Proceedings of the Second West Coast Conference on Formal Linguistics, pp. 114–125. CSLI Publications, Stanford (1983)
10. Hendriks, H.: Information Packaging: From Cards to Boxes. In: van Deemster, K., Kibble, R. (eds.) Information Sharing, pp. 1–33. CSLI Publications, Stanford (2002)
11. Hinds, J.: Thematization, Assumed Familiarity, Staging, and Syntactic Binding in Japanese. In: Hinds, J., Maynard, S.K., Iwasaki, S. (eds.) Perspectives on Topicalization: The Case of Japanese 'WA', John Benjamins, Amsterdam (1987)
12. Kempson, R., Meyer-Vior, W., Gabbay, D.: Dynamic Syntax. Blackwell, Oxford (2001)
13. Kuno, S.: The Structure of the Japanese Language. MIT Press, Cambridge (1973)
14. Masuoka, T.: Meidai-no Bunpoo–Nihon-Bunpoo-Josetsu. Kuroshio Publishers, Tokyo (1987)
15. Miyagawa, S.: Wa and the WH Phrase. In: Hinds, J., Maynard, S.K., Iwasaki, S. (eds.) Perspectives on Topicalization: the Case of Japanese 'WA', pp. 185–217. John Benjamins, Amsterdam (1987)
16. Morrill, G.: Incremental Processing and Acceptability. Computational Linguistics 26(3), 319–338 (2000)
17. Morrill, G.: Type Logical Grammar. In: The course reader distributed at ESSLLI 2004, Nancy (2004)
18. Nakamura, H.: Topic-Comment Articulation in Japanese: A Categorial Approach. In: Tingting, H., Maosong, S., Qunxiu, C. (eds.) Proceedings of the 20th PACLIC, pp. 198–205. Tsinghua University Press, Beijing (2006)
19. Noda, H.: "Wa" and "Ga". Kuroshio-Publishers, Tokyo (1996)
20. Partee, B.H.: Noun Phrase Interpretation and Type-shifting Principles. In: Groenendijk, J., de Jongh, D., Stokhof., M. (eds.) Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers, pp. 93–123. Foris Publications, Dordrecht (1986)
21. Roorda, D.: Resource Logics: Proof-theoretical Investigation, Ph. D. Thesis. Universiteit van Amsterdam (1991)
22. Shirai, K.: Keishiki-imiron Nyuumon, Sangyoo-Tosyo, Tokyo (1985)
23. Steedman, M.: Surface Structure and Interpretation. MIT Press, Cambridge (1996)
24. Steedman, M.: The Syntactic Process. MIT Press, Cambridge (2000)
25. Takami, K., Kamio, A.: Topicalization and Subjectivization in Japanese: Characterization and Identificational Information. Lingua 99, 207–235 (1996)
26. Uechi, A.: Toward Syntax-Information Mapping. In: Japanese/Korean Linguistics, vol. 5, pp. 387–403. CSLI Publications, Stanford (1996)
27. Uechi, A.: An Interface Approach to Topic/Focus Structure in Japanese. Ph. D. Thesis. The University of British Columbia, Vancouver (1998)
28. Vallduví, E.: The Information Component. Garland (1992)
29. Vallduví, E., Engdahl, E.: The Linguistic Realization of Information Packaging. Linguistics 34, 459–519 (1996)

Incremental Processing and Design of a Parser for Japanese: A Dynamic Approach

Masahiro Kobayashi

University Education Center, Tottori University,
4-101 Koyama-chō Minami, Tottori, 680-8550 Japan
kobayashi@uec.tottori-u.ac.jp

Abstract. This paper illustrates a parser which processes Japanese sentences in an incremantal fashion based on the Dynamic Syntax framework. In Dynamic Syntax there has basically been no algorithm which optimizes the application of transition rules: as it is, the rules can apply to a current parsing state in an arbitrary way. This paper proposes both partitioned parsing states allowing easier access to some kind of unfixed nodes and an algorithm to apply transition rules for Japanese. The parser proposed in this paper is implemented in Prolog. The parser is able to process not only simple sentences but also relative clause constructions, scrambled sentences and complex (embedded) sentences.

1 Introduction

Incremental and dynamic approaches to sentence comprehension and grammar formalisms have attracted a great deal of attention both in psycholinguistics and natural language processing. For example, Lombardo, Mazzei and Sturt (2004) discuss the relationship between competence and performance in incremental parsing based on Tree Adjoing Grammar. This paper is meant to describe a parser and its algorithm for a fragment of Japanese based on the incremental, left-to-right parsing formalism in the Dynamic Syntax framework (Kempson, Meyer-Viol and Gabbay 2001, Cann, Kempson and Marten 2005) and show how the parser processes some of the constructions in Japanese. In advocating Dynamic Syntax, Kempson et al. (2001) and Cann et al. (2005) have shown that the formalism is able to cope with many constructions regarding word order phenomena, such as cross-over constructions and topicalizations in English and also that it has a broad typological perspective. However there is no algorithm specified for computational implementation. In other words, Dynamic Syntax is an abstract grammar formalism with lexical rules and transition rules, and no algorithm has been proposed which optimizes the application of transition rules. In addition, the formalism makes the parser keep NPs unfixed within the tree structures at a certain parsing stage and narrow down the possibilities of outputs by specifying their positions. This means that the parser needs to deal with not only usual binary tree structures but also unfixed tree structures whose positions must be fixed in the final tree structure.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 174–186, 2008.

This paper has two aims. The first is to propose partitioned parsing states with easier access to fixed and unfixed nodes. The second aim is to propose an algorithm to implement the application of lexical rules and transition rules for Japanese. Purver and Otsuka (2003) and Otsuka and Purver (2003) proposed a parsing and generation model for English based on the Dynamic Syntax formalism, but as far as I know, there is no research so far which attempts to implement Dynamic Syntax for Japanese. The parser can currently process not only simple sentences but also relative clause constructions and embedded clauses. A Prolog implementation of the parser and efficient parsing algorithm will also be presented including a comparison to Purver and Otsuka (2003) and Otsuka and Purver (2003).

The outline of this paper is as follows: the subsequent section will be devoted to a brief look at the Dynamic Syntax formalism and issues this paper will deal with. This section will also describe previous studies. Section 3 will illustrate the parser for Japanese on the basis of Dynamic Syntax and propose an algorithm as well as show how the parser processes sentences. Section 4 discusses the effectiveness of this approach and further issues. Section 5 concludes the discussion.

2 Dynamic Syntax and Problems of Parsing

2.1 Dynamic Syntax and Formalism

Before getting into the discussion of the issues I address, a brief illustration of the formalism of Dynamic Syntax (hereafter DS) is needed. The DS formalism allows the parser to process a sentence in a left-to-right fashion: it consumes the words from the onset, building up semantic representation as the scan of the string proceeds. The grammar formalism is called goal-directed in the sense that in each node there exists $requirement(s)$ prefixed by "?" and every requirement needs to be satisfied or canceled until the parsing has been finished. The initial state of parsing consists of the single node $\{Tn(a),\ ?Ty(t),\ \Diamond\}$ (quoted from Kempson et al. 2001: p.57), where Tn is the tree node identifier and $?Ty(t)$ means that this node will be associated with a type t formula; the parser is going to parse the sentence. The node includes the pointer $\Diamond$ indicating that the pointed node is highlighted or active so that lexical rules and transition rules are applied to this node. Summing up, processing is defined as (1). A parsing state, which is a (partial) tree structure, shifts to the next one and grows larger and larger from the initial state $\mathcal{T}_0$ to the final state $\mathcal{T}_n$ through the application of lexical and transition rules.

(1) $\mathcal{T}_0 \rightarrow$ rule application $\rightarrow \mathcal{T}_1 \rightarrow \ldots\ldots \rightarrow$ rule application $\rightarrow \mathcal{T}_n$

In the tree structure, nodes are represented as sets, and the relation between nodes are represented by the tree node predicate Tn and the node modalities. The lexical items themselves are defined as rules and take the form of **IF** $Action_1$ **THEN** $Action_2$ **ELSE** $Action_3$ which updates the current partial tree structure. In Cann et al. (2005) the main transition rules are as follows: Local ∗Adjunction, Generalised Adjunction, ∗Adjunction, Introduction, Prediction, Elimination, Anticipation, Completion, Thinning, Link Adjunction, Link Evaluation and Merge.

The other characteristic worth noting is *syntactic* underspecification. When the parser processes the string in a left-to-right fashion, especially in head-final languages like Japanese, the parser cannot specify whether the NP at the beginning of a sentence will be fixed to a main or subordinate or relative clause. For example, the initial NP *hon-o* "book-ACC" of (2a) is the object NP of a main clause but in (2b) it is the object of a relative clause in a subordinate clause.

(2) a. Hon o gakusei ga katta.
book ACC student NOM buy-PAST
"The student bought the book."

b. Hon o katta gakusei ga koronda to Taro ga itta.
book ACC buy-PAST student NOM fall down-PAST COMP Taro NOM say-PAST
"Taro said that the student who bought the book fell down."

DS has a mechanism which enables the parser to keep NPs unfixed in the partial tree structure. This paper adopts three transition rules; Local ∗Adjunction, Generalised Adjunction and ∗Adjunction (Cann et al. 2005). Local ∗Adjunction provides an unfixed position for local scrambling, and Generalised Adjunction is used to introduce a subordinate clause because it provides an unfixed node associated with a type t formula and the unfixed node dominates all the words which belong to the subordinate clause. ∗Adjunction rule provides a tree structure with an unfixed node for long distance scrambling. The snapshot of the partial tree consuming *hon* of (2a) is shown in Figure 1 and that of (2b) in Figure 2. The unfixed node introduced by Local ∗Adjunction is connected to the mother node by the dashed line, while the dotted line means that the unfixed node is provided by the Generalised Adjunction rule in each figure. Such unfixed nodes are required to find their fixed positions by the end of the processing with tree-update actions. Readers who wishes to see a more detailed description of the DS formalism, are referred to Kempson et al. (2001) and Cann et al. (2005).

2.2 Problems of Parsing in DS

This subsection will outline the issues addressed in this paper. This paper mainly deals with two problems. The first is the issue of how we should express (partial) tree structures in implementing the DS grammar. To put it another way, the problem is how we should efficiently implement the grammar of head-final languages such as Japanese within the DS framework in Prolog. In Prolog notation, a complete binary tree structure is often represented as `node(s, node(np, [], []), node(vp, [], []))`, which is the equivalent of the usual notation of the tree structure in which S goes to NP and VP. When unfixed nodes,

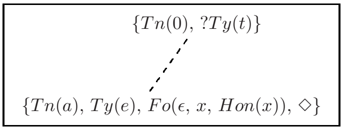


Fig. 1. The parsing tree of *hon* "book" in (2a)

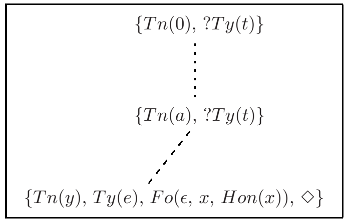


Fig. 2. The parsing tree of *hon* "book" in (2b)

which are introduced by LOCAL ∗ADJUNCTION, GENERALISED ADJUNCTION and ∗ADJUNCTION, come in this notation, then it would not be easy to search a pointed node and merge the unfixed nodes with other fixed positions. Particularly in a head-final language like Japanese, the number of nodes dominated by the unfixed node can be far larger than in English because, for example, the type t node of a sentence sometimes needs to be unfixed until the complementizer *to* has been consumed. If the parser represents the tree structures as sets of nodes, as in DS originally, it would not be easy to parse the string efficiently. This problem has a close relation to how we should represent the processing state.

The second issue this paper addresses is an algorithm for the application of the transition rules and lexical rules. DS is a grammar formalism that allows a parsing state to be updated to a subsequent state, and currently in the DS framework transition rules can apply to any state in an arbitrary way; there is no algorithm which specifies how and when lexical and transition rules are applied. As illustrated in Figures 1 and 2, application of different transition rules leads to different partial tree structures. For example in Figure 2, LOCAL ∗ADJUNCTION is applied after applying GENRALISED ADJUNCTION, while only LOCAL ∗ADJUNCTION is applied in Figure 1. In implementing the grammar, we need an algorithm to derive such partial structures as these two. This paper proposes an algorithm of rule applications for the implementation of the DS parser and provides some improvements for efficiency.

2.3 Previous Studies

Although the generation and parsing model of Purver and Otsuka (2003) and Otsuka and Purver (2003) based on DS formalism can be found in the literature, they mainly deal with English. Purver and Kempson (2004) and Purver, Cann and Kempson (2006) go a step further and propose a context-dependent parsing and generation model to account for the transition from a hearer to a speaker and vice versa with reference to shared utterances and some context-dependent phenomena such as VP ellipsis.

A brief description of their model and implementation is as follows. In their implementation in Prolog, the tree structure is represented as a set of nodes. Figure 3 is their Prolog notation of a partial tree structure, while its DS conterpart is illustrated as Figure 4 (both are cited from Otsuka and Purver 2003: p.98).

```
tree( [node('0', [?ty(t), ?([\/0],ty(e)), ?([\/1],ty(e>t))] ),
       node('00', [ty(e), fo(john), +male] ),
       node('01', [ty(e>t), fo(X^snore(X)), +pres] )] ).
```

Fig. 3. Prolog notation of Otsuka and Purver (2003: p.98)

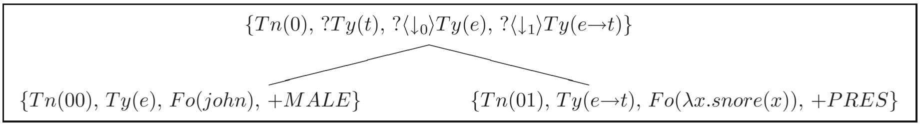


Fig. 4. DS notation of Figure 3 (Otsuka and Purver 2003: p.98)

Although the tree structure is represented as a set of nodes, as Figure 3 shows, the relation between the mother node and its daughter node is represented as the tree node identifier (Tn): the root node has the identifier $Tn(0)$ and the argument daughter has the identifier $Tn(00)$, and the functor $Tn(01)$.

There are two problems I can point out about their model, ones related to the two issues illustrated in subsection 2.2. The first is about Prolog tree notation and a formal definition of the parsing state. According to Purver and Kempson (2004: p.77) a parsing state of their parser is defined as a set of triples $\langle T, W, A \rangle$, where T is a tree, W, words which have already been consumed and A, rules which have already been applied. As described above, any node (and any partial tree) is defined as an element of set T. Therefore, once we attempt to search a certain node, e.g., the pointed node, among the partial tree, it turns out to be inefficient especially as the tree gets bigger and bigger. Moreover, as described in the previous subsection, in Japanese the parser needs to cope with many unfixed nodes other than fixed nodes; this makes the treatment of a tree as a set of nodes inefficient for processing head-final languages. In the subsequent section of this paper a slightly different, structured parsing state approach to DS tree structure will be proposed which enables easier access to the nodes.

The second problem is related to the algorithm and the application of transition rules. Otsuka and Purver (2003) and Purver and Otsuka (2003) try to improve the efficiency of parsing by assuming `always_rules` which apply forcibly and `possible_rules` which don't necessarily apply. In the subsequent section of this paper I propose a different transition rule application system for Japanese and an algorithm to process fragments of Japanese sentences.

3 Incremental Parsing for Japanese

3.1 Definition of the Parser

This subsection describes development of the parser proposed in this paper, an issue closely related to the representation of the tree structures we have already discussed. Unlike the model of Purver and Kempson (2004), the model proposed here has the following parsing state structure: it is a set of triples $\langle W, S, P\rangle$ where W is a string to be consumed and P a node address of the pointed node at a current stage including the feature `fixed, local, gen` and `link`; `fixed` means that the pointer is in the fixed tree structure, `local` that the pointer is in a node introduced by Local $*$Adjunction, `gen` that the pointer is in a node introduced by Generalised Adjunction, and `link` that the pointer is in the linked structure. S is a multitier parsing state which consists of a set of doubles $\langle R, T\rangle$ where R is a set of transition and lexical rules which have been used to establish the current tree. T consists of a triple $\langle F, G, L\rangle$ where F is a fixed tree structure, G a tree structure whose root node is introduced by Generalised Adjunction, and L a linked tree structure. F has the binary tree structure as shown in section 2, such as `node(Mother, node(Argument, [], []), node(Functor, [], []))`, but each node, e.g., `Mother`, has a single place for a node introduced by Local $*$Adjunction, while Otsuka and Purver (2003) treat trees as a set of nodes. The pictorial image of the parsing state is illustrated in Figure 5. The initial parsing state is $\langle W_0, S_0,$ `pn(fixed, [root])`$\rangle$; S_0 consists of a double $\langle \phi, A\rangle$ where A is the initial parsing state, which contains only the requirement $?Ty(t)$ and the pointer $\diamond$. The final parsing state is $\langle \phi, S_n,$ `pn(fixed, [root])`$\rangle$. In S_n the generalised tree structure and the link structure are empty, and the semantic representation is established at the root node of the fixed tree structure. The pointer needs to go back to the root node of the fixed tree structure for the parsing to be successful as indicated by `pn(fixed, [root])`.

The separate treatment of fixed, generalised, and link structure and the pointed node address indicator P enable us to search the pointed node and manipulate tree-update actions efficiently: this approach helps the parser process SVO languages like Japanese, particularly when a considerable amount of nodes is dominated by the unfixed mother node.

3.2 Transition Rules and Algorithm

This subsection deals with the second issue I addressed: the application of the transition rules and its algorithm. The parser currently has the following transition rules: Local $*$Adjunction, Generalised Adjunction, $*$Adjunction, Introduction, Prediction, Elimination, Anticipation, Completion,

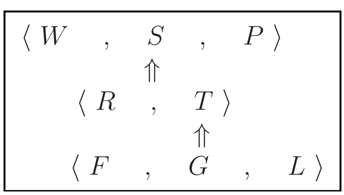


Fig. 5. Structure of parsing state

```
tree_expansion_rules = {generalised_adjunction/local_*adjunction/
                        link_adjunction}
node_update_rules = {merge, link_evaluation, elimination,
                     thinning, completion}
```

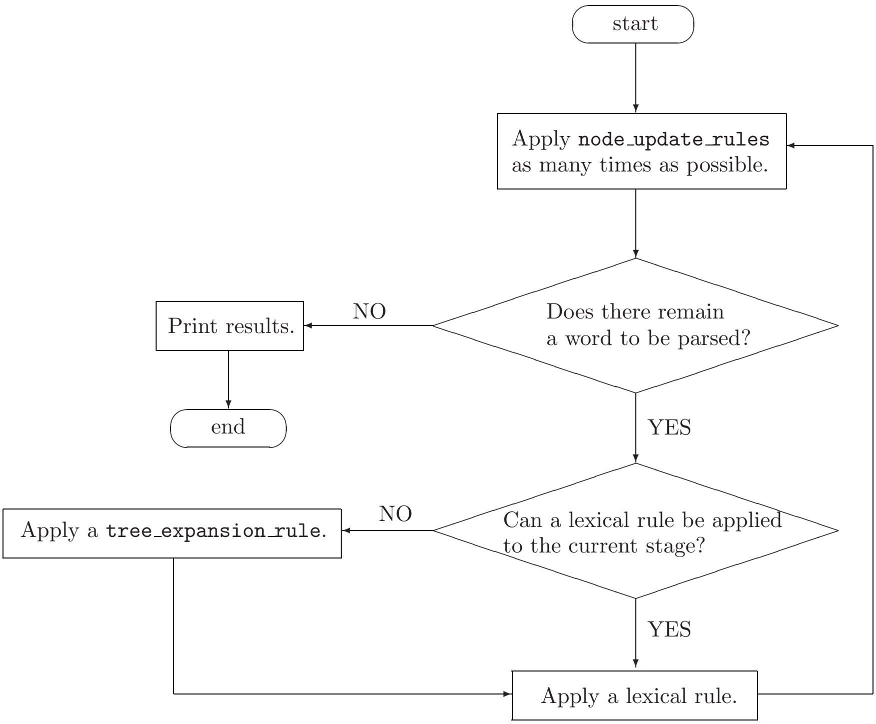


Fig. 6. Algorithm of processing

Thinning, Link Adjunction, Link Evaluation, and Merge. This paper assumes that among those rules Introduction, Prediction, and Anticipation are not used for processing Japanese sentences.

As mentioned earlier, there is no algorithm which specifies when and how the transition rules apply: they can be applied to any state in an arbitrary way in the current DS formalism. This subsection presents an algorithm, and the subsequent subsection shows that the algorithm is sufficient to parse simple sentences, relative clause constructions, and embedded clauses. Unlike Otsuka and Purver (2003) and Purver and Otsuka (2003), the approach proposed here divides the transition rules into two groups: one is those rules which expand partial tree structure, `tree_expansion_rules`, and the other is those which do not expand the tree but update the node information, `node_update_rules`. The former group, `tree_expansion_rules`, includes Local *Adjunction, Generalised

```
ds_parse([], [[R, L1]|Result], [[R2, L2], [R, L1]|Result]) :-
     apply_tree_expansion_rules([[[]], L1], [], [R2, L2]),
     satisfy(L2), %Checks the tree has no requirement
     pretty_print([[R2, L2], [R, L1]|Result]), nl.

ds_parse([H|T], [[X, T1]|T2], Result) :-
     apply_node_update_rules([[[]], T1], [], [R, Mid]),
     (lexical_rule(H, [R, Mid], [R1, Mid3]);
      (tree_expansion_rule([R, Mid], [R2, Mid2]),
       lexical_rule(H, [R2, Mid2], [R1, Mid3])
      )
     ),
     ds_parse(T, [[R1, Mid3], [X, T1]|T2], Result).
```

Fig. 7. Prolog code of the algorithm

Adjunction, ∗Adjunction, and Link Adjunction. Other transition rules, Merge, Link Evaluation, Elimination, Thinning, and Completion belong to `node_update_rules`. The shared characteristic among the former `tree_expansion_rules` group is that they require that the pointed node have the requirement of type t.

The algorithm proposed here is diagramed as Figure 6, and the algorithm implemented in the Prolog code is illustrated in Figure 7. The upper four lines of Figure 7 are the base rule, and the bottom eight lines are the recursive rule. The first argument of the `ds_parse` predicate is the string to be consumed: when there remains no word to be consumed, the processing ends. The second argument is the list consisting of the pair of rules applied in that stage and the tree structure. The algorithm defined in Figure 6 considerably improves the efficiency of parsing because the `tree_expansion_rules` are not applied if the pointed node does not have the requirement of type t. The important characteristic worth noting is the application order of the `node_update_rules`. The parser tries to apply rules in the following order; Merge, Link_Evaluation, Elimination, Thinning, and Completion, though there might be the possibility that some of them cannot apply. When one of the rules, e.g., Merge, cannot apply, the tree structure is passed to the next rule, in this case Link_Evaluation, without any modification to the tree structure. This ensures that there is always an output tree structure at the end of each rule application in `node_update_rules`. This application order especially works in the situation where the parser brings the type and semantic information up to the mother node with the β-reduction after consuming the verb: Elimination generates the type and formula of the node and Thinning deletes the requirement, and Completion brings the information up to the mother node.

3.3 How the Parser Works

This subsection illustrates how the parser works. The parser is written in SWI-Prolog on Linux. The current grammar is small but the source code is composed

```
?- parse([john, ga, hon, o, katta], X).

Step 0
Nothing applied.
Pointer: pn(fixed, [root])
Root: [tn([0]), an([?ty(t)]), []]
Gen_adj: []
Linked: link([[], [], []])

Step 1
local_adj, john applied.
Pointer: pn(fixed, [root, local])
Root: [tn([0]), an([?ty(t)]), [loc([fo(john), ty(e), ?ty(e)])]]
Gen_adj: []
Linked: link([[], [], []])

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Step 5
katta applied.
Pointer: pn(fixed, [root, 1, 1])
Root: [tn([0]), an([?ty(t), \/[1, ty((e->t))], \/[0, fo(john)], \/[0, ty(e)]]), []]
          [tn([0, 0]), an([fo(john), ty(e)]), []]
          [tn([0, 1]), an([\/[0, ty(e)], \/[0, fo(book)], ty((e->t))]), []]
                    [tn([0, 1, 0]), an([fo(book), ty(e)]), []]
                    [tn([0, 1, 1]), an([fo(lambda(book, lambda(john, buy(john, book)))),
                     ty((e->e->t))]), []]
Gen_adj: []
Linked: link([[], [], []])

Step 6
completion, elimination, completion, elimination, thinning applied.
Pointer: pn(fixed, [root])
Root: [tn([0]), an([ty(t), fo(buy(john, book)), \/[1, fo(lambda(john, buy(john, book)))],
       \/[1, ty((e->t))], \/[0, fo(john)], \/[0, ty(e)]]), []]
          [tn([0, 0]), an([fo(john), ty(e)]), []]
          [tn([0, 1]), an([fo(lambda(john, buy(john, book))),
           \/[1, fo(lambda(book, lambda(john, buy(john, book))))],
           \/[1, ty((e->e->t))], \/[0, ty(e)], \/[0, fo(book)], ty((e->t))]), []]
                    [tn([0, 1, 0]), an([fo(book), ty(e)]), []]
                    [tn([0, 1, 1]), an([fo(lambda(book, lambda(john, buy(john, book)))),
                     ty((e->e->t))]), []]
Gen_adj: []
Linked: link([[], [], []])

Semantic Representation: fo(buy(john, book))
```

Fig. 8. Snapshot of the output

of about 1,200 lines except the lexicon, and the total size is 41.1 KB. The parser can process simple sentences such as (3a, b) as well as also the relative clause constructions (3c) and complex sentences (3d).

(3) a. John ga hon o katta.
John NOM book ACC buy-PAST
"John bought the book."

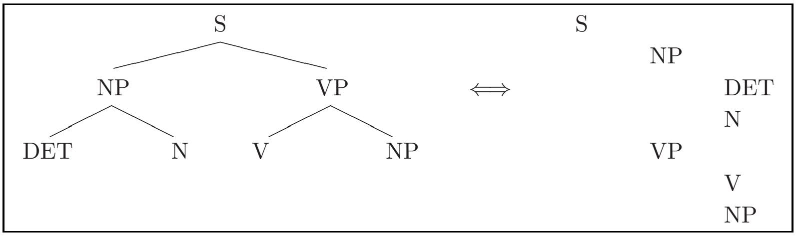


Fig. 9. Tree notation and command line notation

b. Hon o John ga yonda.
book sc acc John NOM read-PAST
"John read the book."

c. Hon o katta gakusei ga hashitta.
book ACC buy-PAST student NOM run-PAST
"The student who bought the book ran."

d. John ga hon o katta to Tom ga itta.
John NOM book ACC buy-PAST COMP Tom NOM say-PAST
"Tom said that John bought the book."

The parser returns all steps of the processing, N+1 steps in total, from step 0 (the initial state) to step N+1, for the N words input. Figure 8 is a simplified snapshot of the processing of the example sentence (3a). The parsing result is printed from step 0 (the initial state) to step 6 (final state), although steps 2, 3 and 4 are abbreviated. The semantic formula is printed at the bottom line as `fo(buy(john, book))`, which means *John bought the book* (ignoring tense). Some other notational conventions are given in Figure 10. As you can see in the figure the tree structure grows from the onset to the final stage monotonically. In each step, the tree structure is represented as lines whose dominance relationships are shown with the tab spacing. Figure 9 shows us the corresponding, standard DS representation.

The parser can return more than one semantic representation for ambiguous sentences. As Kempson et al. (2001: p.70) discusses, when an appropriate context is given, the initial NP in (4) can be the subject of the main clause or that of the subordinate clause or both of them.

(4) John ga hon o katta to itta.
John NOM book ACC buy-PAST COMP say- PAST
1: "John said that (Tom or Mary or someone else) bought the book."
2: "(Tom or Mary or someone else) said that John bought the book."
3: "John$_i$ said that he$_i$ bought the book."

The parser returns two semantic representations by Prolog's backtracking after processing (4); one is `fo(say(john, buy(meta_v, book)))`, and the other `fo(say(meta_v, buy(john, book)))`, where `meta_v` is a meta-variable which will be substituted for *John* or other entities in the context by pragmatic actions.

DS notation	Parser notation
$\langle\downarrow_0\rangle Ty(e)$	`\/[0, ty(e)]`
$\langle\uparrow_1\rangle Ty(t)$	`/\[1, ty(t)]`
$\langle\downarrow_0\rangle Fo(John)$	`\/[0, fo(john)]`
$Tn(0110)$	`tn([0, 1, 1, 0])`

Fig. 10. Notational convention of features in DS and parser

Let me explain briefly how the algorithm in Figure 6 processes the sentence using one of the possible readings in (4), "John said that (Tom or Mary or someone else) bought the book." As Figure 6 shows, the first step the parser attempts to take is to apply `node_update_rules` as many times as possible. This step results in a vacuos application; in this case, it returns tree structures without any modification compared to the previous stage, since none of the `node_update_rules` are applied. Then in the next step the lexical specification of the NP *John* fails to apply to the current tree structure where the pointer is situated in the type t node, since *John* requires that the type e node be the pointed node. As the next step `tree_expansion_rules`, in this case Generalised Adjunction or Local ∗Adjunction, apply because the pointed node is of type t (when Generalised Adjunction is set to work, the semantic representation would be `fo(say(meta_v, buy(john, book)))`, while Local ∗Adjunction leads to the other reading `fo(say(john, buy(meta_v, book))))`. Let us assume that the Local ∗Adjunction applies to the current stage. After consuming the NP *John*, `tree_expansion_rules` takes on a role again, because the pointer has gone back to the root, type t node, and Generalised Adjunction gives rise to another type t node which would be the root node of the embedded clause in a later stage. After scanning the verb within the embedded clause *katta* "buy-PAST", the repetitive application of `node_update_rules` generates the meaning of the subordinate clause `fo(buy(meta_v, book))` in the type t node introduced by Generalised Adjunction, bringing the pointer back to the type t node for the subordinate clause. Then the complementizer *to* integrates the unfixed subordinate tree structure to the fixed, topmost type t node. After returning the pointer to the type t node, as shown in Figure 6, not `tree_expansion_rules` but the lexical rule *itta* "say-PAST" applies to the current node because `tree_expansion_rules` are triggered only if a lexical rule cannot be applied.

As this algorithm shows, the application of these transition rules is restricted by the types of the pointed node. In Japanese, rules which expand tree structures are able to apply to the tree only if the pointed node is of type t.

4 Discussion

This section is devoted to discussing my approach and possibilities for future research. This paper presented a partitioned parsing state for Japanese based on the DS framework. This approach to locally unfixed nodes and non-local unfixed nodes, such as a type t node which establishes a new embedded clause in the course of incremental processing, realizes more efficient actions for merging unfixed nodes with fixed nodes as well as easier access to the pointed node.

This paper also presented the algorithm of application of the transition rules and lexical rules. I speculate that the algorithm is basically applicable to other head-final languages like Korean. Purver and Otsuka (2003) proposes the generation and parsing model for English on the basis of DS, but as far as I know this paper is the first implementation for Japanese within the DS framework.

The parser (and the DS framework itself) has some problems to be overcome in future research. The most important one to be addressed in the future is the quantitive evaluation of the parser. Although the grammar and lexicon are still small, the parser will be evaluated using corpus, and some stochastic treatment should be added to improve its efficiency.

In the DS formalism the core engine of the parser does not provide any pragmatics-related action, and there have been few in-depth studies presented on this issue. Furthermore, the idea and treatment of *discourse* or *context* is not very clear in DS. Therefore, the other important issue would be the formalization and implementation of *pragmatic actions* giving rise to the merging of meta-variables with appropriate entities in the context.

5 Conclusion

This paper delineated a parser for fragments of Japanese sentences on the basis of Dynamic Syntax. The parser proposed separate treatments of nodes in fixed (binary) trees and unfixed nodes for Japanese, a proposal essential for head-final languages which often may have multiple unfixed nodes at some parsing stages. This allows easier access to the pointed node. This paper also proposed the algorithm of an application of transition rules which makes the parsing more efficient. The transition rules have been classified into `node_update_rules` and `tree_expansion_rules`. The application of these rules is restricted by the position of the pointer in the tree structure. The parser, implemented in Prolog, is able to process simple sentences as well as relative clause constructions and complex sentences.

Acknowledgements

I would like to thank Kei Yoshimoto for valuable suggestions and comments on an earlier draft. I would also like to thank anonymous reviewers for their comments.

References

Cann, R., Kempson, R., Marten, L.: The Dynamics of Language. Elsevier, Amsterdam (2005)

Kempson, R., Meyer-Viol, W., Gabbay, D.: Dynamic Sytnax: The Flow of Language Understanding. Blackwell Publishers, Oxford (2001)

Lombardo, V., Mazzei, A., Sturt, P.: Competence and Performance Grammar in Incremental Processing. In: Keller, F., et al. (eds.) Proceedings of the ACL Workshop Incremental Parsing: Bringing Engineering and Cognition Together, pp. 1–8 (2004)

Otsuka, M., Purver., M.: Incremental Generation by Incremental Parsing. In: Proceedings of the 6th Annual CLUK Research Colloquium, pp. 93–100 (2003)

Purver, M., Cann, R., Kempson, R.: Grammars as Parsers: Meeting the Dialogue Challenge. Research on Language and Computation 4, 2–3, 259–288 (2006)

Purver, M., Kempson., R.: Incremental Parsing, or Incremental Grammar? In: Proceedings of the ACL Workshop on Incremental Parsing., pp. 74–81 (2004)

Purver, M., Otsuka., M.: Incremental Generation by Incremental Parsing: Tactical Generation in Dynamic Syntax. In: Proceedings of the 9th European Workshop on Natural Language Generation, pp. 79–86 (2003)

Breaking Quotations⋆

Emar Maier

University of Amsterdam
emar.maier@gmail.com
www.ru.nl/ncs/~emar/

Abstract. Quotation exhibits characteristics of both use and mention. I argue against the recently popular pragmatic reductions of quotation to mere language use (Recanati [1]), and in favor of a truly hybrid account synthesizing and extending Potts ([2]) and Geurts & Maier ([3]), using a mention logic and a dynamic semantics with presupposition to establish a context-driven meaning shift. The current paper explores a "quote-breaking" extension to solve the problems posed by non-constituent quotation, and anaphora, ellipsis and quantifier raising across quotation marks.

1 Varieties of Quotation

Natural language provides a number of ways to report what someone else has said. Let me illustrate this by categorizing some reports of a real-life Bushism.[1]

Direct discourse provides the most faithful representation of the original speech act, even at the expense of correct grammar and morphology:

(1) [Bush:] "I've, I've got a eckullectic reading list."[2]

If we don't feel the need to convey Bush's exact wording, but pay attention only to what is expressed, we opt for indirect discourse:

(2) Bush said that he has an eclectic reading list.

A third way to report an utterance is *mixed quotation*, an indirect report in which only a particular phrase is quoted verbatim:

(3) The president said he has an "ecelectic" reading list.[3]

Semantically, direct discourse is commonly treated as a kind of *mention*: the reporter employs quotation marking to *mention* a sentence, i.e. to refer to Bush's utterance rather than *use* the sentence to express a proposition herself. In section 2,

⋆ I thank the audiences of LENLS 2007 (Miyazaki), LeGO Colloquium (Amsterdam), and the Frankfurt GK colloquium for valuable comments and suggestions on parts of this work. Further thanks go to the organizers and referees of LENLS 2007.

[1] www.youtube.com/watch?v=Jd0lIduBBoU

[2] www.thehookny.com/issue/47.pdf

[3] www.thecarpetbaggerreport.com/archives/8339.html

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 187–200, 2008.

I make this more precise and argue against the reduction of direct discourse to pure mention. Indirect discourse on the other hand is usually analyzed on a par with attitude reports (*believes that*, *hopes that*, etc.) as an intensional semantic operator, $\mathsf{SAY}_{\mathtt{x}}\varphi$, e.g. as defined by Kaplan ([4]). Mixed quotation finally is the principal subject of sections 3 and 4.

2 Direct Discourse and the Logic of Mentioning

Tests for transparency vs. opacity suggest a fundamental distinction between direct and indirect reports: (i) quoted errors do not affect the overall acceptability of the report, as seen in (1); (ii) quoted indexicals likewise retain their original form (while in indirect speech they must be adapted to the report context, cf. (2)'s embedded subject *he* vs. (1)'s *I*); and (iii) quotation blocks quantifying in and *wh*-extraction:

(4) a. What kind of reading list did Bush say he has? [from (2)]
b. *What kind of reading list did Bush say: "I've, I've got"? [from (1)]

It is suggested that these observations are explained by an analysis of direct speech as mention (and indirect speech as use, i.e. as intensional operator), but what exactly is the logic of mentioning?

I follow the standard Quine/Tarski analysis where quotation marks turn any linguistic entity into a name for that entity. This requires the addition of those linguistic items to the domain of semantic interpretation. Following Potts ([2]) I introduce a new type u (in addition to the usual e and t) and a corresponding domain D_u for these linguistic entities, and say that quotation marks turn a linguistic entity of any type into one of type u that refers to that linguistic object:

(5) a. $D_u = \{\text{eclectic}, \text{ekullectic}, \text{I've got an ekullectic reading list}, \ldots\}$
b. If $\sigma \in D_u$ then $\ulcorner\sigma\urcorner$ is an expression of type u and $[\![\ulcorner\sigma\urcorner]\!] = \sigma$

In this logic, direct discourse may be represented with the relation $\mathtt{say}$:

(6) $\mathtt{say}$ is an expression of type $\langle u, \langle e, t\rangle\rangle$ and if σ is a sentence $\in D_u$ then $[\![\mathtt{say}(\mathtt{x}, \ulcorner\sigma\urcorner)]\!] = 1$ iff $[\![\mathtt{x}]\!]$ utters σ

Indeed, this intuitive analysis of direct quotations as terms (of type u), *naming* the enclosed utterances, would account neatly for the opacity effects observed in (i)-(iii).

A closer look at the actual data however casts some doubt on the clean separation of quotation as mention and reporting as use. Ad (i): direct speech occasionally adapts various features such as the language (cf. Yomiuri Shimbun quoting Bush), but also some linguistic errors (cf. the official transcript of the reading list interview)[4] of the original utterance to the report context. Ad (ii):

[4] www.msnbc.msn.com/id/14576012

there has been a lot of recent interest in the discovery that some languages allow indexicals to shift, i.e. to keep them in their original form, as if mentioned:

(7) John$_i$ said that I$_i$ am a hero [Amharic (Schlenker 5)]

The standard account of this involves a modification to the syntax and semantics of indirect speech (allowing "monsters"), but perhaps an easier account can be had if we assume the feature retention really is due to some interfering mentioning effect?[5] Test (iii) seems to hold up pretty well, but there are other tests for the 'semantic inertia' of mentioning. For instance, (iv) it should block anaphoric and elliptical dependencies, which as Partee (6) already noted, is clearly false for direct discourse:[6]

(8) "My girlfriend bought me this tie," said John, but I don't think she did

Finally, (v) from a direct quotation we often infer the corresponding indirect version where the words are used, e.g. (1)$\models$(2), or, more generally: x says $\ulcorner\sigma\urcorner \models x$ says that σ', where σ' is a variant of σ with all indexicals and errors adapted to the report context.

None of these latter facts are explained in the simple *mention* analysis. I conclude that contrary to traditional wisdom (cf. Davidson 8, Kaplan 4) even direct discourse involves both use and mention. Before giving an adequate formal analysis we turn to mixed quotation where this is even clearer.

3 Mixed Quotation: Use and Mention

Davidson was the first to recognize the dual nature of mixed quotations:

> (9) Quine said that quotation "has a certain anomalous feature"
> Are the quoted words used or mentioned? Obviously mentioned since the words are Quine's own, and I want to mark the fact. But equally obvious is the fact that the words are used; if they were not, what follows the word 'quotation' would be a singular term, and this cannot be if I have produced a grammatical sentence. [(Davidson 8)]

In the system outlined in section 2, we distinguish mentioned expressions (type u) from ordinary singular terms (type e), but the argument still holds. So, we need to combine both aspects: *"ecelectic"* in (3) refers to Bush's words, but still plays the role of a predicate modifier in the report.

To be absolutely sure, let's go through the other transparancy tests. In mixed quotation (i) errors are allowed (cf. (3), (10a)), (ii) indexicals retained (cf. (10a)), and (iii) extraction blocked (cf. (10b)):

(10) a. Bush said the enemy "misunderestimates me"
b. *What kind of feature did Quine say quotation "has" [from (9)]

[5] I leave this to another occasion.

[6] Watters & Gundel (7) have explored the effect of quotation marks on anaphora resolution in a magazine corpus. They found none, thus (unwittingly) corroborating Partee's findings.

Tests (i)-(iii) thus indicate mention, but we have evidence for use also: (iv) anaphora and ellipsis are possible, (v) use-inference very strong ((3) $\models$ (2)), and (vi), the Davidsonian argument, the quoted consituent must have the semantic type that fits in the report. Since (vi) is often taken to definitively preclude a simple mention analysis along the lines of (5), some have gone the other way and tried to reduce the semantics of mixed quotation to pure reportative language use (e.g. Stainton [9], Recanati [1]). These may be grouped together as 'pragmatic' analyses of quotation for they take it that the sole contribution of quotation marks is to signal that there is something special about this part of the sentence, and leave it to pragmatics to decide what that something is. As far as semantics proper is concerned, the quotation marks may simply be omitted. However, the data we found in checking (i)-(iii) now become highly problematic, for how can we interpret (10a)'s quoted *me* as referring to anything other than the reporter if we were to neglect quotation marks semantically? And how to interpret quoted non-words like *ecelectic* or *misunderestimated* at all when the sentence without quotation marking[7] is simply ungrammatical?

3.1 Potts

Potts ([2]) proposes an account with both a mention and a use component separated on two dimensions of meaning. The first consists of the proposition that so-and-so uttered the phrase in quotation marks, the second is computed as the ordinary semantic value of the quoted expression:

(11) $$\left\langle \begin{array}{c} \mathtt{say}(\mathtt{x}, \ulcorner\text{ecelectic}\urcorner) \\ \mathtt{eclectic} \end{array} \right\rangle$$

In the full, compositional system the mention component always projects up, while the material in the second dimension remains *in situ*, or rather, is employed in further compositional computations. Finally, at clause level the two dimensions are collapsed and interpreted as conjoined:

(12) $$\mathtt{say}(\mathtt{x}, \ulcorner\text{ecelectic}\urcorner) \wedge \mathsf{SAY}_{\mathtt{x}} \exists \mathtt{y}[(\mathtt{eclectic}(\mathtt{book\_list}))(\mathtt{y}) \wedge \mathtt{have}(\mathtt{x}, \mathtt{y})]$$

In words: Bush uttered the phrase "ecelectic" and he said that he has an eclectic book list.

Though arguably the first and most precise formal semantic account of quotation, there are a number of objections to Potts' analysis. First is a small methodological worry. Potts' goal is to give a (directly) compositional treatment of the semantics of quotation. The result of processing a mixed quoted example however is a logical form (henceforth: lf) where the mention component ($\mathtt{say}(\mathtt{x}, \ulcorner\text{ecelectic}\urcorner)$) has moved from the embedded position where it was generated by the quotation marks, all the way up to the main clause level. It seems that the two-dimensional machinery was devised solely for the purpose of incorporating this kind of movement, which makes this strategy slightly ad hoc.

[7] Note that quotation marking need not involve quotation marks or fingerdance quotes, other forms of quotation marking include capitals and italics, or, in spoken language, a rather distinctive intonational contour.

My second, more substantial objection concerns the quotation data associated with tests (i), errors, and (ii), indexicals: How exactly do we determine the use component of such a quoted item? In the formalism sketched above, the use dimension should contain a logical representation of the part mentioned in the first component. To facilitate further discussion, let's introduce a special (meta-language) notation for this:

(13) If $\alpha \in D_u$, then $|\alpha|$ denotes the logical form corresponding to α.

Note, by way of a methodological interlude, that Potts and many others simply use the semantic evaluation brackets to denote the logical form of an expression. This is likely due to the Montagovian idea that logical forms can be dispensed with in favor of a system where syntactically enriched fragments of natural language are interpreted in a model directly. I on the other hand assume a slightly different, DRT inspired setup in which we distinguish the following three levels: (i) unanalyzed surface forms, (ii) formulas (representing truth conditions) in some logical language such as first-order predicate logic or DRT, and (iii) model-theoretic interpretations.

For future reference we can now state the rule behind (11) more precisely:

(14) If $\alpha \in D_u$ then the 2D contributions of "α" and α itself are as follows:

$$\text{“}\alpha\text{”} \rightsquigarrow \left\langle \begin{array}{c} \mathtt{say}(\mathtt{x}, \ulcorner\alpha\urcorner) \\ |\alpha| \end{array} \right\rangle \;;\; \alpha \rightsquigarrow \left\langle \begin{array}{c} \\ |\alpha| \end{array} \right\rangle$$

We finish our reconstruction of Potts' system by adding the usual rules of semantic composition, operating (non-vacuously) only on the second dimension, plus an extra rule to collapse the two levels into a proper, conjoined logical form as soon as the second dimension reaches type t.

Back to the evaluation of this system. Since we're adding the lf contribution of the quoted part in the second dimension, we run into trouble with indexicals if we then evaluate the end result lf relative to the actual context (c):

(15) a. Bush said he would "recommit my heart to Jesus Christ"[8]

b. $$\left\langle \begin{array}{c} \mathtt{say}(\mathtt{x}, \ulcorner\text{recommit my heart to Jesus Christ}\urcorner) \\ \lambda \mathtt{x}[\mathtt{recommit}(\mathtt{x}, \imath \mathtt{y}[\mathtt{heart}(\mathtt{y}) \wedge \mathtt{poss}(\mathtt{i}, \mathtt{y})], \mathtt{jesus})] \end{array} \right\rangle$$

$$\rightsquigarrow \quad \mathtt{bush}(\lambda \mathtt{x}[\mathtt{say}(\mathtt{x}, \ulcorner\text{recommit my heart to Jesus Christ}\urcorner) \wedge \mathsf{SAY}_{\mathtt{x}}[\mathtt{recommit}(\mathtt{x}, \imath \mathtt{y}[\mathtt{heart}(\mathtt{y}) \wedge \mathtt{poss}(\mathtt{i}, \mathtt{y})], \mathtt{jesus})]])$$

c. $[\![(15b)]\!]^c$ = the proposition that Bush uttered "recommit my heart to Jesus Christ" and said that he recommitted Emar's (!) heart to Jesus.

We clearly need to evaluate the quoted material's lf contribution relative to the original reported utterance context instead of the actual c, but how? Now this situation is actually very reminiscent of what we see with Schlenker's (2003) examples of shifted indexicals in Amharic indirect reports, cf. (7): we have an

[8] A slight simplification of: *Bush writes of his decision to "recommit my heart to Jesus Christ."* (www.utilitarian.net/singer/by/200303--.htm)

embedded normal use occurrence of *I* (lf: $\texttt{i}$) that is interpreted as referring to a reported speaker. Schlenker's analysis involves a reinterpretation of the indirect speech report operator as a quantifier over contexts. Now, simply plug in Schlenker's monstrous lfs (where SAY denotes a quantifier over context variables ($\texttt{c}'$) and every predicate gets an added parameter for such evaluation contexts) into our second dimension. Then, add a stipulation that all quoted material gets evaluated relative to the shifted context $\texttt{c}'$, and we get quoted indexicals right:

(16) a. $\texttt{bush}(\lambda \texttt{x}[\texttt{say}(\texttt{x}, \ulcorner\text{recommit my heart to Jesus Christ}\urcorner, \texttt{c}^*) \wedge \textsc{say}_\texttt{x}\texttt{c}'[\texttt{recommit}(\texttt{x}, \imath\texttt{y}[\texttt{heart}(\texttt{y}, \texttt{c}') \wedge \texttt{poss}(\texttt{i}(\texttt{c}'), \texttt{y}, \texttt{c}')], \texttt{jesus}, \texttt{c}')]])$

b. $[\![\text{(16a)}]\!]^{\texttt{c}^* \mapsto c}$ = the proposition that Bush uttered "recommited my heart to Jesus Christ" and, in all contexts c' compatible with what Bush said, Bush recommitted c''s agent's heart (≈ Bush' heart) to Jesus

Note that this solution requires context quantification in the object language (=monsters), a controversial, yet independently motivated mechanism. The current formulation differs from Potts' own, though I think it captures the essence (although Potts doesn't talk about monsters, it's evident that something at least as strong is needed.)

The problem with quoted errors is more serious:

(17) Bush said Saddam "misunderestimated" him

To evaluate the complement we have to compute the 2D contribution of a non-word, which in turn requires that we provide a logical form for it:

(18) $$\left\langle \begin{array}{c} \texttt{say}(\texttt{x}, \ulcorner\text{misunderestimated}\urcorner) \\ |\text{misunderestimated}| \end{array} \right\rangle$$

But what is |misunderestimated|? Apart from the past tense morpheme it does not have internal structure, so we must rely on the lexicon to provide us with an lf. But the whole point of the example was to emphasize Bush use of an incorrect word, so can we really expect to find it in a lexicon? Potts would have to say yes, so in his lexicon we find non-words and mispronounced words like *eck-a-lectic* as separate entries, each paired with a meaning, i.e.: ⟨eck-a-lectic, $\texttt{eclectic}$⟩, ⟨eclectic, $\texttt{eclectic}$⟩, ⟨misunderestimated, $\texttt{underestimated}$⟩, etc. Apart from the size of such a lexicon, there is a more fundamental objection. Take Bush' *misunderestimate*. In this context we assume he meant the enemy *underestimated* him, but on another occasion he or someone else could use the same word to mean *misunderstood*, or something still different. Should we complicate our already rich lexicon even further by putting in everything anybody can mean at some occasion or other (with utterance context parameters added to each item)? In section 3.2 I introduce a different account which regulates who meant what with a certain phrase pragmatically, rather than lexically.

Another problem for Potts' account comes from the fact that the two dimensions are completely separate. Even if we could make sense of a pragmatic, intention dependent (idio)lexicon, the end-result for (17) would be that Bush uttered "misunderestimated" and said that Saddam underestimated him, but there is nothing that says that Bush used that utterance of "misunderestimated" to mean *underestimated*. This apears to be the source of the 'binding problems' addressed by Geurts & Maier ([3]) [=:G&M], to which I refer the interested reader.

3.2 Geurts and Maier

G&M propose a "1D" approach which bears a noteworthy resemblance to Potts' but takes some of the load off the lexicon, shifting it to pragmatics, while integrating the two levels of meaning, thus addressing the objections raised in the previous section. Their DRT formalization can be seen as a minimal variant of Potts with the use component left underspecified and the mention component put into a presupposition. Semantic underspecification of the use contribution requires adding a third argument to the say relation:

(19) $[\![\mathtt{say}(\mathtt{x}, \ulcorner\alpha\urcorner, \mathtt{P})]\!] = 1$ iff $[\![\mathtt{x}]\!]$ utters α to express $[\![\mathtt{P}]\!]$

An obvious restriction on x using α to express $[\![\mathtt{P}]\!]$ is that the syntactic category of α matches the semantic type of P. In cases of 'relatively normal use of words', something stronger may be expected hold: $[\![\alpha]\!] = [\![\mathtt{P}]\!]$ (or, in (semi-)object language: $|\alpha|$=P). That would mean that the quoted speaker associated the same meaning with the quoted phrase as we do. The crucial difference with Potts is that we do not *require* this strengthened assumption to hold, it is merely a defeasible, pragmatic inference, licensed in certain contexts. Semantically, we don't need to be able to interpret α at all. Instead of computing *the* lf and semantic value of the quoted expression, the semantics now relies on *what x means with her use of the expression*:[9]

(20) Bush said that he has a book list with ∂[the property he expressed with the phrase "ecelectic"] [∂ = presupposition marker]

This idea is formalized quite naturally in DRT+PA: the dynamic framework of Discourse Representation Theory extended with Van der Sandt's ([11]) theory of 'Presupposition as Anaphora' (PA). In this framework, DRS-formulas represent contexts, and interpretation is modeled as incrementation of *input contexts*, by successively adding new sentence representations (=*preliminary DRSs*, playing the role of dynamic logical forms) and *resolving* their presuppositions. Resolving a presupposition triggered in a preliminary DRS can be done in either of two ways: *binding* (find an appropriate antecedent for the presupposed variable in the extended context DRS, and equate presupposition and antecedent), or, if that

[9] There is an interesting parallel with the speech act based philosophy of language (Grice [10]), where the unit of meaning is not intension or extension (or context change), but the speaker's intention. The current analysis captures the intention-dependence of quoted expressions, and may be extended to other speech act phenomena such as ironic language use. Note also that we use the intention based meaning concept in our object language, while retaining a classical (or rather, dynamic) semantics of that formal language.

fails, *accommodation* (*create* a suitable antecedent by simply adding the presupposed content and variable at a suitable position in the (context) DRS). Note that presuppositions are generated in situ in the construction of the preliminary DRS, but tend to float up to the main, unembedded context in the process of resolution.

So, what does a DRT+PA formalization of (20) look like? That is, what is the preliminary DRS of the mixed quotation example (3) on G&M's theory? The quotation marks introduce a presupposition to the effect that someone used the given quoted words to express something, while the rest of the sentence provides the asserted contribution. Crucially, in DRT+PA the asserted part contains a free variable P of the appropriate type (here: predicate modifier: $(et)et$) filling in as a placeholder for the quoted, presupposed part:[10]

(21) $$\left[\mathtt{x} \middle| \begin{array}{l} \mathtt{bush(x)}, \\ \mathsf{SAY}_{\mathtt{x}} \left[\mathtt{y} \middle| \begin{array}{l} \mathtt{have(x,y)},\ \mathtt{(P(book\_list))(y)} \\ \partial \left[\mathtt{P} \middle| \mathtt{say(x,} \ulcorner \text{ecelectic} \urcorner \mathtt{,P)} \right] \end{array} \right] \end{array} \right]$$

Now, the presupposition, marked with a ∂, needs to be resolved. Let's just assume a minimal context which doesn't add anything not already in (21). In such a meagre context, there is no salient property of the same type as P, let alone one that matches the presupposed content associated with P, so binding fails and we fall back on (global) accommodation:[11]

(22) $$\left[\mathtt{x,P} \middle| \begin{array}{l} \mathtt{bush(x)},\ \mathtt{say(x,} \ulcorner \text{ecelectic} \urcorner \mathtt{,P)}, \\ \mathsf{SAY}_{\mathtt{x}} \left[\mathtt{y} \middle| \mathtt{have(x,y)},\ \mathtt{(P(book\_list))(y)} \right] \end{array} \right]$$

In words: $[\![(22)]\!] = 1$ iff Bush utters the phrase "ecelectic" to express a predicate modifier P and he says that he has a P reading list.

In a more realistically represented context we may be able to *infer* what property it was that Bush intended (or, what comes to the same thing, *bind* the presupposition rather than *accommodate* it). In cases of well-formed English expressions we already said there may be an automatic default implication that P denotes the property we, or the dictionary, associate with the quoted expression. For quotations involving longer, grammatically well-formed fragments α, we might continue our derivation by computing $|\alpha|$ and adding the condition $\mathtt{P} = |\alpha|$. However, this should be seen only as a defeasible strengthening of the basic, semantic interpretation represented in (22).

The important difference with Potts is therefore that it depends on the context (on resolution and pragmatic, defeasible reasoning) what Bush really expressed with the quoted phrase. If, in some discourse context, it were given that when Bush says "ecelectic", he really means "dyslexic", we would bind P to that contextually

[10] Note how the presupposed part corresponds more or less with Potts' first dimension, while the asserted part corresponds with the use dimension, with the crucial difference that the variable P connects the two levels, which is exactly what saves the current account from the binding problems discussed by G&M.

[11] Global accommodation is generally assumed to be preferred over local and intermediate.

available reading difficulty. In this way we account for the problem of quoted errors, the main objection against Potts, in a principled way, involving some pragmatic reasoning instead of an intention- and context-dependent superlexicon.

The next thing to note is that we have gotten an account of indexical shifting, (ii), for free, i.e. without the need for monsters or related context preservation methods. A quoted *me* is represented as "the individual x refers to as 'me'", which, when x is resolved to the quote source in the main clause, will simply refer to that reported speaker, i.e. Bush in (10a) and (15).[12]

Let's check the other use and mention characteristics identified in sections 2-3. The prohibition of extraction, (iii), is derived from the fact that, semantically, the quotation is analyzed as a single higher-order variable. The final two use characteristics, the use-inference (v), and Davidson's type argument (vi), are also trivially satisfied. But (iv) is problematic: How can ellipsis and anaphora reach into quotations and pick up a part seemingly hidden deep inside a mentioned phrase (or in the corresponding presupposed variable associated with it). I will solve this by "quote-breaking" in section 4.

Let me summarize the comparison of the two semantic accounts of quotation. We've replaced Potts' monsters, 2D lfs, stipulated projection, and intentional superlexicon, with an independently motivated and formalized dynamic theory of presuppositions that are generated in situ and resolved pragmatically. This resulted in a superior account of quoted indexicals and errors. The analysis of (i)-(iii) also proves the superiority of a hybrid account like G&M to the pragmatic accounts of e.g. Recanati. Anaphoric and elliptical links remain problematic. In the next section we add two more problems before proposing a uniform solution.

4 Quote-Breaking

The main objection that has been raised against these kinds of semantic analyses is the false prediction that only constituents can be quoted (Cumming 12, Abbott 13, De Brabanter 14). This prediction comes from the type-theoretic restriction inherent in definition (19), viz. that the semantic type of what's expressed by a quotation is determined by the syntactic category of the quoted expression itself, which presupposes that the quoted expression *has* a category which in turn means that it must be a constituent. I originally considered this an advantage, and used it to explain the infelicity of (23):

(23) *Life "is what" happens while you're making other plans
[(judgment from Maier & Geurts 15)]

However it was soon pointed out that there are plenty of counterexamples to this 'constituent hypothesis':

(24) a. She said the dog ate "strange things, when left to its own devices"
[(Abbott 13)]

[12] This has potentially profound implications for the analysis of shifted indexicality in languages like Amharic, cf. fn. 5.

b. David said that he had donated "largish sums, to several benign institutions" [(Abbott 13, De Brabanter 14)]
c. Yet Craig remains confident that the pitching 'will come round sooner or later. We just have to hope everybody stays healthy.' [taken from the *Cambridge Grammar* by Cumming (12:87)]

These quotations all involve two distinct constituents of the larger discourse. In each case no syntactic theory[13] would allow the two constituents to combine into a larger, interpretable constituent, so the data in (24) (which I assume to be fully felicitous) are indeed counterexamples to G&M.

In addition to this, we have already encountered another obstacle for G&M in the fact (empirically proven by Watters & Gundel 7) that anaphora and ellipsis can pick up material opaquely hidden inside quotation marks (cf. discussion of (8) in section 3.2). A third objection comes from Cumming (12), who points out that G&M cannot account for quantifier raising out of quotations, as in the reverse scope reading of (25) (i.e. with a different student for each professor):

(25) The dean asked that a student 'accompany every professor' [(Cumming 12:81)]

Again, the data seems real enough, and indeed G&M would block such a reading. G&M needs revision.

4.1 Breaking Non-constituents

Let's start with the non-constituents, focusing on example (24a). Where exactly does the presuppositional account break down? If we proceed constructing a preliminary DRS as sketched in 3.2 above, we see that, in order to fill the object NP position of 'ate', we must construct a presupposition ($\mathtt{z}$) of type e (ignoring distracting intricacies introduced by the semantics of bare plurals, read $(et)t$ (generalized quantifier) for e if you will) on the basis of the quotation. Thus, we get:

$$\left[\mathtt{x}\middle|\begin{array}{l}\mathtt{she(x)}\\ \mathsf{SAY}_{\mathtt{x}}\left[\mathtt{y}\middle|\begin{array}{l}\mathtt{dog(y),\ ate(y,z)}\\ \partial\left[\mathtt{z}\middle|\mathtt{say(x,}\ulcorner\text{strange things, when left to its own devices}\urcorner\mathtt{,z)}\right]\end{array}\right]\end{array}\right]$$

But the mention condition in the presupposed contribution of the quotation does not fulfill the wellformedness requirement of (19): the syntactic category of the mentioned phrase does not match the required type e of the object position of 'ate'. In fact, the mentioned phrase does not even have a single identifiable syntactic category—it is not a constituent. Thus, we are unable to construct a proper preliminary DRS.

[13] Paul Dekker (p.c.) has pointed out to me that there are more flexible theories, like the Lambek calculus, that may be able to interpret such non-constituents, thus in effect solving this whole problem. These theories obviously have their share of problems, but this might be an interesting application. T.b.c.

Let's take a closer look at the internal structure of this problematic quotation. Basic syntax shows the quotation to be made up out of two incompatible constituents, a plural NP and a *when*-clause (CP), neatly separated by a comma:

(26) the dog [$_{\text{VP}}$[ate [$_{\text{NP}}$"strange things]]],[$_{\text{CP}}$when left to its own devices"]

The two main constituents making up the quotation correspond to types e and $(et)et$, respectively, clearly showing, again, their incompatibility.

A first impulse may be to revise our theory of constituent structure (i.e. syntax) to accommodate these examples. Of course it is *theoretically possible* to devise a grammar that does that. For example, if we allow product types, we could analyze the quotation as a product of its two constituents, i.e. as having type $e \cdot (et)et$, and then give 'ate' the type $(e \cdot (et)et)et$. Instead of our type e object z we'd use a presupposed variable P of type $e \cdot (et)et$. But what is the motivation for type-shifting the transitive verb 'ate' from eet into $(e \cdot (et)et)et$, other than fixing a flaw in our theory? This ad hoc shifting really is a non-starter.

My proposal is to break up non-constituent quotes into appropriate constituents, in this case into the NP and the CP, and to consider them as separate quotations:

(27) the dog [$_{\text{VP}}$[ate [$_{\text{NP}}$"strange things"]]],[$_{\text{CP}}$"when left to its own devices"]

$$\left[\mathtt{x}\,\middle|\,\begin{array}{l}\mathtt{she}(\mathtt{x})\\ \mathsf{SAY}_{\mathtt{x}}\left[\mathtt{y}\,\middle|\,\begin{array}{l}\mathtt{dog}(\mathtt{y}),(\mathtt{P}(\mathtt{ate}_{eet}(\mathtt{z})))(\mathtt{y})\\ \partial\left[\mathtt{z}_e\middle|\mathtt{say}(\mathtt{x},\ulcorner\text{strange things}\urcorner,\mathtt{z})\right]\\ \partial\left[\mathtt{P}_{(et)et}\middle|\mathtt{say}(\mathtt{x},\ulcorner\text{when left to its own devices}\urcorner,\mathtt{P})\right]\end{array}\right]\end{array}\right]$$

In words: the asserted contribution of the complement clause contains now two free variables, P (of type $(et)et$) and z (e), corresponding to the two presuppositions formed from the two quoted parts. Both these presuppositions are fully in accordance with the type-theoretic restriction imposed on say in (19). Accommodation gives:

$$\rightsquigarrow\quad\left[\mathtt{x},\mathtt{z},\mathtt{P}\,\middle|\,\begin{array}{l}\mathtt{she}(\mathtt{x})\\ \mathtt{say}(\mathtt{x},\ulcorner\text{strange things}\urcorner,\mathtt{z})\\ \mathtt{say}(\mathtt{x},\ulcorner\text{when left to its own devices}\urcorner,\mathtt{P})\\ \mathsf{SAY}_{\mathtt{x}}\left[\mathtt{y}\middle|\mathtt{dog}(\mathtt{y}),(\mathtt{P}(\mathtt{ate}(\mathtt{z})))(\mathtt{y})\right]\end{array}\right]$$

Very roughly, this output may be paraphrased as: She said that when, as she put it, "left to its own devices" the dog ate, what she called, "strange things". Moreover, since both quotations are perfectly fine English phrases, we may assume that x used them to express their normal meanings, i.e. we can defeasibly fix z and P by adding $\mathtt{z} = |\text{strange things}|$ and $\mathtt{P} = |\text{when left to its own devices}|$. I will not further specify which lfs those are as that would involve too many irrelevant assumptions on the interpretation of *when*-clauses and bare plurals.

But then, can we still derive the infelicity of the non-constituent quote in (23)? The examples in (24) suggest a restriction on quote-breaking based on the observation that all those felicitous cases involve a clearly marked, natural breaking point in the form of a comma or period. However, the data in the next subsections demonstrate the need for a generalized quote-breaking procedure[14] that is not so constrained. Perhaps, then, the infelicity of (23) is merely a pragmatic affair: why would anyone ever want to quote precisely those two words, *is what*, of Lennon's original utterance?

4.2 Anaphora and Ellipsis

Without quote-breaking, we'd assign the following DRS (with highly oversimplified account of VP ellipsis as higher-order anaphora) to our earlier ellipsis/anaphora example:

(28) "My girlfiend bought me this tie," said John, but I don't think she did [=(8)]

$$\left[\begin{array}{l|l} \mathtt{x} & \mathtt{john}(\mathtt{x}), \mathsf{SAY}_{\mathtt{x}}\mathtt{p} \\ & \partial\left[\mathtt{p}\,\middle|\,\mathtt{say}(\mathtt{x}, \ulcorner\text{my girlfriend bought me this tie}\urcorner, \mathtt{p})\right] \\ & \neg\mathsf{THINK}_{\mathtt{i}}\left[\begin{array}{l|l} & \mathtt{P}(\mathtt{y}) \\ & \partial[\mathtt{y}|\mathtt{she}(\mathtt{y})],\ \partial[\mathtt{P}|\mathtt{did}(\mathtt{P})] \end{array}\right] \end{array}\right]$$

The problem with this representation is that there is no way to bind the presuppositions $\mathtt{y}$ (from the *she* in the second clause) and $\mathtt{P}$ (from *did*), even after succesful accommodation of the quotation induced presupposition $\mathtt{p}$.

Quote-breaking would solve this as follows:

(29) "My girlfiend" "bought me this tie," said John, but I don't think she did

$$\left[\begin{array}{l|l} \mathtt{x} & \mathtt{john}(\mathtt{x}) \\ & \mathsf{SAY}_{\mathtt{x}}\left[\begin{array}{l|l} & \mathtt{Q}(\mathtt{z}) \\ & \partial[\mathtt{z}|\mathtt{say}(\mathtt{x}, \ulcorner\text{my girlfriend}\urcorner, \mathtt{z})] \\ & \partial[\mathtt{Q}|\mathtt{say}(\mathtt{x}, \ulcorner\text{bought me this tie}\urcorner, \mathtt{Q})] \end{array}\right] \\ & \neg\mathsf{THINK}_{\mathtt{i}}\left[\begin{array}{l|l} & \mathtt{P}(\mathtt{y}) \\ & \partial[\mathtt{y}|\mathtt{she}(\mathtt{y})],\ \partial[\mathtt{P}|\mathtt{did}(\mathtt{P})] \end{array}\right] \end{array}\right]$$

$$\rightsquigarrow \left[\begin{array}{l|l} \mathtt{x}, \mathtt{z}, \mathtt{Q} & \mathtt{john}(\mathtt{x}),\ \mathtt{say}(\mathtt{x}, \ulcorner\text{my girlfriend}\urcorner, \mathtt{z}) \\ & \mathtt{say}(\mathtt{x}, \ulcorner\text{bought me this tie}\urcorner, \mathtt{Q}) \\ & \mathsf{SAY}_{\mathtt{x}}[\,|\mathtt{Q}(\mathtt{z})] \\ & \neg\mathsf{THINK}_{\mathtt{i}}[\,|\mathtt{P}(\mathtt{y})\ \partial[\mathtt{y}|\mathtt{she}(\mathtt{y})],\ \partial[\mathtt{P}|\mathtt{did}(\mathtt{P})]] \end{array}\right] \rightsquigarrow \begin{array}{l} \mathtt{P}\mapsto\mathtt{Q} \\ \mathtt{y}\mapsto\mathtt{z} \end{array}$$

$$\rightsquigarrow \left[\begin{array}{l|l} \mathtt{x}, \mathtt{z}, \mathtt{Q} & \mathtt{john}(\mathtt{x}),\ \mathtt{say}(\mathtt{x}, \ulcorner\text{my girlfriend}\urcorner, \mathtt{z}) \\ & \mathtt{say}(\mathtt{x}, \ulcorner\text{bought me this tie}\urcorner, \mathtt{Q}) \\ & \mathsf{SAY}_{\mathtt{x}}[\,|\mathtt{Q}(\mathtt{z})] \\ & \neg\mathsf{THINK}_{\mathtt{i}}[\,|\mathtt{Q}(\mathtt{z})] \end{array}\right]$$

That is, in the final interpretation of the *but*-clause: *she* = *the person John referred to as 'my girlfriend'*, and *did* = *what John referred to as 'bought me this tie'*.

---

[14] Perhaps even so general as breaking a quote all the way down to its atomic constituents, as suggested by Ede Zimmermann (p.c.).

4.3 Quantifier Raising

Finally, we turn to Cumming's scope objection. If we analyze the quotation in (25) as a single constituent, triggering a presupposed property, we can derive only the reading where the student has wide scope:

(30) The dean asked that a student "accompany every professor" [≈(25)]

$$\left[\mathtt{x}\ \middle|\ \begin{array}{l}\mathtt{dean(x)}\\ \mathsf{ASK}_{\mathtt{x}}\left[\mathtt{y}\ \middle|\ \begin{array}{l}\mathtt{student(y), P(y)}\\ \partial\left[\mathtt{P}\middle|\mathtt{say(x,}\ulcorner\text{accompany every professor}\urcorner\mathtt{, P)}\right]\end{array}\right]\end{array}\right]$$

After we've accommodated $\mathtt{P}$ we're done: the dean asked that there be a student with the property the dean referred to as "accompany every professor". Even if we were to add $\mathtt{P}=|\text{accompany every professor}|$, with appropriate lf for the quoted phrase, we'd never get the student inside the scope of *every*.

Breaking the quotation into a transitive verb *"accompany"* and object *"every professor"*, reduces the problem posed by (25) to the general problem of deriving inverse scope readings. We could, for instance, invoke Hendriks' ([16]) compositional, semantic treatment of inverse scope in terms of argument raising. To do this, we have to analyze the contributions of both the subject and the object as generalized quantifiers (which in turn requires that we break the quote to get to the object NP).[15]

(31) [...] asked that $[_{\text{NP}}$a student$]$ $[_{\text{VP}}$ $[_{\text{V}}$"accompany"$]$ $[_{\text{NP}}$"every professor"$]]$

$$\rightsquigarrow \left[\mathtt{x}\ \middle|\ \mathsf{ASK}_{\mathtt{x}}\left[\ \middle|\ \begin{array}{l}\mathtt{R}\left(\lambda\mathtt{P}\left[\mathtt{y}\middle|\mathtt{student(y), P(y)}\right]_{(et)t}\ ,\ \mathtt{X}_{(et)t}\right)\\ \partial\left[\mathtt{R}_{eet}\middle|\mathtt{say(x,}\ulcorner\text{accompany}\urcorner\mathtt{, R)}\right]\\ \partial\left[\mathtt{X}_{(et)t}\middle|\mathtt{say(x,}\ulcorner\text{every professor}\urcorner\mathtt{, X)}\right]\end{array}\right]\right]$$

We resolve the quotation presuppositions $\mathtt{R}$ and $\mathtt{X}$ globally, while in the embedded DRS we have to raise both the subject and object arguments of $\mathtt{R}$ in order to fit in the two given arguments. As Hendriks shows, the orders in which we can raise these two arguments gives us exactly the two scope possibilities that the sentence has. In particular, if we raise the second, subject, argument first, we get reverse scope. Note that more syntactically driven analyses of scope ambiguities in terms of movement may work just as well, but also requires that the quotation be broken up (for mentioning in the technical sense of $\ulcorner\ldots\urcorner$, would block all movement).

In conclusion, I've shown how quotation cannot be analyzed as pure, metalinguistic mention (contra e.g. Kaplan [4]), nor as pragmatic language use (contra e.g. Recanati [1]). A hybrid use-mention account is needed. The first formally explicit hybrid semantic attempt, Potts' ([2]) 2D semantics, was shown to have difficulty dealing with quoted indexicals and errors. Geurts and Maier's ([3]) presuppositional system solves these issues satisfactorily, shifting some of the load off the lexicon, into pragmatics. Remaining problems for a G&M-style account

[15] In DRT, the generalized quantifier *a student* looks like this: $\lambda\mathtt{P}\left[\mathtt{y}\middle|\mathtt{student(y), P(y)}\right]$.

included quoted non-constituents, anaphoric and elliptical dependencies across quotation marks, and quantifier raising out of quoted matter. These are addressed by adding a quote-breaking procedure, that breaks a quotation into appropriate smaller constituents and reassembles the corresponding quote-shifted (*use*) interpretations compositionally.

References

Recanati, F.: Open quotation. Mind 110, 637–687 (2001)

Potts, C.: The dimensions of quotation. In: Barker, C., Jacobson, P. (eds.) Direct Compositionality, pp. 405–431. Oxford University Press, Oxford (2007)

Geurts, B., Maier, E.: Quotation in context. Belgian Journal of Linguistics 17, 109–128 (2005)

Kaplan, D.: Demonstratives. In: Almog, J., Perry, J., Wettstein, H. (eds.) Themes from Kaplan, pp. 481–614. Oxford University Press, Oxford (1989)

Schlenker, P.: A plea for monsters. Linguistics and Philosophy 26, 29–120 (2003)

Partee, B.: The syntax and semantics of quotation. In: Anderson, S., Kiparsky, P. (eds.) A Festschrift for Morris Halle, Holt, Rinehart and Winston, pp. 410–418 (1973)

Watters, S., Gundel, J.: An empirical investigation of the relation between coreference and quotations. In: Branco, A. (ed.) DAARC 2007. LNCS (LNAI), vol. 4410, Springer, Heidelberg (2007)

Davidson, D.: Quotation. Theory and Decision 11, 27–40 (1979)

Stainton, R.: Remarks on the syntax and semantics of mixed quotation. In: Murasugi, K., Stainton, R. (eds.) Philosophy and Linguistics, pp. 259–278. Westview Press, Boulder (1999)

Grice, P.: Meaning. The Philosophical Review 66, 377–388 (1957)

van der Sandt, R.S.: Presupposition projection as anaphora resolution. Journal of Semantics 9, 333–377 (1992)

Cumming, S.: Two accounts of indexicals in mixed quotation. Belgian Journal of Linguistics 17, 77–88 (2005)

Abbott, B.: Some notes on quotation. Belgian Journal of Linguistics 17, 13–26 (2005)

de Brabanter, P.B.: Problems with semantic accounts of hybrid quotations [abstract for talk at Sinn und Bedeutung 10, Berlin] (2005)

Maier, E., Geurts, B.: The semantics of subclausal quotation [handout TaBu day, Groningen] (2004)

Hendriks, H.: Studied flexibility: Categories and types in syntax and semantics. PhD thesis, University of Amsterdam (1993)

A Modifier Hypothesis on the Japanese Indeterminate Quantifier Phrase

Mana Kobuchi-Philip*

UiL-OTS, Utrecht University
mana.kobuchi@let.uu.nl

Abstract. Shimoyama's [20, 21] analysis of the Japanese indeterminate quantifier construction (e.g. *dono gakusei-mo odotta.* 'Every student danced') is superior to previous analyses in the sense that it closely adheres to the Principle of Compositionality. However, under this analysis the *mo*–phrase of the form [[…wh...]$_{NP}$-*mo*], is analyzed as a generalized quantifier of semantic type <<e,t>,t>, and this gives rise to a type-theoretical problem. In order to overcome the difficulties, this paper proposes an alternative analysis in which the *mo*-phrase is treated as a modifier of semantic type <<e,t>,<e,t>>. It also discusses sentences in which *mo* combines with a PP and an IP and argues that here the modifier hypothesis of the *mo*-phrase can still be maintained.

1 *Mo*-Phrase as a Generalized Quantifier

Shimoyama [20, 21] offers a new semantic analysis of Japanese indeterminate quantifier constructions with *mo* of the type exemplified in (1):

(1) a. [[*dono gakusei-no okaasan*]$_{NP}$-*mo*] *odotta*.
which student GEN mother MO danced
'Every mother of a student danced.'

b. [[*doko -kara*]$_{PP}$-*mo*] *shootaijoo -ga todoita*.
where from MO invitation card NOM arrived
'An invitation card arrived from every place.'

c. *Taro-wa* [[*dare -ga denwashite*]$_{IP}$-*mo*] *deru*.
Taro TOP who NOM call MO answer
'Taro answers no matter who calls.'

In accordance with earlier work (e.g. Ohno [18], Nishigauchi [16, 17], Watanabe [23], von Stechow [22], Hagstrom [4]), *mo* is analyzed as a universal quantifier in Shimoyama's theory. However, unlike earlier authors, Shimoyama proposes that the restrictor of the universal quantifier *mo* is directly provided by the whole sister phrase of *mo*. Thus, in (1a), the restrictor of *mo* is the entire NP *dono gakusei-no okaasan*

* For critical comments and constructive suggestions, I thank the two anonymous reviewers as well as the audience of LENLS 2007, Taalkunde in Nederland 2007, the 134th meeting of the Linguistic Society of Japan, and Quantifier Modification workshop at ESSLLI 2007, where parts of the materials in this paper were presented. All shortcomings are mine.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 201–213, 2008.

'which student's mother' (the direct restrictor view), rather than the embedded NP *dono gakusei* 'which student' (the embedded restrictor view). Under the embedded restrictor view, the fact that the quantifier *mo* takes the embedded NP as its restrictor gives rise to a serious compositionality problem, since the quantifier and the embedded NP are separated from each other. In contrast, under Shimoyama's direct restrictor view, the composition of *mo* with its restrictor turns out to be completely straightforward in the surface syntax, since *mo* and its sister phrase combine directly with each other, by definition. Note that the sister phrase of *mo* is not always an NP. As shown in (1b) and (1c), *mo* may also compose with a PP or an IP. In order to account for the apparent cross-categoriality of its sister phrase, Shimoyama suggests that *mo* is a quantifier of type $<<\tau,t>,<<\tau,t>,t>>$, where τ is a variable ranging over any semantic type, as defined in (2a). This entails that the *mo*-phrase as a whole is also cross-categorial, as shown in (2b):

(2) a. MO = $\lambda P \lambda Q \forall x[P(x) \rightarrow Q(x)]$, where $x \in D_\tau$, and P, Q $\in D_{<\tau,t>}$.
 b. [[X]-MO] = $\lambda Q \forall x[P(x) \rightarrow Q(x)]$, of type $<<\tau,t>,t>$.

2 Problems

Despite its elegance, Shimoyama's analysis becomes problematic with respect to type assignment, as well as with the syntactic status of the *mo*-phrase, when one considers the following data:

(3) a. [*gakusei-no okaasan*] *-ga* [[*dono hito*]$_{NP}$ *-mo*] *odotta.*
 student GEN mother NOM which person MO danced
 b. [[[*gakusei -no okaasan*]$_{NP}$ [[*dare*]$_{NP}$*-mo*]] *-ga*]$_{DP}$ *odotta.*
 student GEN mother who MO NOM danced
 c. [[[*dono gakusei-no okaasan*]*-mo*] *-ga*]$_{DP}$ *odotta.*
 which student GEN mother MO NOM danced
 'All the mothers of the students danced.'

In (3a), the sister phrase of *mo* is *dono hito* 'which person'. Adopting Hamblin's [5] semantics of questions, Shimoyama treats the NP containing an indeterminate pronoun as denoting the set of the alternative individuals in the context, i.e. type <e,t>. Thus, the *mo*-phrase in (3a) is an ordinary generalized quantifier of type <<e,t>,t>, just like the *mo*-phrase in (1a). However, (3a) differs from (1a) in that it contains another phrase, namely the subject DP, which is case-marked by nominative *ga*. Since a case marker is generally postpositioned in Japanese, the following *mo*-phrase cannot be part of the subject DP. This only leaves the possibility that the *mo*-phrase composes with the predicate. However, in this case, this composition already yields a truth value, without including the subject in the calculation. Consequently, (3a) is incorrectly predicted to be ill-formed under Shimoyama's analysis.

Next, in (3b), since the sister phrase of *mo* is the NP *dare* 'who', the *mo*-phrase is again a generalized quantifier. However, here it syntactically composes with the NP *gakusei-no okaasan* 'student's mother', and the resulting constituent is case-marked by nominative *ga* in its entirety, forming the subject DP. The productivity of this particular construction seems to be limited, and some native speakers of Japanese find it slightly marked.[1] However, a Google search shows that this construction does appear in written documents frequently enough, whereby I assume that this is syntactically possible in Japanese grammar. Given this, if the *mo*-phrase is of type <<e,t>,t>, it must somehow compose with the preceding NP and must somehow form a subject DP (<<e,t>,t>). This process is quite unclear.

Finally, consider (3c). (3c) is exactly the same as (1a) except that the nominative case-marker *ga* has been included. Under Shimoyama's account, (1a) and (3c) are not distinguishable; both the *mo*-phrase in (1a) and the subject DP in (3c) are generalized quantifiers. However, a syntactic distinction between the bare *mo*-phrase and the nominative-marked *mo*-phrase can rather clearly be seen in the following set of data involving an imperative construction:[2]

(4) a. *hashire!*
run
'Run!'

b. *omae-ga hashire!*
you NOM run
'(Not someone else but) YOU run!'

c. *doitsu -mo hashire!*
which one MO run
'All run!'

d. *doitsu -mo -ga hashire!*
which one MO NOM run
'(Not some but) EVERYONE run!'

An imperative sentence is most natural when the subject is omitted, as shown in (4a). Adding a subject such as *omae-ga* 'you-NOM' to an imperative sentence, as in (4b), is pragmatically restricted to a situation in which the speaker adds the contrastive emphasis 'not someone else but YOU' to the command 'do X'. The markedness of this form can be detected by the fact that the subject in such a case is almost always stressed.[3] Thus, (4b) is marked while (4a) is unmarked. Similarly, (4c), with a

---

[1] Takao Gunji, Ikumi Imani, p.c. Gunji also points out that a sentence such as (i) would be well-formed and questions the validity of the assumption that *ga*-marked nominal elements are generally generalized quantifiers:

(i) *gakusei-ga dare-mo-ga odotta.*
student NOM who MO NOM danced
'All the students danced.'

Such a sentence is acceptable as a type of the multiple nominative constructions. Here it does not seem that both *ga*-marked nominal elements have equal status as the subject of the sentence, as has been suggested in the literature (e.g. Heycock and Lee [8], Heycock [7]). Formal details aside, I speculate that *dare-mo-ga* itself is the subject of the predicate *odotta* 'danced', and *gakusei* 'student' is related to *daremo* 'everyone' in the sense that the latter is part of the former in an abstract sense. See (ii) as a clearer example of this relation:

(ii) *John -ga otooto -ga kekkonshita.*
John NOM younger brother NOM got married
'John is such that (his) younger brother got married.'

Note here that the sentence asserts two propositions, namely that (John's) younger brother has the property of having got married and that John has the property that his younger brother got married. However, the precise semantics of the multiple-nominative construction in languages such as Japanese and Korean is beyond the scope of this paper and its examination has to be postponed to another occasion.

[2] I thank Masaaki Kamiya for pointing this out to me.

[3] The same general phenomenon appears to arise in English as well.

mo-phrase, is a perfectly grammatical and unmarked imperative sentence, while (4d), where the nominative case-marker *ga* has been added to the *mo*-phrase, is grammatical though marked. In this case, ‘everyone’ is emphasized, as shown in the gloss. These observations show that a *mo*-phrase with the nominative case marker is quite distinct from one without a case marker. Specifically, the former is a syntactic subject but the latter is not in sentences like (4). This strongly suggests that the *mo*-phrase in (1a) and the *mo*-phrase with a nominative case marker in (3c) have distinct syntax.

In sum, the analysis in (2) raises some type-theoretical and empirical problems with respect to data such as (3a)-(3c). Under Shimoyama’s [21] analysis, the type of the *mo*-phrase is elevated to <<<e,t>,t>,t>, closely adhering to Hamblin semantics. However, the same problems remain since every element is reanalyzed in an equally higher-order type system.

3 Preverbal *mo*-Phrase as a Modifier

Given the problems discussed in the previous section, I suggest an alternative analysis in which the *mo*-phrase in (3a-c), as well as in (1a), is a modifier of type <<e,t>,<e,t>>, rather than a generalized quantifier of type <<e,t>,t>. This analysis is motivated by the hypothesis that the *mo*-phrase in (3a) is a VP-modifier. I offer three pieces of empirical evidence in support of this. First, as I have already mentioned above, a case marker is generally postpositioned in Japanese, and thus a right-adjacent *mo*-phrase cannot be part of the subject DP. Second, that the *mo*-phrase does not compose with the subject can be seen from the fact that a direct object can be placed left-adjacent to it:

(5) *gakusei-no okaasan-ga* [*ima naratta bakari-no waltz -o*]$_i$
student GEN mother NOM now learned just GEN waltz ACC
dono hito -mo t_i *joozuni odotta*
which person MO skillfully danced
‘The students’ mothers all skillfully danced the waltz that they have just learned now.’

This sentence, in which the direct object is scrambled from the position indicated by the trace, is perfectly well-formed syntactically. Assuming that the landing site of the direct object indicates the left edge of the maximal projections which contain the main verb, this clearly shows that the *mo*-phrase in (5) belongs to the verbal domain. The interpretation would be calculated on the basis of the basic unscrambled form (Lasnik and Saito [15]); thus the *mo*-phrase here composes with a VP consisting of a direct object, a manner adverb (*joozuni* ‘skillfully) and a verb (*odotta* ‘danced’). Since VP is of type <e,t>, the *mo*-phrase must be a VP-modifier, namely of type <<e,t>,<e,t>>. The third piece of evidence supporting the modifier hypothesis is the fact that it is possible for two or more *mo*-phrases to co-occur in a single sentence:

(6) *Kono jikken -noojoo -de Taroo-wa*
this experiment farm at Taro TOP

dono uma -no taijuu-mo dono ushi-no taijuu -mo
which horse GEN weight MO which cow GEN weight MO
dono hitsuji-no taijuu-mo maitsuki ichido kirokusuru.
which sheep GEN weight MO every month once record
'At this experimental farm, Taro records the weight of every horse, every cow and every sheep once a month.'

In (6), the weight of the three animals is thematically the theme of Taro's recording. However, it is not conceivable that the three *mo*-phrases are three occurrences of a direct object, under the basic Case-theoretical consideration. On the other hand, if a *mo*-phrase is assumed to be adverbial, naturally it would have no problem of multiple occurrences. On the basis of these observations, I assume that at least the preverbal *mo*-phrase is a VP-modifier.

Incorporating this syntactic hypothesis into the semantics, a formal analysis for (3a) can be as shown in (7):

(7) a. [$\varnothing_{the}$ [*gakusei-no okaasan*]-*ga*] [[*dono hito*] -*mo*] [*odotta*]]].
student GEN mother NOM which person MO danced
'Every mother of a student danced.'

b. dono hito 'which person': $\lambda x[\text{PERSON}_{AT}(x)]$ (=a set of persons)
mo 'MO': $\lambda R\lambda P\lambda x[x \prod \oplus(R \cap AT(\oplus P)) \wedge |AT(x)| \geq 2]$
odotta 'danced': $\lambda x[\text{DANCED}(x)]$
gakusei-no okaasan
'student's mother': $\lambda x[\text{STUDENT'S MOTHER}(x)]$
$\varnothing_{the}$: $\lambda X\lambda Y[Y(\oplus X)]$ (Link [14], Landman [13])

c. $\oplus\text{STUDENT'S MOTHER} \prod \oplus(\text{PERSON}_{AT} \cap AT(\oplus\text{DANCED})) \wedge |AT(\oplus\text{STUDENT'S MOTHER})| \geq 2$

Following Shimoyama, I adopt Hamblin's semantics for the NP containing an indeterminate pronoun and assume that this NP provides the restrictor for the quantifier *mo*. Thus, the NP *dono hito* 'which person' denotes the set of people in the context. However, I claim that, given that this NP functions as a restrictor, its denotation must be a set of only atoms in the sense of Link [14], rather than a set containing both atoms and sums, as is the case for an ordinary NP (Link [14], Landman [13]). This is because the restrictor denotation of a quantifier construed with a count noun must satisfy a basic logical condition of atomicity, i.e. the quantified objects must be discrete (Kratzer [12], Chierchia [3], Landman [13], Kobuchi-Philip [9, 10]). Now, the quantifier *mo* composes with its sister phrase NP *dono hito* 'which person', and the outcome is a modifier of type <<e,t>,<e,t>>. This composes with the predicate *odotta* 'danced', forming a larger VP of type <e,t>. Turning to the subject, the bare NP *gakusei-no okaasan* 'student's mother' denotes the set of mothers of students in the context. Of course, this NP cannot by itself compose with the predicate, since it is an <e,t> element. However, in a language such as Japanese which generally lacks determiners, there is a need for a general mechanism that type-lifts an NP to a generalized quantifier. As discussed in Kobuchi-Philip [9, 10], this can be formulated as the assumption that Japanese contains phonetically null definite and indefinite determiners.[4] Adopting this hypothesis, I assume that the null determiner here is definite, given that *dono hito-mo* 'every person' is a strong quantifier. This definite DP yields the supremum of the set denoted by the NP (Link [14], Landman [13]). Note that, under this analysis, the denotation of the quantifier *mo* does not contain a universal quantifier. However, observe that the universality is embedded in the meaning of *mo*, i.e. the part relation between the definite subject DP and the predicate. This part relation logically entails a subset relation, which is also logically entailed by a universal quantifier. This is what is expressed in (7c): The sum of students' mothers is part of the sum of every dancing person. In plain English, every student's mother danced.

Assuming (7) for (3a), let us now consider again (1a), repeated here:

(1) a. [[*dono gakusei-no okaasan*]$_{NP}$-*mo*] *odotta.*
which student GEN mother MO danced
'Every mother of a student danced.'

Note that this construction solely consists of a *mo*-phrase and a VP. Adopting what has been suggested in the syntax literature (e.g. Hasegawa [6]. Aoyagi [1]), I assume that this is a construction in which the entire DP is missing. Thus, I assume its structure as shown in (8b), in comparison to (8a) which represents (3a):

(8) a. [$\varnothing_{Det}$ NP]$_{DP}$ [MoP VP]$_{VP}$ (=3a)
b. [$\varnothing_{Det}$ [$\varnothing_{NP}$]]$_{DP}$ [MoP VP]$_{VP}$ (=1a)

The structural parallelism in (8) indicates that the interpretation of the two constructions can be calculated basically in a parallel fashion. This unification is quite an advantage. However, to assume a null DP entails the presence of an NP in addition to a null determiner. As before, this null determiner can be assumed to be definite, equivalent to English *the*. What, however, is the identity of the NP, whose presence is logically necessary for the semantic computation?

What I hypothesize here is that the lexical content of this null NP is inherited from the sister phrase of *mo*, as represented in (9):

(9) [$\varnothing_{the}$[***gakusei-no okaasan***]$_{NP}$ –*ga*]$_{DP}$
student GEN mother NOM
[[*dono* ***gakusei-no okaasan***]-*mo*] *odotta.*
which student GEN mother MO danced

This may at first appear ad hoc. However, there is independent motivation for positing the existence of a general inheritance mechanism of this sort, e.g. with regard to the partitive construction (Barker [2], Kobuchi-Philip [11]).[5] More importantly, however,

[4] A Japanese bare NP is ambiguous both with respect to plurality and definiteness. Assuming a null determiner and the plurality theory of Link [14] and Landman [13] accounts for the four-way ambiguity in the simplest manner.

[5] For example, in Kobuchi-Philip [11], *three of the boys* is analyzed with an empty N in association with the numeral, whose lexical value is inherited from the whole-denoting N *boys*.

this is empirically motivated for the particular case under discussion. Recall that Shimoyama [20] had treated the *mo*-phrase in (1a) as a generalized quantifier, i.e. as equivalent to the subject of the sentence. In fact, many native speakers of Japanese find such an analysis intuitively plausible. On the basis of this intuition, it is rather natural to hypothesize that there is some close semantic relation between the *mo*-phrase and the null subject in question. The hypothesis turns out to be quite plausible when we consider the entailment relationship between a sentence with and without the subject in (10):[6]

(10) a. *dono gakusei-no peepaa-mo A datta.*
which student GEN paper MO A was
←→ *gakusei-no peepaa-ga dore -mo A datta.*
student GEN paper NOM which MO A was
'All the papers by a student got an A.

b. *dare-ga kaita e -mo homerareta.*
who NOM drew picture MO was praised
←→ *dareka -ga kaita e -ga dore -mo homerareta.*
someone NOM drew picture NOM which MO was praised
'The pictures someone drew were all praised.'

c. *doko -kara amerika -ni hairu hikooki -no jookyaku-mo*
where from America- to enter airplane GEN passenger MO
orita kuukoo-de pasupootokensa sareru.
get off airport at passport examination done
←→ *dokoka -kara amerika -ni hairu hikooki -no jookyaku-ga*
somewhere from America to enter airplane GEN passenger NOM
dare-mo orita kuukoo-de pasupootokensa sareru.
who MO get off airport at passport examination done
'The passengers of an airplane which comes into America from anywhere are all examined his passport.'

d. *dare -ga itsu doko -de kaita shoosetsu-mo koko-ni kiroku sareru.*
who NOM when where at wrote novel MO here in record done
←→ *dareka -ga itsuka dokoka -de kaita shoosetsu-ga*
Who NOM sometime somewhere at wrote novel NOM
dore -mo koko-ni kiroku sareru.
which MO here in record done
'The novels which someone wrote sometime somewhere are all recorded here.'

The first sentence in each of (10a-d) lacks a subject DP (*ga*-phrase). Given such a construction, it is always possible to construct a sentence corresponding to the second of each in (10a-d) that does have a subject. The lexical content of this subject is drawn from the NP inside the *mo*-phrase in the first sentence. Let us then tentatively adopt

[6] In (10), the entailment relationship holds under a non-specific reading of *dareka* 'someone', *itsuka* 'sometime', and *dokoka* 'somewhere' in the sentence on the right hand side of the entailment relation sign. It is normally more natural to assign a specific reading to such lexical elements in Japanese, since they occupy a position taking a wide scope.

this 'inherited lexical value'-hypothesis. Under this hypothesis, (1a) can formally be analyzed as shown in (11):

(11) a. [[$\varnothing_{\text{the}}$ $\varnothing_{\text{NP}}$] [[*dono gakusei-no okaasan*] –*mo*] [*odotta*]]].
which mother GEN mother MO danced
'Every mother of a student danced.'

b. dono gakusei-no okaasan
'which mother of student': $\lambda x[\text{STUDENT'S MOTHER}_{\text{AT}}(x)]$
mo 'MO': $\lambda R\lambda P\lambda x[x \prod \oplus(R\cap AT(\oplus P))\wedge|AT(x)|\geq 2]$
odotta 'danced': $\lambda x[\text{DANCED}(x)]$
→gakusei-no okaasan $\lambda x[\text{STUDENT'S MOTHER}(x)]$
'student's mother'
$\varnothing_{\text{the}}$: $\lambda X\lambda Y[Y(\oplus X)]$

c. $\oplus\text{STUDENT'S MOTHER} \prod \oplus(\text{STUDENT'S MOTHER}_{\text{AT}}\cap AT(\oplus\text{DANCED})) \wedge|AT(\oplus\text{STUDENT'S MOTHER})|\geq 2$

It should be noted here that the denotation of the null NP (indicated by → in 11) is not exactly the same as that of the sister phrase of *mo*. Recall that the denotation of the NP sister to *mo* was a set of only atoms, due to the atomicity condition of the restrictor. This atomicity condition does not apply to the denotation of a null NP simply because it is not functioning as a restrictor of a quantifier. Thus, the inherited NP denotes a set containing both atoms and sums, just like any ordinary Japanese NP. The logical representation in (11c) asserts that the supremum of the set denoted by students' mothers in the context is part of the sum of students' mothers who danced. In short, the students' mothers all danced. In this manner, (1a) and (3a) can be analyzed in a parallel fashion.

4 DP-Internal *mo*-Phrase as a Modifier

Under the simplest hypothesis, if the *mo*-phrase in (1a) and (3a) must be a predicate modifier for type-theoretical reasons, the other occurrences of the *mo*-phrase in (3b-c), namely the *mo*-phrase in a DP, should also have this property. This is what I pursue in what follows. Let us consider (3b), repeated here:

(3) b. [[[*gakusei -no okaasan*]$_{\text{NP}}$ [*dare-mo*]] *-ga*]$_{\text{DP}}$ *odotta.*
student GEN mother who MO NOM danced
'All the mothers of the students danced.'

In this sentence, we have an NP *gakusei-no okaasan* 'student's mother' and a *mo*-phrase in the subject DP. Recall that we assume the presence of a null determiner. In this case, it turns out that we immediately obtain a type-theoretically well-formed analysis of (3b) as shown in (12):

(12) [$\varnothing_{\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle}$ [[*gakusei -no okaasan*]$_{\langle e,t\rangle}$ [*dare-mo*]$_{\langle\langle e,t\rangle,\langle e,t\rangle\rangle}$] *-ga*] *odotta*$_{\langle e,t\rangle}$.
student GEN mother who MO NOM danced
'All the mothers of the students danced.'

We obtain this simply following the logic of our proposal. However, it might be objected that *dare-mo* 'everyone' is a strong quantifier, and as such cannot be interpreted unless its domain is specified. Indeed, in (3b), *gakusei-no okaasan* 'student-GEN mother' refers to a specific set of students' mothers; the entire set of the students' mothers in the context. The definiteness of the nominal element that composes with *dare-mo* is a general necessity due to the strength of the quantifier, and we can clearly observe it in the following examples as well:

(13) a. [[*John-to Bill*] [*dochira-mo*]]-*ga* *chikokushita.*
John and Bill both MO NOM arrived-late
'Both John and Bill arrived late.'
b. [[*omae-ra*][*doitsu- mo*]]-*ga* *gookakushitara meshi ogotteyaru.*
you-PL which.guy MO NOM pass.if meal treat
'If you all pass, I will treat you a meal.'

John-to Bill 'John and Bill' in (13a) and *omae-ra* 'you-PL' in (13b) are apparently definite DPs. Type-theoretically, a DP in the standard analysis is of type <<e,t>,t>, or else e, and it cannot compose with the *mo*-phrase if it is <<e,t>,<e,t>>. Is it necessary then to give up the predicate modifier analysis of the *mo*-phrase at this point? Perhaps not. As is well-known, a DP can be assigned type <e,t> in a predicative context (Partee [19]). If so, the *mo*-phrase can indeed compose with a DP, thanks to this potential for a predicative type. In this sense the *mo*-phrase which is local to DP can still be considered a predicate modifier. This definite DP, assigned the predicative type <e,t>, denotes a singleton set. Thus, the definite *gakusei-no okaasan* 'student's mother' denoting λx[⊕STUDENT'S MOTHER(x)] is a set of just one element, the supremum of the students' mothers in the context. The formal analysis of (3b), then, would be as shown in (14):

(14) a. [[$\varnothing_{the}$ [[*gakusei-no okaasan*] [*dare-mo*]]-*ga*] [*odotta*]].
student GEN mother who MO NOM danced
'Every mother of a student danced.'
b. dare 'who': $\lambda x[\text{PERSON}_{AT}(x)]$[7]
mo 'MO': $\lambda R\lambda P\lambda x[\ x\prod\oplus(R\cap AT(\oplus P))\wedge|AT(x)|\geq 2]$
→gakusei-no okaasan
'student's mother': $\lambda x[(\oplus\text{STUDENT'S MOTHER})(x)]$
$\varnothing_{the}$: $\lambda X\lambda Y[Y(\oplus X)]$
odotta 'danced': $\lambda x[\text{DANCED}(x)]$
c. $\text{DANCED}(\oplus\lambda x[x\prod\oplus(\text{PERSON}_{AT}\cap AT(\oplus(\oplus\text{STUDENT'S MOTHER}))) \wedge|AT(x)|\geq 2])$

The nominal element that composes with a *mo*-phrase local to DP (indicated by →) is a DP but has the predicative type of <e,t>. It denotes a singleton set. When this is calculated, the final logical representation in (14c) happens to have a couple of sum-operators. However, this is equivalent to the same formula with a single sum-operator.

---

7 *Dare* 'who' can be considered to denote a set of persons, thus it can be treated as identical in denotation to *dono hito* 'which person', which appears in (7) above.

In sum, the logical representation asserts that the sum of the students' mothers is an element of the set denoted by 'danced'. In plain English, every mother of a student danced.

Assuming (14) for sentence (3b), let us now go on to (3c), repeated here:

(3) c. [[[*dono* *gakusei-no* *okaasan*]-*mo*] -*ga*]$_{DP}$ *odotta*.
which student GEN mother MO NOM danced
'All the mothers of the students danced.'

This sentence consists of a *mo*-phrase that is associated with a nominative case marker and a VP. This is a sentence similar to (3b), differing only in that it lacks the nominal element in the position preceding the *mo*-phrase in the DP. Thus, I assume that (3b) and (3c) have the same structure, as represented in (15):

(15) a. [$\varnothing_{Det}$ [DP MoP]$_{NP}$]$_{DP}$ VP (=3b)
b. [$\varnothing_{Det}$ [$\varnothing_{DP}$ MoP]$_{NP}$]$_{DP}$ VP (=3c)

Note that the relationship between (3b) and (3c) is similar to that between (3a) and (1a) discussed earlier, the difference being that the nominal element, DP here, is overt in (3b) and covert in (3c). The interpretation of these sentence types can be calculated in the same fashion as before, the latter making use of the inherited lexical value mechanism discussed above. That is, the lexical value of the null DP is inherited from the nominal element inside the *mo*-phrase, as follows:

(16) [$\varnothing_{the}$ [[***gakusei-no okaasan***]$_{DP}$ [*dono* ***gakusei-no*** ***okaasan***-*mo*] –*ga*]$_{DP}$
student GEN mother which student GEN mother MO NOM
odotta.
danced

The DP supplied in (16) denotes the set of students' mothers in the context, and denotes a singleton set. The formal analysis of (3c), then, is as follows:

(17) a. [[$\varnothing_{the}$ [[*dono* *gakusei-no* *okaasan*] -*mo*] -*ga*] [*odotta*]].
which student GEN mother MO NOM danced
'Every mother of a student danced.'
b. dono gakusei-no okaasan
'which student's mother': $\lambda x[\text{STUDENT'S MOTHER}_{AT}(x)]$
mo 'MO': $\lambda R\lambda P\lambda x[\, x\prod\oplus(R\cap AT(\oplus P))\wedge|AT(x)|\geq 2]$
→gakusei-no okaasan
'students' mother': $\lambda x[(\oplus\text{STUDENT'S MOTHER})(x)]$
$\varnothing_{the}$: $\lambda X\lambda Y[Y(\oplus X)]$
odotta 'danced': $\lambda x[\text{DANCED}(x)]$
c. $\text{DANCED}(\oplus\lambda x[x\prod\oplus(\text{STUDENT'S MOTHER}_{AT}\cap AT(\oplus(\oplus\text{STUDENT'S MOTHER})))\wedge|AT(x)|\geq 2])$

The logical representation in (17c) asserts that the sum of the students' mothers is an element of the set denoted by 'danced'. In other words, it asserts that every mother of a student danced.

In this and the last section I have shown how the modifier hypothesis of the *mo*-phrase can formally be implemented, using the plurality theory of Link-Landman and supplemented by Partee's notion of possible predicative type assignment for a definite DP. The proposed analysis is advantageous in the following ways: (i) It gives rise to no type-theoretical problems such as those which Shimoyama's analysis encounters; (ii) it provides a basically uniform analysis of the sentence types illustrated in (1a), (3a), (3b) and (3c); (iii) this analysis adheres closely to the Principle of Compositionality; and (iv) it provides distinct semantics for the sentence with a bare *mo*-phrase and that with a *mo*-phrase associated with a case-marker, on the basis of their syntactic difference.

5 *Mo*-Phrase Containing a Non-NP as a Modifier

The principal motivation for the type-flexibility of *mo* in Shimoyama's [20] hypothesis, summarized in (2a) and repeated below, is the fact that the sister phrase of *mo* could be other grammatical categories aside from NP, as illustrated in (1b-c), repeated here:

(2) a. MO = $\lambda P \lambda Q \forall x[P(x) \rightarrow Q(x)]$, where $x \in D_\tau$, and P, Q $\in D_{<\tau,t>}$.

b. $[[X]\text{-MO}] = \lambda Q \forall x[P(x) \rightarrow Q(x)]$, of type <<τ,t>,t>.

(1) b. [[*doko -kara*]$_{PP}$-*mo*] *shootaijoo -ga todoita*.
where from MO invitation card NOM arrived
'An invitation card arrived from everywhere.'

c. *Taro-wa* [[*dare -ga denwashite*]$_{IP}$-*mo*] *deru*.
Taro TOP who NOM call MO answer
'Taro answers no matter who calls.'

Above I suggested that the *mo*-phrase which composes with an NP is a modifier of type <<e,t>,<e,t>>. This hypothesis leaves open the question concerning the apparent cross-categoriality of *mo*. Although I do not currently have a precise analysis for the *mo*-phrase composing with a PP or an IP, I would like to make some remarks on the matter.

The question is: Can our hypothesis that the *mo*-phrase is a modifier be maintained even covering PP-*mo* and IP-*mo* sentence types? I suspect that the answer is positive. Let us first consider the case of IP-*mo*, as in (1c). The interpretation of the IP-*mo* phrase here is 'no matter who calls'. This *mo*-phrase is actually a subordinate clause of the sentence and the main clause is *Taro-wa deru* 'Taro answers'. That is, the *mo*-phrase provides a further qualification to the proposition described by the main clause. In this sense, it can be considered as a sort of adverb of quantification. It is, then, a sentence modifier, if not a predicate modifier. If so, we may find a way to pursue our hypothesis under a situation semantics.

What about the case of the PP-*mo* in (1b)? The subject of this sentence is *shootaijoo-ga* 'invitation card-NOM', and the verb is *todoita* 'arrived'. The *mo*-phrase *doko-kara-mo* ‘from everywhere’ does indeed modify the verb. This seems supportive of the proposed modifier hypothesis. However, the difficulty here is that a PP without *mo*, such as *doko-kara* ‘from where’, would normally also be assumed to be a predicate modifier of type <<e,t>,<e,t>>, since a PP typically modifies an NP or a VP. If so, in (1b) *mo* would have to be of type <<<e,t>,<e,t>>,<<e,t>,<e,t>>>, contrary to what was suggested above. This might pose a tough technical problem. On the other hand, there might be a different type assignment possibility for PPs. Specifically, a PP may also be used as a predicate, e.g. *The letter is from the mayor* and *Tegami-wa shichoo-kara-da* ‘the letter is from the mayor’. In such cases, a PP could almost be considered as a property. If so, it might be claimed that a PP also has a potential of being of type <e,t>. Thus, there might be a way for our hypothesis to survive in the case of the sentence with PP-*mo*. I will leave the investigation of such a possibility to future research.

6 Summary

In this paper, I pointed out some type-theoretical and syntactic difficulties with Shimoyama’s [20, 21] analysis of Japanese indeterminate quantification. Specifically, we observed that, when there was an overt subject or an DP combining the *mo*-phrase within the nominal domain, a type-theoretical problem arose. Furthermore, there was another problem in terms of the fact that *mo*-phrase with and without *ga* could not be distinguished, even though there is a real syntactic difference between the two. In order to overcome these difficulties, I suggested an alternative analysis which treats the *mo*-phrase containing an NP as an element of semantic type <<e,t>,<e,t>>. I showed how this hypothesis could technically be implemented. Leaving the formal semantic analysis of the *mo*-phrase containing a PP and an IP for future research, I suggested that the *mo*-phrase in these cases is also a modifier. The latter seems to be a modifier categorized as an adverb of quantification, which perhaps can receive an adequate analysis within situation semantics. In the case of the former, although it can plausibly be considered to be a modifier, a precise type-theoretical analysis of the PP itself may hold the key to a solution consistent with the proposed modifier hypothesis.

References

1. Aoyagi, H.: On Association with Focus and Scope of Focus Particles in Japanese. In: Koizumi, M., Ura, H. (eds.) Formal Approaches to Japanese Linguistics 1, pp. 23–44. MITWPL, Cambridge, MA (1994)
2. Barker, C.: Partitives, Double Genitives and Anti-Uniqueness. Natural Language and Linguistic Theory 16, 679–717 (1998)
3. Chierchia, G.: Reference to Kinds across Languages. Natural Language Semantics 6, 339–405 (1998)
4. Hagstrom, P.: Decomposing Questions. Ph.D. Diss. MIT (1998)
5. Hamblin, C.: Questions in Montague English. Foundations of Language 10, 41–53 (1973)
6. Hasegawa, N.: Floating Quantifiers and Bare NP Expressions. In: Hasegawa, N. (ed.) Japanese Syntax in Comparative Grammar, pp. 115–145. Kuroshio, Tokyo (1993)

7. Heycock, C.: Syntactic Predication in Japanese. Journal of East Asian Linguistics 2, 167–211 (1993)
8. Heycock, C., Lee, Y.-S.: Subject and Predication in Korean and Japanese. In: Hoji, H. (ed.) Japanese /Korean Linguistics, pp. 239–253. CSLI, Stanford (1990)
9. Kobuchi-Philip, M.: Distributivity and the Japanese Floating Numeral Quantifier. Ph.D. Diss. CUNY Graduate Center (2003)
10. Kobuchi-Philip, M.: Identity of the Domain of Quantification for the Numeral. In: Tasmowski, L., Vogeleer, S. (eds.) Non-Definiteness and Plurality, pp. 267–290. John Benjamins, Amsterdam (2006)
11. Kobuchi-Philip, M.: Semantics of the Cardinal Partitive. In: Puig-Waldmüller, E. (ed.) Proceedings of the Conference Sinn und Bedeutung, vol. 11, pp. 390–402. Universitat Pompeu Fabra, Barcelona (2007)
12. Kratzer, A.: An Investigation of the Lumps of Thought. Linguistics and Philosophy 12, 607–653 (1989)
13. Landman, F.: Events and Plurality. Kluwer, Dordrecht (2000)
14. Link, G.: The Logical Analysis of Plural and Mass Terms: A Lattice-theoretical Approach. In: Bäuerle, R., Schwarze, C., von Stechow, A. (eds.) Meaning, Use, and Interpretation of Language, pp. 302–323. de Gruyter, Berlin (1983)
15. Lasnik, H., Saito, M.: Move <Emphasis Type="Italic">α</Emphasis>: Conditions on Its Application and Output. MIT Press, Cambridge (1992)
16. Nishigauchi, T.: Quantification in Syntax. Ph.D. Diss. University of Massachusetts, Amherst (1986)
17. Nishigauchi, T.: Quantification in the Theory of Grammar. Kluwer, Dordrecht
18. Ohno, Y.: Mo. In: Bach, E., Kratzer, A., Partee, B.H. (eds.) Papers on Quantification. NSF Grant Report. Department of Linguistics, University of Massachusetts, Amherst. Amherst, MA: GLSA, pp. 224–250 (1989)
19. Partee, B.: Noun Phrase Interpretation and Type-Shifting Principles. In: Groenendijk, J., de Jongh, D., Stokhof, M. (eds.) Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers, pp. 115–143. Foris, Dordrecht (1987)
20. Shimoyama, J.: Wh-Constructions in Japanese. Ph.D. Diss. University of Massachusetts, Amherst (2001)
21. Shimoyama, J.: Indeterminate Phrase Quantification in Japanese. Natural Language Semantics 14, 139–173 (2006)
22. von Stechow, A.: Against LF Pied-Piping. Natural Language Semantics 4, 57–110 (1996)
23. Watanabe, A.: Wh-in-Situ, Subjacency, and Chain Formation. In: WWW 1995, MITWPL, Cambridge MA (1992)

A Presuppositional Analysis of Definite Descriptions in Proof Theory

Koji Mineshima

Keio University
minesima@phil.flet.keio.ac.jp

Abstract. In this paper we propose a proof-theoretic analysis of presuppositions in natural language, focusing on the interpretation of definite descriptions. Our proposal is based on the natural deduction system of ε-calculus introduced in Carlström [2] and on constructive type theory [11,12]. Based on the idea in [2], we use the ε-calculus as an intermediate language in the translation process from natural language into constructive type theory. Using this framework, we formulate the process of presupposition resolution as the process of searching for a derivation in a natural deduction system. In particular, we show how to treat presupposition projection and accommodation within our proof-theoretic framework.

1 Introduction

In this paper we propose a proof-theoretic analysis of presuppositions in natural language, focusing on the interpretation of definite descriptions. Our proposal is based on constructive type theory [11,12]. While an original aim of constructive type theory is to analyze informal reasoning in mathematics, it is also applicable to the realm of natural language semantics. In particular, Ranta [13] shows that it provides us with an alternative framework to the standard dynamic theories of discourse semantics such as DRT [9] and DPL [5]. A central problem here is how to formulate the translation procedure from natural language discourses into formal representations in constructive type theory. In particular, the problem is how to deal with the resolution process of presupposition and anaphora. Ranta [13] deals with this problem by adopting the 'generation'-based approach, where the translation proceeds from formulas in constructive type theory to natural language sentences. In what follows, we present a usual 'parsing'-based translation procedure. The main idea is to use the natural deduction system of intuitionistic ε-calculus introduced in Carlström [2], which serves as an intermediate language in the translation process from natural language into constructive type theory. A formula of ε-calculus plays the role of an *underspecified* representation of the content of an utterance. This makes it possible to separate the entire translation process into the deterministic and the non-deterministic parts.

The structure of this paper is as follows. In section 2, we review a proof-theoretic approach to the presupposition theory of definite descriptions, which was originally developed by Stenlund [14,15]. In section 3, we introduce ε-calculus, which was originally proposed in Carlström [2], and in section 4 we

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 214–227, 2008.

review how it is translated into constructive type theory. In section 5, we present a proof-theoretic analysis of presupposition resolution, restricting our attention to existential presuppositions triggered by definite descriptions. In particular, we explain how to treat presupposition projection and accommodation within our proof-theoretic framework. Comparisons with model-theoretic approaches including DRT [17] and the satisfaction theory of presuppositions [7,6] are also discussed.

2 Presupposition Theory of Definite Descriptions

In [14,15], Stenlund develops the presupposition theory of definite descriptions within a proof-theoretic framework. It consists of two ideas: (i) a definite description is a referring expression (not a quantificational device as in the Russellian analysis), and (ii) it can be properly used under the *assumption* that it refers to an individual. The comparison with traditional approaches is useful here. Consider the following classical example involving an empty description:

(1) The present king of France is bald.

On the Russellian analysis, (1) is simply false. On the presuppositional analysis due to Frege and Strawson, (1) fails to express a proposition. On Stenlund's proof-theoretic analysis, which can be regarded as a refinement of the Frege-Strawson analysis, (1) can be used to express a proposition under the assumption that there is a king of France. On this account, a sentence involving an empty description can express a proposition simply because we can use any assumption we want in the course of an inference. As we will see later, the notion of assumption is crucial when we deal with the projection problem for presuppositions.

In [14], Stenlund introduces the natural deduction system of classical first order logic with $\imath$-operator, where $\imath x A x$ is intended to capture the use of 'the A' in mathematical reasoning. In [15], Stenlund also introduces an intuitionistic version of the system. In [2], Carlström presents a similar natural deduction system of intuitionistic first order logic with ε-operator, and shows how it can be translated into constructive type theory. Interestingly, Carlström suggests that not only indefinite descriptions, but also definite descriptions in ordinary discourse can be represented by ε-terms. This is consistent with the claim in the philosophy and linguistics literature that uniqueness implications associated with definite descriptions should be derived pragmatically (cf. [16]). Following Carlström's suggestion, we will analyze a definite description 'the F' as '$\varepsilon x\, F(x)$'. As we will see later, ε-terms can be naturally interpreted in terms of existential quantification in constructive type theory.

Before moving on, let us make a cautionary note. We take it that the presuppositional analysis adopted here applies only to a definite description occurring in the argument position of a predicate, as in (1). It is widely observed in the literature that there are at least two types of constructions in which a definite description has a 'non-referential' use. One is the case in which a definite description occurs in the predicative position as in (2), and the other is the case in which a definite description occurs in an existential construction as in (3).

(2) John is the king of France.
(3) The king of France does not exist.

We take it that the presuppositional analysis does not apply to these constructions. Thus, it is important to recognize that ε-term employed here does not provide a translation of a definite description *per se*; rather, it comes from a particular role a definite description plays in the argument position of a predicate.

3 ε-Calculus

In this section we introduce the natural deduction system of ε-calculus in [2]. In the next section, we briefly review how it can be translated into constructive type theory. In order to formalize the presupposition theory of definite descriptions introduced above, we need to formally represent an inference like (4).

$$(4)\qquad \cfrac{\cfrac{\text{`There is a king of France' is true}}{\text{`The king of France' refers to an individual}}}{\text{`The king of France is bald' expresses a proposition}}$$

These statements are called 'judgement', which are usually taken to be implicit in the meta-language of a logical system. Here we need three kinds of judgements. We write 'A : **true**' for 'a formula A is true', 'A : **prop**' for 'a formula A expresses a proposition', and 'a : **ind**' for 'a term a refers to an individual'. Thus the informal derivation in (4) can be represented as in (5).

$$(5)\qquad \cfrac{\cfrac{\exists x\,\mathbf{king}(x,\mathbf{france}) : \mathbf{true}}{\varepsilon x\,\mathbf{king}(x,\mathbf{france}) : \mathbf{ind}}}{\mathbf{bald}(\varepsilon x\,\mathbf{king}(x,\mathbf{france})) : \mathbf{prop}}$$

Here we analyze 'the king of France' as '$\varepsilon x\,\mathbf{king}(x,\mathbf{france})$', assuming that 'king' is a relational noun and that '$\mathbf{king}(x, y)$' reads as 'x is a king of y'.

The natural deduction system of ε-calculus consists of three kinds of rules: formation rule (of individual-denoting terms and propositions), introduction rule, and elimination rule. The formation rules are the following: [1]

$$\frac{t_1 : \mathbf{ind} \quad \cdots \quad t_n : \mathbf{ind}}{P(t_1,\ldots,t_n) : \mathbf{prop}}\,PF \qquad \frac{}{\bot : \mathbf{prop}}\,\bot F \qquad \frac{\exists x A(x) : \mathbf{true}}{\varepsilon x A(x) : \mathbf{ind}}\,\varepsilon F$$

$$\frac{A : \mathbf{prop} \quad \begin{matrix}[A : \mathbf{true}]\\ \vdots \\ B : \mathbf{prop}\end{matrix}}{A \wedge B : \mathbf{prop}}\,\wedge F \qquad \frac{A : \mathbf{prop} \quad \begin{matrix}[A : \mathbf{true}]\\ \vdots \\ B : \mathbf{prop}\end{matrix}}{A \supset B : \mathbf{prop}}\,\supset F$$

[1] The rules for disjunction are omitted here. For the full treatment of this system, readers should consult [2].

$$\dfrac{\begin{array}{c}[x:\mathbf{ind}]\\ \vdots\\ A(x):\mathbf{prop}\end{array}}{\forall x A(x):\mathbf{prop}}\ \forall F \qquad \dfrac{\begin{array}{c}[x:\mathbf{ind}]\\ \vdots\\ A(x):\mathbf{prop}\end{array}}{\exists x A(x):\mathbf{prop}}\ \exists F$$

The rule PF is a schematic rule for each n-ary predicate. The negation '$\neg A$' is defined to be the formula '$A \supset \bot$'.

The introduction and elimination rules are the following:

$$\dfrac{A:\mathbf{true} \quad B:\mathbf{true}}{A \wedge B:\mathbf{true}}\ \wedge I \qquad \dfrac{A \wedge B:\mathbf{true}}{A:\mathbf{true}}\ \wedge El \qquad \dfrac{A \wedge B:\mathbf{true}}{B:\mathbf{true}}\ \wedge Er$$

$$\dfrac{A:\mathbf{prop} \quad \bot:\mathbf{true}}{A:\mathbf{true}}\ \bot E \qquad \dfrac{\exists x A(x):\mathbf{true}}{A(\varepsilon x A(x)):\mathbf{true}}\ \varepsilon I$$

$$\dfrac{A:\mathbf{true} \quad A \supset B:\mathbf{true}}{B:\mathbf{true}}\supset E \qquad \dfrac{A:\mathbf{prop} \quad \begin{array}{c}[A:\mathbf{true}]\\ \vdots\\ B:\mathbf{true}\end{array}}{A \supset B:\mathbf{true}}\supset I$$

$$\dfrac{\begin{array}{c}[x:\mathbf{ind}]\\ \vdots\\ A(x):\mathbf{true}\end{array}}{\forall x A(x):\mathbf{true}}\ \forall I \qquad \dfrac{\begin{array}{c}[x:\mathbf{ind}]\\ \vdots\\ A(x):\mathbf{prop}\end{array} \quad t:\mathbf{ind} \quad A(t):\mathbf{true}}{\exists x A(x):\mathbf{true}}\ \exists I$$

$$\dfrac{\forall x A(x):\mathbf{true} \quad t:\mathbf{ind}}{A(t):\mathbf{true}}\ \forall E \qquad \dfrac{\exists x A(x):\mathbf{true} \quad \begin{array}{c}[x:\mathbf{ind}][A(x):\mathbf{true}]\\ \vdots\\ B:\mathbf{true}\end{array}}{B:\mathbf{true}}\ \exists E$$

We assume that the usual variable restrictions apply to the rules that discharge an assumption of the form 'x : **ind**'. We also assume that the special restrictions apply to $\supset E$ and $\exists I$ as stated in [2].

4 Constructive Type Theory

The basic idea of constructive type theory is to understand the *truth* of a proposition in terms of the existence of a *proof*, where a proof is taken to be an individual. Thus, the judgement of the form 'A : **true**' is replaced by '$p : A$', which reads as 'a proposition A has a proof p'. Here, following the 'propositions-as-sets' interpretation (Curry-Howard Correspondence), the meaning of a proposition is identified with the set of proofs (evidences) of that proposition. In particular, the meaning of an existential proposition is given by a pair consisting of an individual (witness) and a proof of the corresponding proposition. This is captured by the following introduction and elimination rules.

$$\frac{\begin{array}{c}[x:A]\\ \vdots\\ B(x):\mathbf{prop}\end{array}\quad a:A\quad b:B(a)}{(a,b):(\exists x:A)\,B(x)}\ \exists I$$

$$\frac{c:(\exists x:A)\,B(x)}{l\,c:A}\ \exists El \qquad \frac{c:(\exists x:A)\,B(x)}{r\,c:B(l\,c)}\ \exists Er$$

Here, l and r denote the left and right projections, which are computed in the following way:

$$l(a,b)\longrightarrow a \qquad r(a,b)\longrightarrow b.$$

The formation rules for existential and universal quantification are the following:

$$\frac{A:\mathbf{prop}\quad \begin{array}{c}[x:A]\\ \vdots\\ B(x):\mathbf{prop}\end{array}}{(\exists x:A)B(x):\mathbf{prop}}\ \exists F \qquad \frac{A:\mathbf{prop}\quad \begin{array}{c}[x:A]\\ \vdots\\ B(x):\mathbf{prop}\end{array}}{(\forall x:A)B(x):\mathbf{prop}}\ \forall F$$

The introduction and elimination rules for universal quantification are the following:

$$\frac{\begin{array}{c}[x:A]\\ \vdots\\ b(x):B(x)\end{array}}{\lambda x b(x):(\forall x:A)B(x)}\ \forall I \qquad \frac{b:(\forall x:A)B(x)\quad a:A}{\mathrm{apply}(b,a):B(a)}\ \forall E$$

The corresponding computation rule is:

$$\mathrm{apply}(\lambda x b(x),a)\longrightarrow b(a).$$

We define $A\wedge B$ as $(\exists x:A)B(x)$ and $A\supset B$ as $(\forall x:A)B(x)$, when B does not depend on x.

The translation of ε-calculus into constructive type theory can be done as described in [2]. In particular, $\wedge F$ and $\supset F$ in ε-calculus are respectively interpreted by $\exists F$ and $\forall F$ in constructive type theory. The epsilon rules εF and εI are interpreted by $\exists El$ and $\exists Er$ respectively, following the suggestion of Martin-Löf [11]. As an illustration, consider the derivation in ε-calculus given in (5) above. The derivation in (5), shown to the left below, can be translated into constructive type theory as shown to the right.

$$\frac{\dfrac{\exists x\,\mathbf{king}(x,\mathbf{france}):\mathbf{true}}{\varepsilon x\,\mathbf{king}(x,\mathbf{france}):\mathbf{ind}}\ \varepsilon F}{\mathbf{bald}(\varepsilon x\,\mathbf{king}(x,\mathbf{france})):\mathbf{prop}}\ PF \quad\Longrightarrow\quad \frac{c:(\exists x:\mathbf{ind})\,\mathbf{king}(x,\mathbf{france})}{\dfrac{lc:\mathbf{ind}}{\mathbf{bald}(lc):\mathbf{prop}}\ PF}\ \exists El$$

It is important to recognize that the translation of a proposition A into constructive type theory is determined by a *derivation* of 'A : **prop**' in ε-calculus, not by the proposition A itself.

5 Linguistic Applications

5.1 Presupposition Projection

From the linguistic point of view, a formula A in ε-calculus can be taken as an *underspecified* representation of the content of an utterance, which is disambiguated in the derivation of '$A : \mathbf{prop}$'. Here, the process of finding a derivation of '$A : \mathbf{prop}$' can be understood as the process of presupposition resolution. The point of the translation into constructive type theory is to make the anaphoric dependencies explicit. Thus, we have the following equations:

– to understand an utterance of a sentence S, i.e., to specify the content of an utterance of S (through presupposition resolution) = to search for a derivation of '$A : \mathbf{prop}$'
– what is presupposed by an utterance of S = open assumptions in a derivation of '$A : \mathbf{prop}$'
– what is asserted by an utterance of S = the judgement '$A : \mathbf{true}$'
– to evaluate the content of an utterance of S = to search for a derivation of '$A : \mathbf{true}$'

where A is a formula in ε-calculus which formalizes the sentence S.

A task of specifying and evaluating the content of an utterance is done in a particular context. Such a context can be represented by a sequence of judgements. These judgements work as local axioms in a derivation. Thus, the hearer's task of understanding an utterance of a sentence S in a context Γ can be taken as a process of searching for a derivation of '$A : \mathbf{prop}$' in the bottom-up way under the context Γ, where the process starts with the judgement '$A : \mathbf{prop}$', and stops when all assumptions upon which '$A : \mathbf{prop}$' depends are closed or contained in Γ.

As an illustration, consider a simple example in which presupposition projection is blocked.

(6) If John has a wife, John's wife is happy.

This sentence is mapped into the formula of ε-calculus '$\exists x\, \mathbf{wife}(x, \mathbf{john}) \supset \mathbf{happy}(\varepsilon x\, \mathbf{wife}(x, \mathbf{john}))$' by a usual kind of semantic decoding procedure, with which we are not concerned in this paper. Then, it can be proved to be a proposition in the following way.

$$\dfrac{\dfrac{\dfrac{[x : \mathbf{ind}]^1 \qquad \mathbf{john} : \mathbf{ind}}{\mathbf{wife}(x, \mathbf{john}) : \mathbf{prop}}\,PF}{\exists x\, \mathbf{wife}(x, \mathbf{john}) : \mathbf{prop}}\,\exists F,\, 1 \qquad \dfrac{\dfrac{[\exists x\, \mathbf{wife}(x, \mathbf{john}) : \mathbf{true}]^2}{\varepsilon x\, \mathbf{wife}(x, \mathbf{john}) : \mathbf{ind}}\,\varepsilon F}{\mathbf{happy}(\varepsilon x\, \mathbf{wife}(x, \mathbf{john})) : \mathbf{prop}}\,PF}{\exists x\, \mathbf{wife}(x, \mathbf{john}) \supset \mathbf{happy}(\varepsilon x\, \mathbf{wife}(x, \mathbf{john})) : \mathbf{prop}}\,\supset F,\, 2$$

By translating the derivation into the one in constructive type theory, as shown below, the anaphoric link between the individual introduced in the antecedent and the one referred to in the consequent can be made explicit.

$$\dfrac{\dfrac{\dfrac{[x:\mathbf{ind}]^1 \qquad \mathbf{john}:\mathbf{ind}}{\mathbf{wife}(x,\mathbf{john}):\mathbf{prop}}\,PF}{(\exists x:\mathbf{ind})\,\mathbf{wife}(x,\mathbf{john}):\mathbf{prop}}\,\exists F,\,1 \qquad \dfrac{\dfrac{[y:(\exists x:\mathbf{ind})\,\mathbf{wife}(x,\mathbf{john})]^2}{ly:\mathbf{ind}}\,\exists El}{\mathbf{happy}(ly):\mathbf{prop}}\,PF}{(\forall y:(\exists x:\mathbf{ind})\,\mathbf{wife}(x,\mathbf{john}))\,\mathbf{happy}(ly):\mathbf{prop}}\,\forall F,\,2$$

In this derivation, the assumption '**john** : **ind**' remains open. This is consistent with the intuition that the utterance of (6) presupposes that the domain of discourse contains an individual named 'John'. In this example, we can see that the translation process is divided into the deterministic and the non-deterministic parts as summarized in (7):

(7) a. If John has a wife, John's wife is happy.
   ↓ Semantic decoding : deterministic
 b. $\exists x\,\mathbf{wife}(x,\mathbf{john}) \supset \mathbf{happy}(\varepsilon x\,\mathbf{wife}(x,\mathbf{john}))$
   ↓ Presupposition resolution : non-deterministic
 c. A derivation of '$\exists x\,\mathbf{wife}(x,\mathbf{john}) \supset \mathbf{happy}(\varepsilon x\,\mathbf{wife}(x,\mathbf{john})) : \mathbf{prop}$'
   ↓ Translation into constructive type theory : deterministic
 d. $(\forall y:(\exists x:\mathbf{ind})\,\mathbf{wife}(x,\mathbf{john}))\,\mathbf{happy}(ly)$

For another example, consider the following:

(8) If John is married, John's wife is happy.

A derivation in which the existential presupposition triggered by 'John's wife' does not project must look as follows:

$$\dfrac{\dfrac{\mathbf{john}:\mathbf{ind}}{\mathbf{married}(\mathbf{john}):\mathbf{prop}}\,PF \qquad \dfrac{\dfrac{\begin{matrix}[\mathbf{married}(\mathbf{john}):\mathbf{true}]^1\\ \vdots\ D\\ \exists x\,\mathbf{wife}(x,\mathbf{john}):\mathbf{true}\end{matrix}}{\varepsilon x\,\mathbf{wife}(x,\mathbf{john}):\mathbf{ind}}\,\varepsilon F}{\mathbf{happy}(\varepsilon x\,\mathbf{wife}(x,\mathbf{john})):\mathbf{prop}}\,PF}{\mathbf{married}(\mathbf{john}) \supset \mathbf{happy}(\varepsilon x\,\mathbf{wife}(x,\mathbf{john})):\mathbf{prop}}\,\supset F,\,1$$

Thus, it is predicted that the utterance of (8) presupposes that there is a derivation D from the judgement '**married**(**john**) : **true**' to the judgement '$\exists x\,\mathbf{wife}(x,\mathbf{john}) : \mathbf{true}$'. One straightforward way to fill the gap D is to make use of an additional judgement '$\mathbf{married}(\mathbf{john}) \supset \exists x\,\mathbf{wife}(x,\mathbf{john}) : \mathbf{true}$'. This can be further derived by using a general judgement '$\forall y\,(\mathbf{married}(y) \supset \exists x\,\mathbf{wife}(x,y)) : \mathbf{true}$', which can be assumed to be within the hearer's constant knowledge. The relevant derivation is the following.

$$\dfrac{\mathbf{married}(\mathbf{john}):\mathbf{true} \qquad \dfrac{\forall y\,(\mathbf{married}(y) \supset \exists x\,\mathbf{wife}(x,y)):\mathbf{true} \qquad \mathbf{john}:\mathbf{ind}}{\mathbf{married}(\mathbf{john}) \supset \exists x\,\mathbf{wife}(x,\mathbf{john}):\mathbf{true}}\,\forall E}{\exists x\,\mathbf{wife}(x,\mathbf{john}):\mathbf{true}}\,\supset E$$

Thus, in the overall derivation, two assumptions, namely, '$\forall y(\mathbf{married}(y) \supset \exists x\, \mathbf{wife}(x, y)) : \mathbf{true}$' and '$\mathbf{john} : \mathbf{ind}$', remain open, and hence they are predicted to be presuppositions of (8).

The treatment of the cases like (8) shows an essential difference between the proof-theoretic approach adopted here and the model-theoretic approach in discourse representation theory (DRT, [17]). The problem inherent to DRT is that it lacks inferential rules, so that the process of ordinary inference and the process of presupposition computation are treated at the different levels of the system (meta-language and object-language, respectively). This separation makes it difficult to handle examples like (8) in DRT, as many writers point out (e.g. [1]). In our proof-theoretic framework, both processes can be represented at the same level by making explicit the distinction between propositions and judgements, and the interaction between the two processes can be captured in a formal derivation.

It should be also noted that there is an important difference between the current proof-theoretic approach and the satisfaction theory of presuppositions (Karttunen [7], Heim [6]). Although there is a similarity in their treatment of presupposition projection of conjunction and implication (cf. Fernando [3]), these two theories make different predictions. In the satisfaction theory, the presuppositions of conjunction and implication are predicted as in (9) and (10), where we denote the presupposition of a sentence A by '$\mathbf{ps}(A)$'.

(9) $\mathbf{ps}(A \wedge B) = \mathbf{ps}(A) \wedge (A \supset \mathbf{ps}(B))$
(10) $\mathbf{ps}(A \supset B) = \mathbf{ps}(A) \wedge (A \supset \mathbf{ps}(B))$

Thus, the satisfaction theory predicts that (11) presupposes (12) rather than 'John has a wife'.

(11) If John works hard, his wife is happy.
(12) If John works hard, he has a wife.

As Geurts [4] argues, this prediction is too weak and pragmatic repairs proposed in the literature (e.g. Karttunen and Peters [8]) do not yield correct predictions. By contrast, the formation rules for conjunction and implication in our proof-theoretic setting, repeated here, can make correct predictions.

$$\frac{A : \mathbf{prop} \quad \begin{matrix}[A : \mathbf{true}] \\ \vdots \\ B : \mathbf{prop}\end{matrix}}{A \wedge B : \mathbf{prop}} \wedge F \qquad \frac{A : \mathbf{prop} \quad \begin{matrix}[A : \mathbf{true}] \\ \vdots \\ B : \mathbf{prop}\end{matrix}}{A \supset B : \mathbf{prop}} \supset F$$

This is simply because the judgement '$B : \mathbf{prop}$' may or may not depend on the assumption '$A : \mathbf{true}$'. The prediction of the satisfaction theory in (9) and (10) lacks this flexibility. [2]

[2] For a recent defence of the satisfaction theory, see van Rooij [18].

5.2 Embedded Descriptions

The treatment of cases in which one definite description is embedded within another, as in (13), is somewhat more complicated.

(13) John's aunt's cousin disappeared.

This sentence presupposes that John has an aunt and that John's aunt has a cousin. In order to account for this fact, the embedded description 'John's aunt' needs to be analyzed as '$\varepsilon y\, (\exists x\, \mathbf{cousin}(x, y) \wedge \mathbf{aunt}(y, \mathbf{john}))$'. Let us abbreviate this term as 't'. Thus, (13) is represented as (14).

(14) $\mathbf{disappeared}(\varepsilon x\, \mathbf{cousin}(x, t))$

Then, we have the following derivation, which yields the desired result.

$$\dfrac{\dfrac{\dfrac{\exists y\, (\exists x\, \mathbf{cousin}(x, y) \wedge \mathbf{aunt}(y, \mathbf{john})) : \mathbf{true}}{\exists x\, \mathbf{cousin}(x, t) \wedge \mathbf{aunt}(t, \mathbf{john}) : \mathbf{true}}\ \varepsilon I}{\dfrac{\exists x\, \mathbf{cousin}(x, t) : \mathbf{true}}{\varepsilon x\, \mathbf{cousin}(x, t) : \mathbf{ind}}\ \varepsilon F}\ \wedge E}{\mathbf{disappeared}(\varepsilon x\, \mathbf{cousin}(x, t)) : \mathbf{prop}}\ PF$$

Let us consider how to deal with a more complex example involving a pronoun. Consider:

(15) He disappeared.
(16) His aunt disappeared.
(17) His aunt's cousin disappeared.

We take it that a personal pronoun refers to an individual satisfying the condition associated with its gender feature. Thus a pronoun is also analyzed as an ε-term. This enables us to treat anaphora resolution as a special case of presupposition resolution, in a similar way to the well-known analysis within the framework of DRT [17]. Given the presuppositional analysis of pronouns, 'he' in (15) is represented as '$\varepsilon x\, \mathbf{male}(x)$'. (16) is treated in the same way as (13): 'His aunt' is analyzed as '$\varepsilon x\, \mathbf{aunt}(x, \varepsilon y\, (\exists x\, \mathbf{aunt}(x, y) \wedge \mathbf{male}(y)))$'. For (17), starting with the most deeply embedded one, we can analyze each description (or pronoun) in the following way:

'his' $= \varepsilon z\, (\exists y\, (\exists x\, \mathbf{cousin}(x, y) \wedge \mathbf{aunt}(y, z)) \wedge \mathbf{male}(z))$
'his aunt' $= \varepsilon y\, (\exists x\, \mathbf{cousin}(x, y) \wedge \mathbf{aunt}(y,$ 'his'$))$
'his aunt's cousin' $= \varepsilon x\, \mathbf{cousin}(x,$ 'his aunt'$)$

Under this analysis, it can be shown that (17) presupposes the following judgement:

$$\exists z\, (\exists y\, (\exists x\, \mathbf{cousin}(x, y) \wedge \mathbf{aunt}(y, z)) \wedge \mathbf{male}(z)) : \mathbf{true}.$$

5.3 Informative Presuppositions

Sometimes the use of a presupposing sentence introduces novel information into a discourse, in particular when the presupposed information is uncontroversial. Since the seminal works of Lewis [10] and Heim [6], such a case of *informative presupposition* is analyzed in terms of a mechanism called *accommodation*. In order to deal with informative presuppositions within our framework, we introduce the following operation:

(18) Given a derivation in which a judgement of the form 'B : **prop**' depends on an open assumption of the form 'A : **true**', we may transform the derivation into the one in which 'A : **prop**' is located at some stage in the derivation so that the assumption 'A : **true**' is discharged.

We call this operation simply 'accommodation'. In our terms, accommodation is a process of adjusting the *content*, rather than the *context*, of an utterance. From the hearer's point of view, to perform accommodation is to add an implicitly understood proposition at some stage in the course of a derivation. The process of accommodation as stated in (18) allows us the following transformation:

$$\begin{array}{c} A : \mathbf{true} \\ \vdots \\ B : \mathbf{prop} \end{array} \quad \Longrightarrow \quad \dfrac{A : \mathbf{prop} \qquad \begin{array}{c}[A : \mathbf{true}] \\ \vdots \\ B : \mathbf{prop}\end{array}}{A \wedge B : \mathbf{prop}} \wedge F$$

Using this transformation, we can formulate within our framework the process of the so-called 'global' and 'local' accommodation adopted in standard dynamic theories of presuppositions (cf. [6,17]). Let us illustrate with the classical example (19) how the transformation works.

(19) The king of France is not bald.

(19) is formalized as (20).

(20) $\neg\, \mathbf{bald}(\varepsilon x\, \mathbf{king}(x, \mathbf{france}))$

We can then derive the following:

$$\dfrac{\dfrac{\dfrac{\exists x\, \mathbf{king}(x, \mathbf{france}) : \mathbf{true}}{\varepsilon x\, \mathbf{king}(x, \mathbf{france}) : \mathbf{ind}} \varepsilon F}{\mathbf{bald}(\varepsilon x\, \mathbf{king}(x, \mathbf{france})) : \mathbf{prop}} PF}{\neg\, \mathbf{bald}(\varepsilon x\, \mathbf{king}(x, \mathbf{france})) : \mathbf{prop}} \neg F$$

where the rule $\neg F$ is a derived rule which can be obtained by using $\supset F$ and $\bot F$. Given this derivation, a hearer can proceed in at least three different ways.

First, suppose that the hearer believes that someone, say John, is the king of France. This is to say, the context Γ contains three judgements:

$$\mathbf{john} : \mathbf{ind},\ \mathbf{france} : \mathbf{ind},\ \text{and}\ \mathbf{king}\,(\mathbf{john}, \mathbf{france}) : \mathbf{true}.$$

In this context, the presupposition would be simply satisfied. This reading can be captured in the following derivation.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[x:\mathbf{ind}]^1 \quad \mathbf{france}:\mathbf{ind}}{\mathbf{king}(x,\mathbf{france}):\mathbf{prop}}\,PF \quad \mathbf{john}:\mathbf{ind} \quad \mathbf{king}(\mathbf{john},\mathbf{france}):\mathbf{true}}{\exists x\,\mathbf{king}(x,\mathbf{france}):\mathbf{true}}\,\exists I,{}_1}{\varepsilon x\,\mathbf{king}(x,\mathbf{france}):\mathbf{ind}}\,\varepsilon F}{\mathbf{bald}(\varepsilon x\,\mathbf{king}(x,\mathbf{france})):\mathbf{prop}}\,PF}{\neg\,\mathbf{bald}(\varepsilon x\,\mathbf{king}(x,\mathbf{france})):\mathbf{prop}}\,\neg F$$

On this reading, '$\varepsilon x\,\mathbf{king}(x,\mathbf{france})$' is interpreted as '$l\,(\mathbf{john},p) : \mathbf{ind}$' in constructive type theory, given the assumption that '$\mathbf{john} : \mathbf{ind}$' and '$p : \mathbf{king}(\mathbf{john},\mathbf{france})$'. Here '$l\,(\mathbf{john},p)$' reduces to '$\mathbf{john}$' by left projection. Accordingly, we can obtain the judgement '$\neg\,\mathbf{bald}(\mathbf{john})$'. This means that the utterance of (19) in this context can communicate the proposition that John is not bald.

The second and third options involve accommodation. Consider the context in which the hearer does not know whether or not there is a king of France. In this context, the hearer might perform accommodation, which transforms the original derivation into the following:

$$\dfrac{\dfrac{\dfrac{[x:\mathbf{ind}]^1 \quad \mathbf{france}:\mathbf{ind}}{\mathbf{king}(x,\mathbf{france}):\mathbf{prop}}\,PF}{\exists x\,\mathbf{king}(x,\mathbf{france}):\mathbf{prop}}\,\exists F,{}_1 \quad \dfrac{\dfrac{\dfrac{[\exists x\,\mathbf{king}(x,\mathbf{france}):\mathbf{true}]^2}{\varepsilon x\,\mathbf{king}(x,\mathbf{france}):\mathbf{ind}}\,\varepsilon F}{\mathbf{bald}(\varepsilon x\,\mathbf{king}(x,\mathbf{france})):\mathbf{prop}}\,PF}{\neg\,\mathbf{bald}(\varepsilon x\,\mathbf{king}(x,\mathbf{france})):\mathbf{prop}}\,\neg F}{\exists x\,\mathbf{king}(x,\mathbf{france}) \wedge \neg\,\mathbf{bald}(\varepsilon x\,\mathbf{king}(x,\mathbf{france})):\mathbf{prop}}\,\wedge F,{}_2$$

The translation into constructive type theory yields:

$$(\exists y : (\exists x : \mathbf{ind})\,\mathbf{king}(x,\mathbf{france}))\,\neg\,\mathbf{bald}\,(ly) : \mathbf{prop}.$$

This corresponds to the 'global' accommodation reading in Heim's analysis [6].

The third option is this. Consider the context in which the hearer knows that there is no king of France. Given this context, in order to avoid a contradiction, the hearer must perform accommodation at 'one stage above' in the original derivation. This yields the following result:

$$\dfrac{\dfrac{\dfrac{\dfrac{[x:\mathbf{ind}]^1 \quad \mathbf{france}:\mathbf{ind}}{\mathbf{king}(x,\mathbf{france}):\mathbf{prop}}\,PF}{\exists x\,\mathbf{king}(x,\mathbf{france}):\mathbf{prop}}\,\exists F,{}_1 \quad \dfrac{\dfrac{[\exists x\,\mathbf{king}(x,\mathbf{france}):\mathbf{true}]^2}{\varepsilon x\,\mathbf{king}(x,\mathbf{france}):\mathbf{ind}}\,\varepsilon F}{\mathbf{bald}(\varepsilon x\,\mathbf{king}(x,\mathbf{france})):\mathbf{prop}}\,PF}{\exists x\,\mathbf{king}(x,\mathbf{france}) \wedge \mathbf{bald}(\varepsilon x\,\mathbf{king}(x,\mathbf{france})):\mathbf{prop}}\,\wedge F,{}_2}{\neg\,(\exists x\,\mathbf{king}(x,\mathbf{france}) \wedge \mathbf{bald}(\varepsilon x\,\mathbf{king}(x,\mathbf{france}))):\mathbf{prop}}\,\neg F$$

This corresponds to the 'local' accommodation reading in Heim's analysis.

Now consider the following type of construction involving existential quantification, which is much discussed in the literature.

(21) Some boy loves his cat.

(21) is formalized as

(22) $\exists x\,(\mathbf{boy}(x) \wedge \mathbf{love}(x, \varepsilon y\,\mathbf{cat}(y, x)))$,

where we assume that the embedded pronoun 'his' is bound to the subject 'some boy' and that for simplicity 'cat' is a two-place predicate '$\mathbf{cat}(x, y)$' which reads as 'x is a cat y owns'. We further assume here that for simplicity 'his cat' is represented as '$\varepsilon y\,\mathbf{cat}(y, x)$', rather than '$\varepsilon y\,\mathbf{cat}(y, \varepsilon x\,\exists\, y(\mathbf{cat}(y, x) \wedge \mathbf{male}(x)))$'. Given the representation in (22), our theory predicts that (21) has two distinct readings:

(i) It is presupposed that there is a derivation from the judgement '$x : \mathbf{ind}$' (or two judgements '$x : \mathbf{ind}$' and '$\mathbf{boy}(x) : \mathbf{true}$') to the judgement '$\exists y\,\mathbf{cat}(x, y) : \mathbf{true}$'.
(ii) By way of accommodation the proposition '$\exists x\,(\mathbf{boy}(x) \wedge \exists y\,(\mathbf{cat}(y, x) \wedge \mathbf{love}(x, y)))$' is derived.

The second option is similar to the prediction of Heim's analysis in [6].

We can also introduce an operation analogous to 'intermediate accommodation' proposed in [17], which can be regarded as a special case of the general procedure in (18). It allows us the following transformation.

$$\dfrac{A : \mathbf{prop} \quad \begin{matrix} B : \mathbf{true} \\ \vdots \\ C : \mathbf{prop} \end{matrix}}{A \supset C : \mathbf{prop}} \supset F \quad \Longrightarrow \quad \dfrac{\dfrac{A : \mathbf{prop} \quad B : \mathbf{prop}}{A \wedge B : \mathbf{prop}} \wedge F \qquad \begin{matrix} \dfrac{[A \wedge B : \mathbf{true}]}{B : \mathbf{true}} \wedge E \\ \vdots \\ C : \mathbf{prop} \end{matrix}}{(A \wedge B) \supset C : \mathbf{prop}} \supset F$$

To see how this works, consider the following example discussed in [17].

(23) Every boy loves his cat.

(23) is formalized as

(24) $\forall x\,(\mathbf{boy}(x) \supset \mathbf{love}(x, \varepsilon y\,\mathbf{cat}(y, x)))$,

where we assume that the pronoun 'his' is bound to the subject 'every boy' and that for simplicity 'his cat' is represented as '$\varepsilon y\,\mathbf{cat}(y, x)$' in a similar way as (22). We start with the following derivation, where '$\exists y\,\mathbf{cat}(y, x) : \mathbf{true}$' remains open.

$$\dfrac{\dfrac{\dfrac{[x : \mathbf{ind}]^1}{\mathbf{boy}(x) : \mathbf{prop}} \qquad \dfrac{[x : \mathbf{ind}]^1 \quad \dfrac{\exists y\,\mathbf{cat}(y, x) : \mathbf{true}}{\varepsilon y\,\mathbf{cat}(y, x) : \mathbf{ind}}}{\mathbf{love}(x, \varepsilon y\,\mathbf{cat}(y, x)) : \mathbf{prop}}}{\mathbf{boy}(x) \supset \mathbf{love}(x, \varepsilon y\,\mathbf{cat}(y, x)) : \mathbf{prop}} \supset F}{\forall x(\mathbf{boy}(x) \supset \mathbf{love}(x, \varepsilon y\,\mathbf{cat}(y, x))) : \mathbf{prop}} \forall F,\ 1$$

Restricting our attention to the options of accommodation, we have four possibilities. First, the ‘intermediate’ accommodation we have just introduced yields the following derivation.

$$
\dfrac{
  \dfrac{
    \dfrac{[x:\mathbf{ind}]^1}{\mathbf{boy}(x):\mathbf{prop}}
    \qquad
    \dfrac{\dfrac{[y:\mathbf{ind}]^2 \quad [x:\mathbf{ind}]^1}{\mathbf{cat}(y,x):\mathbf{prop}}}{\exists y\,\mathbf{cat}(y,x):\mathbf{prop}}\;2
  }{\mathbf{boy}(x)\wedge\exists y\,\mathbf{cat}(y,x):\mathbf{prop}}
  \qquad
  \dfrac{[x:\mathbf{ind}]^1 \quad \dfrac{\dfrac{[\mathbf{boy}(x)\wedge\exists y\,\mathbf{cat}(y,x):\mathbf{true}]^3}{\exists y\,\mathbf{cat}(y,x):\mathbf{true}}}{\varepsilon y\,\mathbf{cat}(y,x):\mathbf{ind}}}{\mathbf{love}(x,\varepsilon y\,\mathbf{cat}(y,x)):\mathbf{prop}}
}{
  \dfrac{\mathbf{boy}(x)\wedge\exists y\,\mathbf{cat}(y,x)\supset\mathbf{love}(x,\varepsilon y\,\mathbf{cat}(y,x)):\mathbf{prop}}{\forall x(\mathbf{boy}(x)\wedge\exists y\,\mathbf{cat}(y,x)\supset\mathbf{love}(x,\varepsilon y\,\mathbf{cat}(y,x))):\mathbf{prop}}\;\forall F,\,1
}\;\supset F,\,3
$$

The resulting reading can be glossed as ‘Every boy who has a cat loves it’. Next, the ‘local’ accommodation yields the following derivation.

$$
\dfrac{
  \dfrac{
    \dfrac{[x:\mathbf{ind}]^1}{\mathbf{boy}(x):\mathbf{prop}}
    \qquad
    \dfrac{
      \dfrac{\dfrac{[y:\mathbf{ind}]^2 \quad [x:\mathbf{ind}]^1}{\mathbf{cat}(y,x):\mathbf{prop}}}{\exists y\,\mathbf{cat}(y,x):\mathbf{prop}}\;2
      \qquad
      \dfrac{[x:\mathbf{ind}]^1 \quad \dfrac{[\exists y\,\mathbf{cat}(y,x):\mathbf{true}]^3}{\varepsilon y\,\mathbf{cat}(y,x):\mathbf{ind}}}{\mathbf{love}(x,\varepsilon y\,\mathbf{cat}(y,x)):\mathbf{prop}}
    }{\exists y\,\mathbf{cat}(y,x)\wedge\mathbf{love}(x,\varepsilon y\,\mathbf{cat}(y,x)):\mathbf{prop}}\;\wedge F,\,3
  }{\mathbf{boy}(x)\supset\exists y\,\mathbf{cat}(y,x)\wedge\mathbf{love}(x,\varepsilon y\,\mathbf{cat}(y,x)):\mathbf{prop}}\;\supset F
}{\forall x(\mathbf{boy}(x)\supset\exists y\,\mathbf{cat}(y,x)\wedge\mathbf{love}(x,\varepsilon y\,\mathbf{cat}(y,x)):\mathbf{prop}}\;\forall F,\,1
$$

The result can be glossed as ‘Every boy has a cat and loves it’. Third, using ‘global’ accommodation, we obtain a derivation ending up with the following judgement.

(25) $\exists y\,\mathbf{cat}(y,x)\wedge\forall x(\mathbf{boy}(x)\wedge\mathbf{love}(x,\varepsilon y\,\mathbf{cat}(y,x))):\mathbf{prop}$

This option is ruled out, since the formula contains a free variable. These three options give the same results as van der Sandt’s analysis [17]. Finally, we can perform accommodation at one stage before the application of $\forall F$ in the original derivation. This yields a derivation ending up with the following judgement.

(26) $\forall x(\exists y\,\mathbf{cat}(y,x)\wedge(\mathbf{boy}(x)\supset\mathbf{love}(x,\varepsilon y\,\mathbf{cat}(y,x)))):\mathbf{prop}$

Here the information that every individual in the domain of discourse has a cat is accommodated.

6 Conclusion

We have shown how a proof-theoretic framework based on ε-calculus [2] and constructive type theory [11] accounts for existential presuppositions triggered by definite descriptions in natural language. In particular, we have shown how presupposition projection and accommodation are treated in the proof-theoretic framework. The applications to a wider range of presupposition triggers, such as factive verbs and additive particles, remain open for further investigation.

References

1. Beaver, D.: Presupposition and Assertion in Dynamic Semantics. CSLI Publications, Stanford (2001)
2. Carlström, J.: Interpreting Descriptions in Intensional Type Theory. Journal of Symbolic Logic 70(2), 488–514 (2005)
3. Fernando, T.: Conservative Generalized Quantifiers and Presupposition. Semantics and Linguistic Theory XI, 172–191 (2001)
4. Geurts, B.: Presuppositions and Pronouns. Elsevier, Amsterdam (1999)
5. Groenendijk, J., Stokhof, M.: Dynamic Predicate Logic. Linguistics and Philosophy 14, 39–100 (1991)
6. Heim, I.: On the Projection Problem for Presuppositions. In: Barlow, M., Flickinger, D., Wescoat, M. (eds.) Proceedings of WCCFL, vol. 2, pp. 114–125 (1983)
7. Karttunen, L.: Presupposition and Linguistic Context. Theoretical Linguistics 1, 181–193 (1974)
8. Karttunen, L., Peters, S.: Conventional Implicatures. In: Oh, C., Dineen, D. (eds.) Syntax and Semantics 11: Presupposition, pp. 1–56. Academic Press, New York (1979)
9. Kamp, H.: A Theory of Truth and Semantic Representation. In: Groenendijk, J., Janssen, T., Stokhof, M. (eds.) Formal Methods in the Study of Language. Mathematical Centre Tracts 136, Amsterdam, pp. 277–322 (1981)
10. Lewis, D.: Scorekeeping in a Language Game. Journal of Philosophical Logic 8, 339–359 (1979)
11. Martin-Löf, P.: Intuitionistic Type Theory. Notes by G. Sambin, Bibliopolis, Naples (1984)
12. Nordström, B., Petersson, K., Smith, J.: Programming in Martin-Löf's Type Theory. Oxford University Press, Oxford (1990)
13. Ranta, A.: Type-Theoretical Grammar. Oxford University Press, Oxford (1994)
14. Stenlund, S.: The Logic of Description and Existence. Philosophical Studies, Philosophical Society and the Department of Philosophy, University of Uppsala, Uppsala (1973)
15. Stenlund, S.: Descriptions in intuitionistic logic. In: Kanger, S. (ed.) Proceedings of the Third Scandinavian Logic Symposium, pp. 197–212 (1975)
16. Szabó, Z.G.: Descriptions and Uniqueness. Philosophical Studies 101, 29–57 (2000)
17. Van der Sandt, R.A.: Presupposition Projection as Anaphora Resolution. Journal of Semantics 9, 333–377 (1992)
18. Van Rooij, R.: Strengthening Conditional Presuppositions. Journal of Semantics 24, 289–304 (2007)

Meaning Games

Kôiti Hasida[1], Shun Shiramatsu[2],
Kazunori Komatani[2], Tetsuya Ogata[2], and Hiroshi G. Okuno[2]

[1] ITRI, National Institute of Advanced Industrial Science and Technology (AIST)
[2] Graduate School of Informatics, Kyoto University
hasida.k@aist.go.jp

Abstract. Communication can be accounted for in game-theoretic terms. The *meaning game* is proposed to formalize intentional communication in which the sender sends a message and the receiver attempts to infer its intended meaning. Using large Japanese and English corpora, the present paper demonstrates that *centering theory* is derived from a meaning game. This suggests that there are no language-specific rules on referential coherence. More generally speaking, language use seems to employ Pareto-optimal ESSs (evolutionarily stable strategies) of potentially very complex meaning games. There is still much to do before this complexity is elucidated in scientific terms, but game theory provides statistical and analytic means by which to advance the study on semantics and pragmatics of natural languages and other communication modalities.

1 Introduction

Linguistic communication essentially consists of cooperation among the conversation participants to extend their common beliefs. A rational individual prefers to cooperatively use language so as to communicate much information to her interlocutor. On the other hand, she prefers to simplify her utterances within the inferable range for her interlocutor. For instance, she may use a pronoun if she expects that her interlocutor can disambiguate its referent. Such characteristics of language are considered to have evolved and been optimized through social interactions. We conjecture that such a core aspect of communication in various languages can be accounted for in game-theoretic terms, probably as evolutionarily-stable strategies (ESSs), without recourse to language-specific stipulations.

Game theory is a mathematical theory about interaction among players (multiple autonomous agents) [1,2]. It formulates the players' rational behavior in various strategic interactions.

The *Prisoner's Dilemma* is a well-known strategic situation in which two players may cooperate with or betray each other. Table 1 is its payoff matrix. The values in this matrix represent players' benefits. In this game, betrayal is the individually optimal strategy of each player in response to any strategy of the other player: Each player gains more by betraying whether the other player

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 228–241, 2008.

Table 1. Prisoner's Dilemma

prisoner B \ prisoner A	cooperate	betray
cooperate	2\2	0\3
betray	3\0	1\1 (Nash equilibrium)

cooperates or betrays. Generally speaking, such a combination of individually optimal strategies is called a *Nash equilibrium*. That is, a Nash equilibrium is a combination of players' strategies in which no player is motivated to change her strategy as far as the other players keeps their strategies. The equilibrium in Prisoner's Dilemma, however, does not maximize the players' benefits in total. So individual optimization does not always conform to social optimization.

On the other hand, communication is an inherently collaborative situation. For instance, even enemies must cooperate in order to convey their animosities to each other. To formulate this, let i be the proposition that the sender S intends to communicate some semantic content c to the receiver R. Then i entails that S intends that R should both recognize c and believe i. This is the core of Grice's *nonnatural meaning*[1] [3,4]. This restricted sense of nonnatural meaning implies that communication is inherently collaborative, because both S and R want R to recognize c and i. S of course wants it, and so does R because it is generally beneficial to know what S intends to make R believe or obey.

The *meaning game* [5,6] is a game-theoretic framework to address this aspect of communication — nonnatural meaning in the above sense. To partly exemplify its efficacy, Hasida argues that it derives *centering theory* [7], which is a standard theory of *referential coherence* (i.e., the preference related to S's selection of anaphors and R's disambiguation of those). Below we outline Meaning Game and a game-theoretic account of centering theory based on it. Furthermore, we empirically verify this account of centering theory on the basis of large Japanese and English corpora.

2 Communication and Games

Communication has been discussed in game-theory literature. The following studies are relevant to the present paper.

2.1 Signaling Game

A *signaling game* consists of S's sending a message (a signal) to R and R's doing some action in response to that message. In this game, S should send a costly message to get a large payoff. Table 2 shows some typical cases of signaling

[1] Grice's original notion of nonnatural meaning further entails (S's intention of) R's believing (when c is a proposition or a reference) or obeying (when it is an order or a request) c, but we disregard this aspect here.

Table 2. Typical applications of the signaling game

	Job market	Mate selection
Sender S	Job seeker	Male deer
Receiver R	Employer	Female deer
Type T	Competence	Vitality
Message M	Education	Antler size
Action A	Hiring	Mating

game. In job market signaling [8], a job seeker S signals her competence (type) to a potential employer R with the level of her education as the message, and R decides the salary to offer her. A competent job seeker tend to be highly educated, and the potential employer will offer her a high salary. In mate selection [9], a male deer S indicates its strength to potential mate R by the size of its antlers. A strong deer will grow extra-large antlers to demonstrate its survival competence despite the handicap of keeping those practically useless antlers — the handicap principle [10].

2.2 Collaborative Disambiguation

One linguistic expression may have multiple meanings depending on multiple contexts. This efficiency of language [11] presupposes disambiguation of meanings conveyed by linguistic messages. Here, let m_1 and m_2 be linguistic messages and c_1 and c_2 be their content candidates. Suppose that both c_1 and c_2 can be referenced by m_1, and that only c_2 can be referenced by m_2, then S and R can determine the optimal correspondence such that m_1 refers to c_1 and m_2 refers to c_2. Parikh explains this disambiguation in game-theoretic terms as a collaboration between S and R [12,13]. This sort of collaborative disambiguation may extend to *referential coherence* in general, but Parikh does not consider that direction.

2.3 Centering Theory

Centering theory [7] is a standard theory of referential coherence and anaphora. Referential coherence is the smoothness of attention transition from an utterance to the subsequent one. The sender S and receiver R prefer referentially coherent mappings between referring expressions and their referents through the discourse. Centering theory consists of rules about referential coherence without explicitly considering S and R's cooperative cognitive processes.

Let U_i represent the i-th utterance unit (clause) in the discourse in question. According to centering theory, entities referenced in U_i are ordered by *Cf-ranking*, i.e., the salience ranking of grammatical functions, dependent on the language. Cf-rankings in English and Japanese are as follow [14]:

English Cf-ranking:
Subject $\succ$ Direct object $\succ$ Indirect object $\succ$ Other complements $\succ$ Adjuncts
Japanese Cf-ranking:
Topic (*wa*) $\succ$ Subject (*ga*) $\succ$ Indirect object (*ni*) $\succ$ Direct object (*o*) $\succ$ Others

Table 3. Transition types of Rule 2

	$\mathrm{Cb}(U_{i+1}) = \mathrm{Cb}(U_i)$	$\mathrm{Cb}(U_{i+1}) \neq \mathrm{Cb}(U_i)$
$\mathrm{Cb}(U_{i+1}) = \mathrm{Cp}(U_{i+1})$	Continue	Smooth-Shift
$\mathrm{Cb}(U_{i+1}) \neq \mathrm{Cp}(U_{i+1})$	Retain	Rough-Shift

The entities referenced in U_i, ordered by salience, are defined as *forward-looking centers*, $\mathrm{Cf}(U_i)$. The highest ranked entity in $\mathrm{Cf}(U_{i-1})$ that is referenced in U_i is defined as the *backward-looking center*, $\mathrm{Cb}(U_i)$. The highest ranked entity in $\mathrm{Cf}(U_i)$ is defined as the *preferred center*, $\mathrm{Cp}(U_i)$.

The two rules about referential coherence are:

Rule 1 (pronominalization): If an entity e in $\mathrm{Cf}(U_i)$ is referenced by a pronoun in U_{i+1}, then $\mathrm{Cb}(U_{i+1})$ must also be referenced by a pronoun in U_{i+1} (e may be $\mathrm{Cb}(U_{i+1})$).
Rule 2 (topic continuity): Continue $\succ$ Retain $\succ$ Smooth-Shift $\succ$ Rough-Shift (See Table 3).

Although centering theory gives the above qualitative model of referential coherence, it has the following drawbacks.

1. **No general principles.** Without a cooperative cognitive principle, the two rules explain only superficial phenomena about referential coherence. A standard theory of intentional communication requires a principle of action selection for the sender and receiver. Centering theory does not explain what kind of mechanism is working when S and R cooperatively prefer strong referential coherence.
2. **Many different versions.** Because of the lack of general principles, different researchers have proposed different versions of centering theory [15]. Thier variations do not clearly specify the principle behind referential coherence. Without a principle of cooperative preference, we fear that variations will grow disorderly.
3. **No quantitative predictions.** Centering theory gives no quantitative scale measuring how S and R prefer particular candidate mappings between expressions and their referents in the succeeding utterance. We believe that such a scale is required for automatic discourse compilation systems (e.g., dialogue systems, summarization systems, etc.).

3 General Framework of Meaning Game

The meaning game [5,6] formulates intentional communication to address the core of nonnatural meaning on the basis of game-theoretic principles. Let C be the set of semantic contents and M be the set of the linguistic messages. The sender S sends a message $m \in M$ to convey content $c_S \in C$. The receiver R interprets m as meaning $c_R \in C$. Therefore, a turn of communication is represented as $\langle c_S, m, c_R \rangle \in C \times M \times C$. $c_S = c_R$ is a necessary condition for this turn of communication to be successful.

The utility function Ut_X of player X would thus be a real-valued function from $C \times M \times C$ (the set of turns). Only if the communication is successful are Ut_S and Ut_R mutually positive. $\mathrm{Pr}(c_S, m, c_R)$ is the probability that a turn $\langle c_S, m, c_R \rangle$ occurs (i.e., S selects $\langle c_S, m \rangle$ and R selects $\langle m, c_R \rangle$).

The meaning game assumes the following:

1. $c_S = c_R$, i.e., S and R's communication is always successful, because Ut_S and Ut_R are mutually positive only if the communication is successful.
2. S and R's common belief includes the distribution of $\mathrm{Pr}(c_S, m, c_R)$.

The expected utility of player X in the succeeding turn is defined as follows:

$$\sum_{\langle c_S, m, c_R \rangle \in C \times M \times C} \mathrm{Pr}(c_S, m, c_R) \mathrm{Ut}_X(c_S, m, c_R)$$

S and R respectively decide the strategy $x = \langle c_S, m, c_R \rangle$ according to their expected utilities. Under the above assumptions, S and R cooperatively decide the strategy $x = \langle c, m, c \rangle$ for success of communication. Then, S and R select the effective solution x that should be a *Pareto-optimal evolutionarily stable strategy (ESS)* [16][2] satisfying the following conditions:

$$\mathrm{Ut}_X(x, x) \geq \mathrm{Ut}_X(y, x) \text{ for all } y$$

$$\mathrm{Ut}_X(y, x) = \mathrm{Ut}_X(x, x) \Rightarrow \mathrm{Ut}_X(y, y) < \mathrm{Ut}_X(x, y)$$

Here, $\mathrm{Ut}_X(x, y)$ represents the utility of X when X's strategy is x and the other player's strategy is y. The first condition indicates that an ESS is a Nash equilibrium. In this way, the social optimization of intentional communication can be reduced to the game theoretic account.

4 Meaning-Game Account of Centering Theory

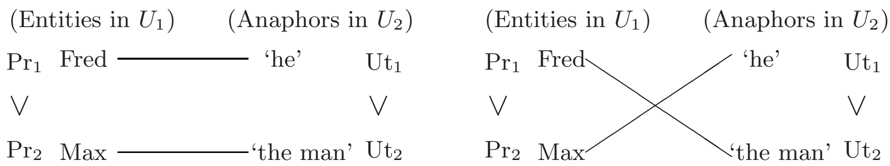
4.1 Derivation of Centering Theory from Meaning Game

Hasida demonstrated that centering theory can be derived from the meaning game [5]. Let us consider the following discourse.

U_1: Fred scolded Max.
U_2: He was angry with the man.

The preferred interpretation of 'he' and 'the man' in U_2 are Fred and Max, respectively, rather than the contrary. This preference, which complies with Rules 1 and 2 of centering theory, is accounted for by the meaning game shown as Figure 1. Let Pr_1 and Pr_2 be the probabilities of Fred and Max being referred to in U_2, respectively. We assume $\mathrm{Pr}_1 > \mathrm{Pr}_2$ because Fred is more salient than Max

[2] When most players' strategy is an ESS x, any other strategy y is disadvantageous.



$$\mathrm{Pr}_1\mathrm{Ut}_1 + \mathrm{Pr}_2\mathrm{Ut}_2 \quad > \quad \mathrm{Pr}_1\mathrm{Ut}_2 + \mathrm{Pr}_2\mathrm{Ut}_1$$
$$\text{because } (\mathrm{Pr}_1 - \mathrm{Pr}_2)(\mathrm{Ut}_1 - \mathrm{Ut}_2) > 0$$

Fig. 1. Derivation of Rule 1 from the meaning game

according to the grammatical functions in U_1. Let Ut_1 and Ut_2 be the utilities of 'he' and 'the man' being used in U_2, respectively. We assume $\mathrm{Ut}_1 > \mathrm{Ut}_2$ because 'he' costs less than 'the man' to perceptually process for S and R. Now we have $(\mathrm{Pr}_1 - \mathrm{Pr}_2)(\mathrm{Ut}_1 - \mathrm{Ut}_2) > 0$ because $\mathrm{Pr}_1 > \mathrm{Pr}_2$ and $\mathrm{Ut}_1 > \mathrm{Ut}_2$. Namely, the combination of mappings in the left-hand side entails a greater joint expected utility.

4.2 Generalized Formulation

Hasida's derivation of centering theory can be generalized to actual examples in a real discourse structure (Figure 2) [17,18]. Hereafter, let w be a target anaphor (i.e., a message) in a succeeding utterance unit and e be an entity (i.e., a semantic content) as a candidate of a referent of the anaphor. Let $\mathrm{pre}(U_i)$ be $[U_1, U_2, \cdots, U_i]$, the preceding discourse of U_i, when S and R predict the succeeding utterance U_{i+1}. In order to generally formulate Hasida's derivation, we define the following two parameters:

Reference probability of e at U_i, represented as $\mathrm{Pr}(e|\mathrm{pre}(U_i))$. It is defined as the conditional probability of e being referred to in a succeeding utterance unit U_{i+1}, given the referential features of e in $\mathrm{pre}(U_i)$ (see Table 5). It represents discourse salience at the moment of U_i (i.e., degree of joint attention to e between S and R) because a salient entity tends to be continuously referred to.

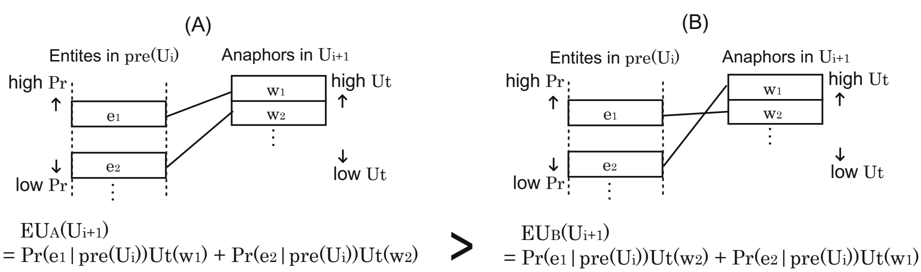


Fig. 2. Preference 1: Generalization of Rule 1 of centering theory

Perceptual utility of w, represented as $\mathrm{Ut}(w)$. It is defined as the reduction of the *perceptual cost* for use of w between S and R. Frequently used expressions (e.g., pronouns) have higher perceptual utilities than rare nouns because S and R are perceptually familiarized with them.

We assume the communication between S and R is successful; i.e., S and R can cooperatively select the turn $\langle w, e\rangle$. Therefore $\mathrm{EU}(U_{i+1})$, S and R's joint expected utility of turns $\langle w, e\rangle$ in U_{i+1}, the succeeding utterance, can be formulated as follows:

$$\mathrm{EU}(U_{i+1}) = \sum_{w \text{ refers to } e \text{ in } U_{i+1}} \Pr(e|\mathrm{pre}(U_i))\mathrm{Ut}(w)$$

.

S and R cooperatively prefer solutions $\langle w, e\rangle$ that have higher expected utility. This is the principle of expected utility. The above meaning-game-based formulation corresponds to the original centering theory, as Table 4 shows. Rules 1 and 2 of centering theory, which represent preference of referential coherence, can be generalized as follows:

Preference 1: In Figure 2, (A) is preferred over (B). That is, w_1 refers to e_1 and w_2 refers to e_2 when $\Pr(e_1|\mathrm{pre}(U_i)) > \Pr(e_2|\mathrm{pre}(U_i))$ and $\mathrm{Ut}(w_1) > \mathrm{Ut}(w_2)$, given that both w_1 and w_2 are in U_{i+1}.
Preference 2: The higher $\mathrm{EU}(U_{i+1})$ is preferred.

These preferences are derived from the principle of expected utility. In other words, the meaning-game-based account of centering theory is under the hypothesis that the referential coherence can be reduced to the principle of expected utility.

Table 4. Correspondence of centering theory's account and the meaning game's account

	centering theory (non-quantified/rule-based)	meaning game (quantified/corpus-based)
Discourse salience	Cf-ranking (Subject>Object> ···)	**Reference probability** $\Pr(e\|\mathrm{pre}(U_i))$
Load reduction	Pronominalization (Pronoun / Non-pronoun)	**Perceptual utility** $\mathrm{Ut}(w)$
Referential coherence	Transition ranking (CONTINUE >RETAIN >SMOOTH-SHIFT >ROUGH-SHIFT)	**Expected utility** $\mathrm{EU}(U_{i+1}) = \sum_{w \text{ refers to } e \text{ in } U_{i+1}} \Pr(e\|\mathrm{pre}(U_i))\mathrm{Ut}(w)$

$\mathrm{pre}(U_i)$: Preceding discourse
U_{i+1}: Succeeding utterance unit
e: Entity referenced in $\mathrm{pre}(U_i)$
w: Anaphor in U_{i+1}

4.3 Statistical Definition of Parameters

Here, we describe a statistical calculation of the parameters, Pr and Ut, on the basis of a corpus.

Calculation of Pr(e|pre(U_i)). Table 5 shows the referential features of e for calculating $\Pr(e|\mathrm{pre}(U_i))$. Figure 3 outlines the basic idea of calculating $\Pr(e|\mathrm{pre}(U_i))$ by using a corpus. Practically, it requires logistic regression in order to cope with data sparseness in the corpus. It requires a training phase, i.e., estimation of the regression weights b_0, b_1, b_2, and b_3 in the logit formula of the logistic regression. The regression weights are calculated with maximum-likelihood method by using training data in the corpus. Then the reference probability of e at U_i can be calculated by the following equation.

$$\Pr(e|\mathrm{pre}(U_i)) = (1 + \exp(-(b_0 + b_1 \mathit{dist} + b_2 \mathit{gram} + b_3 \mathit{chain})))^{-1}$$

Calculation of Ut(w). The perceptual cost of w can be defined as $\mathrm{I}(w) = -\log p(w)$ because S and R are perceptually familiar with the frequent anaphors. Let Ut_0 be the constant utility of successful communication; i.e., S and R cooperatively regard e as the referent of w. Then the perceptual utility of w can be calculated by the following equation.

$$\begin{aligned} \mathrm{Ut}(w) &= Ut_0 - \mathrm{I}(w) \\ &= Ut_0 + \log p(w) \end{aligned}$$

Table 5. Referential features of e in $\mathrm{pre}(U_i)$ for calculation of $\Pr(e|\mathrm{pre}(U_i))$

dist	log((#utterances between U_i and the latest reference to e in $\mathrm{pre}(U_i)$) + 1)
gram	Reference probability preliminarily calculated from only grammatical function of the latest reference to e in $\mathrm{pre}(U_i)$ (see Tables 6 and 7)
chain	log((#references to e in $\mathrm{pre}(U_i)$) + 1)

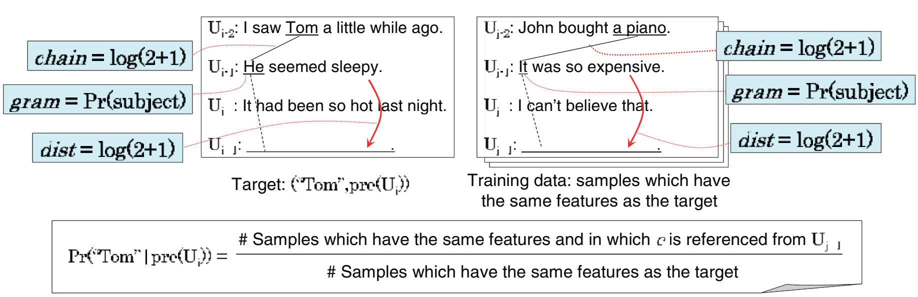


Fig. 3. Basic idea for calculating $\Pr(e|\mathrm{pre}(U_i))$

4.4 Empirical Verification

Now we empirically verify the hypothesis that centering theory can be reduced to the principle of the expected utility in meaning game. We use corpora of newspaper articles for verification. As a Japanese corpus, we use 1,356 articles from Mainichi-Shinbun (*Mainichi*). It contains 63,562 predicate clauses (i.e., utterance units) and 16,728 anaphors. As an English corpus, we use 2,412 articles from Wall Street Journal (*WSJ*). It contains 135,278 predicate clauses (i.e., utterance units) and 95,677 anaphors. The verification requires linguistic annotations to specify the structure of morphology, dependency, and anaphora. Both corpora have been manually annotated according to the Global Document Annotation (GDA) [19] tag set.

Verification of Parameter Definitions. Calculation of the reference probability (see also Figure 3) requires the following two preparations. First, we need to assign a *gram* value (in Table 5 and Figure 3) to each grammatical function. We thus assign average Pr of each grammatical function, which is calculated by counting samples in the *Mainichi* and *WSJ* corpora. Tables 6 and 7 show the *gram* values thus obtained from both corpora. The results show the consistency between the reference probability and salience ranking (i.e., Cf-ranking) of centering theory. Second, we need to obtain the regression weights b_i from the corpora. Table 8 shows the b_i values obtained by logistic regression analysis. The results show the consistency between the obtained regression weights and linguistic heuristics. For instance, the negative values of the weight of *dist*, b_1, are consistent with the heuristics that the entities referred to recently are more salient than the earlier ones.

We can empirically estimate Ut_0 to be 15.1 in the *Mainichi* corpus and Ut_0 to be 12.6 in the *WSJ* corpus. Calculation of the perceptual utilities also result in a valid consistency between the obtained values and the heuristic rules of centering theory [17,18].

Verification of Preference 1. Rule 1 of centering theory is generalized as Preference 1, which is represented as Figure 2. Below, we verify the hypothesis that Rule 1 of centering theory can be reduced to the principle of the expected utility.

If the succeeding utterance U_{i+1} includes a pair anaphors w_1 and w_2, the pair has two possible mappings: (A) and (B) in Figure 2. For the pairs that have a large difference $\mathrm{EU_A}(U_{i+1}) - \mathrm{EU_B}(U_{i+1})$, S and R should strongly prefer (A) over (B). Consequently, the ratio of positive samples (which comply with the preference) should reach almost 100%. On the other hand, for the pairs that have small $\mathrm{EU_A}(U_{i+1}) - \mathrm{EU_B}(U_{i+1})$, S and R should weakly prefer (A). In this case, S and R can select (B) a little less than (A). Hence, the ratio of the positive samples should be a little greater than 50%.

Figures 4 and 5 show the results on the corpora. For $\mathrm{EU_A} - \mathrm{EU_B} \geq 3$, the ratios of the positive pairs to all pairs, were 0.825 in the *Mainichi* corpus and 0.822 in the *WSJ* corpus. For $\mathrm{EU_A} - \mathrm{EU_B} < 0.5$, the ratios of the positive pairs were 0.564 in the *Mainichi* corpus and 0.529 in the *WSJ* corpus. That is to say,

Table 6. Average reference probability for each grammatical function (*Mainichi*)

grammatical function	# samples	# successive references	avg. Pr (*gram*)
Topic (*wa*)	35,329	1,908	5.40×10^{-2}
Subject (*ga*)	38,450	1,107	2.88×10^{-2}
(*no*)	88,695	1,755	1.98×10^{-2}
Object (*o*)	50,217	898	1.79×10^{-2}
Indirect object (*ni*)	46,058	569	1.24×10^{-2}
(*mo*)	8,710	105	1.21×10^{-2}
(*de*)	24,142	267	1.11×10^{-2}
(*kara*)	7,963	76	9.54×10^{-3}
(*to*)	19,383	129	6.66×10^{-3}

Table 7. Average reference probability for each grammatical function (*WSJ*)

grammatical function	# samples	# successive references	avg. Pr (*gram*)
Subject	76,147	16,441	2.16×10^{-1}
(*by*)	5,045	618	1.22×10^{-1}
Indirect object	1,569	184	1.17×10^{-1}
(*with*)	4,272	446	1.04×10^{-1}
(*of*)	23,798	2,145	9.01×10^{-2}
(*from*)	4,005	350	8.74×10^{-2}
Object	42,578	3,703	8.70×10^{-2}
(*to*)	8,449	661	7.82×10^{-2}
(*for*)	7,759	601	7.75×10^{-2}
Complement	7,102	371	5.22×10^{-2}

Table 8. Regression weights in logistic regression for calculating reference probability

Corpus	b_0(const.)	b_1(*dist*)	b_2(*gram*)	b_3(*chain*)
Mainichi	-2.825	-0.7636	9.036	2.048
WSJ	-2.405	-1.411	8.788	3.519

the greater difference ($\mathrm{EU_A} - \mathrm{EU_B}$), the stronger preference. This proves our prediction.

Moreover, Spearman's rank correlation coefficients between $\mathrm{EU_A} - \mathrm{EU_B}$ and the ratio of positive samples were 0.833 in the *Mainichi* corpus and 0.981 in the *WSJ* corpus. These values indicated that the restriction of Preference 1 is strongly associated with $\mathrm{EU_A} - \mathrm{EU_B}$. Therefore, the verification strongly proves the hypothesis that Preference 1 (and Rule 1 of centering theory) can be reduced to the principle of expected utility in both Japanese and English.

Verification of Preference 2. Rule 2 of centering theory is generalized as Preference 2, which is just the principle of expected utility. Below, we verify the

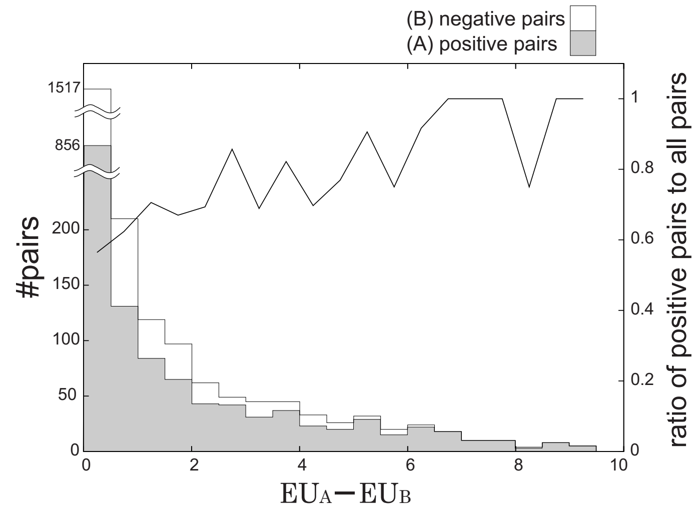


Fig. 4. Preference 1: Ratio of positive pairs is proportional to $\mathrm{EU_A}(U_{i+1}) - \mathrm{EU_B}(U_{i+1})$ (*Mainichi*)

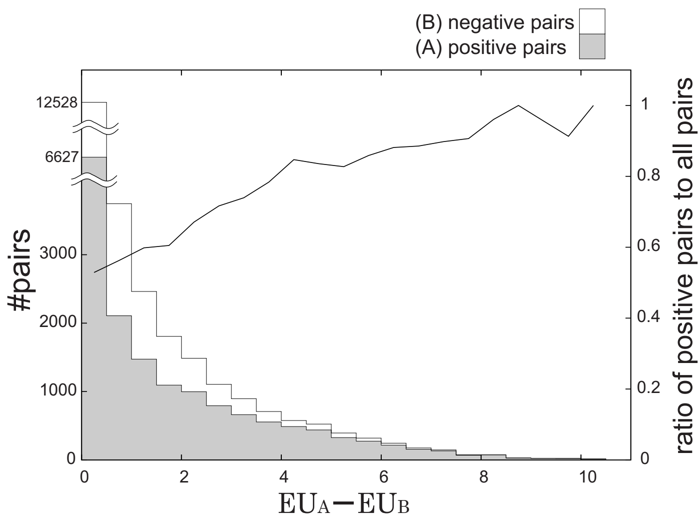


Fig. 5. Preference 1: Ratio of positive pairs is proportional to $\mathrm{EU_A}(U_{i+1}) - \mathrm{EU_B}(U_{i+1})$ (*WSJ*)

hypothesis that Rule 2 of centering theory can be reduced to the principle of the expected utility.

Rule 2 of centering theory is represented as the ranking among the four transition types: Continue $\succ$ Retain $\succ$ Smooth-Shift $\succ$ Rough-Shift. For the verification of the hypothesis, we investigated the consistency between $\mathrm{EU}(U_{i+1})$ and the above transition ranking by using the following procedure:

Table 9. Preference 2: Avg. $\mathrm{EU}(U_{i+1})$ of each transition type (*Mainichi*)

transition type	#sample	Avg. $\mathrm{EU}(U_{i+1})$	95% confidence interval
Continue	1,783	6.89	[6.68, 7.09]
Retain	84	5.07	[3.99, 6.16]
Smooth-Shift	2,704	1.59	[1.49, 1.68]
Rough-Shift	194	0.81	[0.57, 1.05]

Table 10. Preference 2: Avg. $\mathrm{EU}(U_{i+1})$ of each transition type (*WSJ*)

transition type	#sample	Avg. $\mathrm{EU}(U_{i+1})$	95% confidence interval
Continue	13,384	5.90	[5.83, 5.98]
Retain	2,314	3.67	[3.53, 3.80]
Smooth-Shift	18,904	2.96	[2.91, 3.01]
Rough-Shift	5,628	1.17	[1.10, 1.23]

1. Determine $\mathrm{Cb}(U_i)$, $\mathrm{Cb}(U_{i+1})$, and $\mathrm{Cp}(U_{i+1})$ by using the reference probability instead of Cf-ranking as follows:

   $\mathrm{Cb}(U_i)$: The entity in $\bigcup_{k=1}^{i} \mathrm{Cf}(U_k)$ which has the highest $\Pr(e|\mathrm{pre}(U_{i-1}))$
   $\mathrm{Cb}(U_{i+1})$: The entity in $\bigcup_{k=1}^{i+1} \mathrm{Cf}(U_k)$ which has the highest $\Pr(e|\mathrm{pre}(U_i))$
   $\mathrm{Cp}(U_{i+1})$: The entity in $\bigcup_{k=1}^{i+1} \mathrm{Cf}(U_k)$ which has the highest $\Pr(e|\mathrm{pre}(U_{i+1}))$
2. Determine transition types (Continue, Retain, Smooth-Shift, or Rough-Shift) on the basis of $\mathrm{Cb}(U_i)$, $\mathrm{Cb}(U_{i+1})$, and $\mathrm{Cp}(U_{i+1})$.
3. Perform Wilcoxon's rank sum test for verifying consistency between the order of Rule 2 (Continue $\succ$ Retain $\succ$ Smooth-Shift $\succ$ Rough-Shift) and the order of average $\mathrm{EU}(U_{i+1})$ between transition types.
4. Calculate Spearman's rank correlation coefficient between $\mathrm{EU}(U_{i+1})$ and the order of Rule 2.

Tables 9 and 10 shows the average $\mathrm{EU}(U_{i+1})$ for each transition type. Figures 6 and 7 show the average and distribution of $\mathrm{EU}(U_{i+1})$ for each transition type. The order of average $\mathrm{EU}(U_{i+1})$ is consistent with the order of Rule 2 of centering theory in both the *Mainichi* and *WSJ* corpora. Wilcoxon's rank sum test shows consistency between the order of average $\mathrm{EU}(U_{i+1})$ and the order of Rule 2 at a significant level ($< 2.2 \times 10^{-16}$). Spearman's rank correlation coefficients between $\mathrm{EU}(U_{i+1})$ and the order of Rule 2 are as follows: 0.639 (95% confidential interval: $[0.621, 0.655]$) in the *Mainichi* corpus and 0.482 (95% confidential interval: $[0.474, 0.489]$) in the *WSJ* corpus.

These results show the consistency between $\mathrm{EU}(U_{i+1})$ and Rule 2 with the statistical significance. This empirically proves the hypothesis that Rule 2 of centering theory can be reduced to the principle of the expected utility in both Japanese and English.

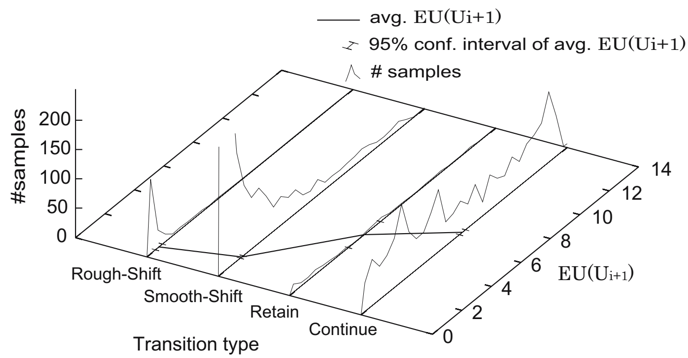


Fig. 6. Preference 2: Avg. EU(U_{i+1}) of each transition type (*Mainichi*)

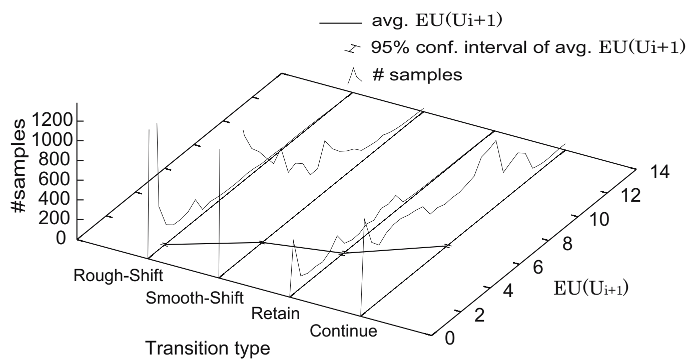


Fig. 7. Preference Rule 2: Avg. EU(U_{i+1}) of each transition type (*WSJ*)

Thus, the hypothesis that centering theory can be reduced to game theory has been empirically verified in both Japanese and English corpora. In conclusion, S and R's cooperative preference of referential coherence does not depend on language because Japanese and English are quite dissimilar from each other.

5 Concluding Remarks

Intentional communication consists of cooperation of rational individuals with each other. We have accounted for this cooperative nature of communication, or in particular linguistic communication, in terms of game theory. In both Japanese and English, we have provided a corpus-based verification of the hypothesis that centering theory is reduced to meaning game. We thus predict that no language-specific rules are needed on referential coherence.

Language use probably consists of Pareto-optimal ESSs of potentially very complex meaning games. There is still much to do before we fully understand this complexity. However, game theory will provide statistical and analytical means to approach this issue in natural languages and other communication modalities.

References

1. von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton (1944)
2. Nash, J.: Non-cooperative games. Annals of Mathematics 54, 286–295 (1951)
3. Grice, H.P.: Meaning. Philosophical Review 66, 377–388 (1957)
4. Grice, H.P.: Utterer's meaning and intentions. Philosophical Review 68(2), 147–177
5. Hasida, K., Nagao, K., Miyata, T.: A Game-Theoretic Account of Collaboration in Communication. In: Proceedings of the First International Conference on Multi-Agent Systems, pp. 140–147 (1995)
6. Hasida, K.: Issues in Communication Game. In: Proceedings of COLING 1996, pp. 531–536 (1996)
7. Grosz, B., Joshi, A., Weinstein, S.: Centering: A Framework for Modeling the Local Coherence of Discourse. Computational Linguistics 21(2), 203–225 (1995)
8. Spence, M.: Job market signaling. Quarterly Journal of Economics 87, 355–374 (1973)
9. Zahavi, A.: Mate selection - A selection for a handicap. Journal of Theoretical Biology 53, 205–214 (1975)
10. Zahavi, A.: The cost of honesty (further remarks on the handicap principle). Journal of Theoretical Biology 67, 603–605 (1977)
11. Barwise, J., Perry, J.: Situations and Attitudes. The MIT Press, Cambridge (1983)
12. Parikh, P.: Situations, games, and ambiguity. In: Kukai, K., Perry, J. (eds.) Situation Theory and Its Applications, I, CSLI Publications, Stanford (1990)
13. Parikh, P.: The Use of Language. CSLI Publications, Stanford, California (2001)
14. Walker, M., Iida, M., Cote, S.: Japanese Discourse and the Process of Centering. Computational Linguistics 20(2), 193–232 (1994)
15. Poesio, M., et al.: Centering: A parametric theory and its instantiations. Computational Linguistics 30(3), 309–363 (2004)
16. Smith, J.M., Price, G.: The logic of animal conflict. Nature 246, 15–18 (1973)
17. Shiramatsu, S., et al.: Meaning-Game-based Centering Model with Statistical Definition of Utility of Referential Expression and Its Verification Using Japanese and English Corpora. In: Branco, A. (ed.) DAARC 2007. LNCS (LNAI), vol. 4410, pp. 121–126. Springer, Heidelberg (2007)
18. Shiramatsu, S., et al.: A Game-Theoretic Model of Referential Coherence and Its Statistical Verification Based on Large Japanese and English Corpora. Journal of Natural Language Processing 14(5), 199–239 (2007)
19. Hasida, K.: Global Document Annotation (GDA) (1998), http://-content.org/GDA

Part III

Risk Informatics

International Workshop on Risk Informatics (RI2007)
- Approach from Data Mining and Statistics -

Takashi Washio[1] and Shusaku Tsumoto[2]

[1] The Institute of Scientific and Industrial Research (I.S.I.R), Osaka University
8-1, Mihogaoka, Ibarakishi, Osaka 567-0047, Japan
[2] Department of Medical Informatics, School of Medicine, Shimane University
89-1 Enya-cho, Izumo, Shimane 693-8501, Japan
washio@ar.sanken.osaka-u.ac.jp, tsumoto@computer.org

Along the enhancement of our social life level, people became to pay more attention to the risk of our society to ensure our life very safe. Under this increasing demand, modern science and engineering now have to provide efficient measures to reduce our social risk in various aspects. On the other hand, the accumulation of a large amount of data on our activities is going on under the introduction of information technology to our society. This data can be used to efficiently manage the risks in the society. The Workshop on Risk Mining 2006 (RM2006) was held in June, 2006 based on these demand and situation while focusing the risk management based on data mining techniques [1,2]. However, the study of the risk management has a long history on the basis of mathematical statistics, and the mathematical statistics is now making remarkable progress in the data analysis field. The successive workshop in this year, the International Workshop on Risk Informatics (RI2007), extended its scope to include the risk management by the data analysis based on both data mining and mathematical statistics.

The primary objectives of this workshop were to bring together researchers who are interested in the areas of risk management, data mining and mathematical statistics, and to have intensive discussions on various aspects on risk management based the data analysis. The workshop program consisted of three sessions focusing on many application domains of the risk management. The first session named “Risk Discovery and Analysis in Society and Business” was the collection of 4 paper presentations to discover and analyze risks of social and business activities. The second session named “Risk Discovery and Analysis in Medicine” had 3 paper presentations to discover and analyze various risk mechanism in medical fields. The third section “Risk Discovery and Analysis in Network” included two paper presentations to discover and manage risks associated with computer and human networks. All areas covered by this workshop play indispensable roles in our modern society.

Among the studies presented, we further applied strict reviews, and selected 5 outstanding papers. They have been further extended, elaborated, and were included in this chapter. The first 3 of these 5 papers are from the first session. The paper titled “Chance Discovery in Credit Risk Management” proposed time ordered KeyGraph to distinguish and express the time order of chain reaction bankruptcy. The paper titled “Risk Bias Externalization for Offshore Software Outsourcing by Conjoint Analysis” utilizes conjoint analysis on questionnaire for the preference of the outsourced project to externalize the tacit knowledge associated with the risk control of the project managers. The third paper named “Extracting Failure Knowledge with Associative Search” applied associative

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 245–246, 2008.

OR search on a Failure Knowledge Database to find cases most analogous to risks concerned by engineers. The other 2 papers are from the second session. The paper titled "Data Mining Analysis of Relationship between Blood Stream Infection and Clinical Background in Patients Undergoing Lactobacillus Therapy" analyzes the effects of lactobacillus therapy and the background risk factors on blood stream infection in patients by using data mining. The last paper named "Discovery of Risky Cases in Chronic Diseases: an Approach Using Trajectory Grouping" presents an approach to finding risky cases in chronic diseases using a trajectory grouping technique. These papers indicate promising feasibility to reduce negative risks and to enhance positive effects of the risk management in our future society.

Finally, the chairs of this workshop would like to express the deep appreciation to Dr. Akinori Abe in Advanced Telecommunications Research Institute International, Prof. Shoji Hirano in School of Medicine, Shimane University, Prof. Yukio Ohsawa in The University of Tokyo, Prof. Katsutoshi Yada in Kansai University and Prof. Kenichi Yoshida in University of Tsukuba for their extensive support on the paper review process.

Program Chairs

Takashi Washio Osaka University, Japan
Shusaku Tsumoto School of Medicine, Shimane University, Japan

References

1. The Workshop on Risk Mining 2006 (RM2006) (2006), http://www.ar.sanken.osaka-u.ac.jp/rm/rm06.html
2. Washio, T., et al. (eds.): JSAI 2006. LNCS (LNAI), vol. 4384, pp. 303–304. Springer, Heidelberg (2007)

Chance Discovery in Credit Risk Management

Time Order Method and Directed KeyGraph for Estimation of Chain Reaction Bankruptcy Structure

Shinichi Goda and Yukio Ohsawa

School of Engineering, The University of Tokyo, 113-8656 Japan
sgoda@mwa.biglobe.ne.jp

Abstract. In this article, chance discovery method is applied to estimate chain reaction bankruptcy structure. Risk of default can be better forecasted by taking chain reaction effect into accont. Time order method and directed KeyGraph are newly introduced to distinguish and express the time order among defaults that is essential information for the analysis of chain reaction bankruptcy. The steps for the data analysis are introduced and result of example analysis with default data in Kyushu, Japan, 2005 is presented.

Keywords: chance discovery, credit risk, chain reaction, bankruptcy.

1 Introduction

Credit risk management based on portfolio theory becomes popular in recent Japanese financial industry. Simulation method comes to be common tool for analysis of credit portfolio and simulation models have been developed for credit risk management [1,2,3,4]. However, there still remain major areas for improvement. Analysis on chain reaction bankruptcies is one of these areas.

Chain reaction bankruptcies are common phenomenon. Most of industry experts regard it necessary to take the effect of chain reaction bankruptcy into accounts when they analyze profile of their credit risk portfolio. By introducing chain reaction factor into analysis, we can expect to better grasp the risk profile of credit risk portfolio. Risk of defaults can be forecasted sooner by using other defaults as leading indicator and amount of loss caused by a default can be better estimated by considering loss of other bankruptcies triggered by the default.

However, majority of simulation models actually used in business do not fully take the chain reaction into account. That is mainly because it is difficult to directly grasp relations among companies since available data is limited. Corporate data related to trade, financial and technological relations among corporations is very limited to public.

Adjustments have been devised in models and simulators to include relation effect, but there is much room for improvement. A major method to take the effect into account is to grasp relations among companies by measuring co-relations

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 247–254, 2008.

among movements of the security price of the companies (In this method, a company's security price is regarded as representative of default probability of the company). But this method is applicable only to security issuing large companies whose number is very small. On the other hand, most of companies in a credit risk portfolio are non-security-issuing and small-mid size.

In this article, we propose a method that grasp relationship among bankrupted companies, without direct information of trade relations or security prices, i.e. We introduce assumptions of the method in section 2 and the steps for the data analysis in section 3. And the result of example analysis with default data in Kyushu, Japan, 2005 is presented in section 4.

2 Grasp Chain Reaction Bankruptcy Structure by Chance Discovery Method

We proposed a method that detects relationship among bankrupted companies, without direct information of trade relations i.e., by chance discovery method in our previous work [5]. (The basic idea of this method follows the method that used for the chance discovery of earthquake by Ohsawa, Y. [6,7].)

In this method, we estimate trade and/or other relations among companies defaulted in a geographical area within a certain time period, by visualizing relations among industry groups that include the defaults with KeyGraph. The method is based on the assumptions as follows;

1. There should be relations that transmitted factors causing defaults among companies that defaulted in a geographical area within a certain time period.
2. The default transmitting relations among companies should be mainly based on and represented by relations among industry groups. General definition of chain reaction bankruptcies is like "bankruptcy triggered by a preceding default of a company that has trade and other relations with the bankrupting company". As seen in this definition of chain reaction bankruptcy, "trade relation" is generally regarded as one of the most influential relations. Trade relation between a pair of industry is generally universal among companies in the paired industries.
3. Default transmitting relations among industries could be paths to transmit default of a company in an industry to other companies in other industries.

Suppose that cloth retailer A and cloth wholesaler B defaulted within a month successively in Kansai district. Suppose other sets of cloth retailer and wholesaler those located in same districts defaulted successively within a month repeatedly. We can estimate with high confidence that there were sets of trade relation between the cloth retailer and the wholesaler defaulted successively and that the sets of trade relation between the two companies caused the successive default, even if there is no public information of sets of trade relations between the two companies. We can estimate so based on expert's knowledge about cloth trading industry and on the observed default patterns analyzed with KeyGraph.

3 Methodology

3.1 Original Method

First, we explain the original method proposed in our previous work [5] and introduce its basic idea. Steps are as follows (see Table 1);

1. Each default event has attributes of default date, geographical area in which the defaulted company is located and an industry to which it belongs. Defaults are grouped by area and sorted in each area by default dates.
2. Transform company data prepared in Step1. to industry data by replacing a company's name to a code of an industry to which the company belongs.
3. Make the default events grouped by area in sentence form by order of their default dates. Each event is denoted by industry name and spaced. Form one sentence starting from a company and ending at a company whose default date is one month after the default date of the starting company.
4. Extract co-occurrence of default events by using KeyGraph with sentences formed default data. Interpret a result of KeyGraph and estimate relations among industries. It is important to read relations expressed in a KeyGraph with experts' knowledge.

Table 1. Example of data and sentence

Step1. Data (1) sorted by date

No.	date of default	comp.	ind. code	area
1	2005/1/4	A	101	north
2	2005/1/15	B	210	south
3	2005/1/30	C	303	east
	⋮			
N	2005/12/28	ZZZ	330	north

Step2. Data (2) grouped by area, and sorted by date

No.	date of default	comp.	ind. code	area
3	2005/1/30	C	303	east
8	2005/3/20	H	440	east
	⋮			
1	2005/4/1	A	101	north
	⋮			
N	2005/12/28	ZZZ	330	north
	⋮			
9	2005/3/29	I	211	west

Step3. Transform to sentence form
Series starting from D1 = 101 110 418 ...
Series starting from D2 = 210 105 330 ...
Series starting from D3 = 303 240 440 ...
⋮

Examples of experts' knowledge about factors that are supposed to work behind relations extracted by KeyGraph are as listed below.

a. Technological and business relations among industries
   (example) An automobile is made of steel, glass, tires and electronic parts.
b. Commonality of customers among industries
   (example) Consumers living in Kyushu region shop at cloth/food retailers and eat restaurant both located in Kyushu.
c. Ownership relation:

3.2 Time Order Method (with Directed KeyGraph)

The original method had problem. It is not able to capture or to express time order among defaults, although chain reaction bankruptcy is time ordered phenomenon. Without time order information, it is difficult to use the result of the analysis for forecasting and simulation.

We newly introduce time order method to deal with above points. The basic idea of the new method is to try to better distinguish the causal relations among defaults with the time order among them.

With time order method, by making a sentence include only a pair of defaults with distinction of earlier event of the two, the time order among defaults can be expressed by KeyGraph with direction – we name it as "directed KeyGraph". Detailed steps are as follows (see Table 2);

1. Make a sentence of a pair of defaults.
2. Put "S" to industry code of the earlier default in a pair to distinguish it as a starting default.
3. Make series of pairs with a selected starting default indicated by "S", and ending defaults each of which follows the starting default within a period.
4. Select another default that occurred next to the first default as the second starting event and take step 3.
5. Repeat step 3. and 4. until all the defaults in the analyzed period are selected as starting events.
6. Make and include linking pairs to link the nodes those are captured as ending events but at the same time are starting ones. For example, in Table 2, default E, that is coded as 10, is included in pair "S20_10", "S30_10", "S40_10" as an ending default and also included in pair "S10_30", "S10_20" as a starting one.

Table 2. Method for making time ordered pair of defalts

Default date		1/1	1/10	1/20	2/1	2/5	2/25	2/28		12/1	12/20	12/30
Name of company		A	B	C	D	E	F	G	···	X	Y	Z
Industry code		10	20	30	40	10	30	20		40	10	50
Pair	Start from 10(A)		S10_20	S10_30								
	Start from 20(B)			S20_30	S20_40	S20_10						
	Start from 30(C)				S30_40	S30_10						
	Start from 40(D)					S40_10	S40_30	S40_20				
	Start from 10(E)						S10_30	S10_20	···			
	···									···		
	Start from 20(W)									S20_40	S20_10	
	Start from 40(X)										S40_10	S40_50

When the paired data is analyzed with directed KeyGraph, starting defaults are indicated by "S" put on the industry code. When a node, "S20" for example, is linked to another node, "10", it means that defaults in the industry 20 occurred before defaults in 10, indicating that defaults in 20 trigger the defaults in 10. In case two nodes of same industry are linked, like "10 – S10", it means ending defaults in "10" are at the same time starting defaults.

3.3 Characteristics and Merits of Time Order Method

Characteristics and possible merits of our new method, when we compare it to Bayesian network that is one of methods to express causal relations among nodes by directed arch[8-12], are as follows;

1. Our new method decides a direction of an arch based on time order of events, whereas Bayesian network decides a direction basically based on information criterion. Time order is generally regarded as one of the most basic information for estimating causal relations. It is especially important for causal estimation of time series event like chain reaction.
2. Our new method is able to express bidirectional/cyclic relations among industries, which might imply cyclic chain reaction of bankruptcies. Suppose the defaults in industry A cause defaults in industry B. Then, due to the defaults in industry B, the other companies in industry A may get defaulted. This positive feedback may trigger the non-linear growth of the defaults in both industries. This kind of non-liner effect is out of the scope of Bayesian network since basic method of Bayesian network is for handling one-direction/noncyclic structure.

4 Case Study – Analysis of Chain Reaction Structure of Bankruptcies in Kyushu, 2005

As a case study, we applied time order method described above to data of bankruptcies in Kyushu district, Japan, 2005. The reason for limiting area for analysis only to Kyushu district is simply to make the size of data controllable.

A. Contents of data
Samples are 343 defaults in Kyushu, a local part of the Japanese samples described below. About 10,400 pairs are made from the above selected 343 default samples.

1. Japanese samples, about 3,400 defaults, are randomly selected from the all defaulted companies that defaulted based on bankruptcy laws and published in the official gazette in Japan, 2005. The area consists of 9 districts including Kyushu. Samples are categorized in about 200 mid-level industry groups by author, based on the industry categories defined by Teikoku Databank, Ltd.
2. Period for pairing one starting default with ending ones is set to one month.
3. About 6,200 linking pairs are added.

B. Analysis
First, the data was analyzed by original method and then by time order method.

(1) Original method (see Fig. 1 for result)
With this KeyGraph by original method, we can understand that defaults in linked industries occurred in a close time period. But it is difficult to estimates

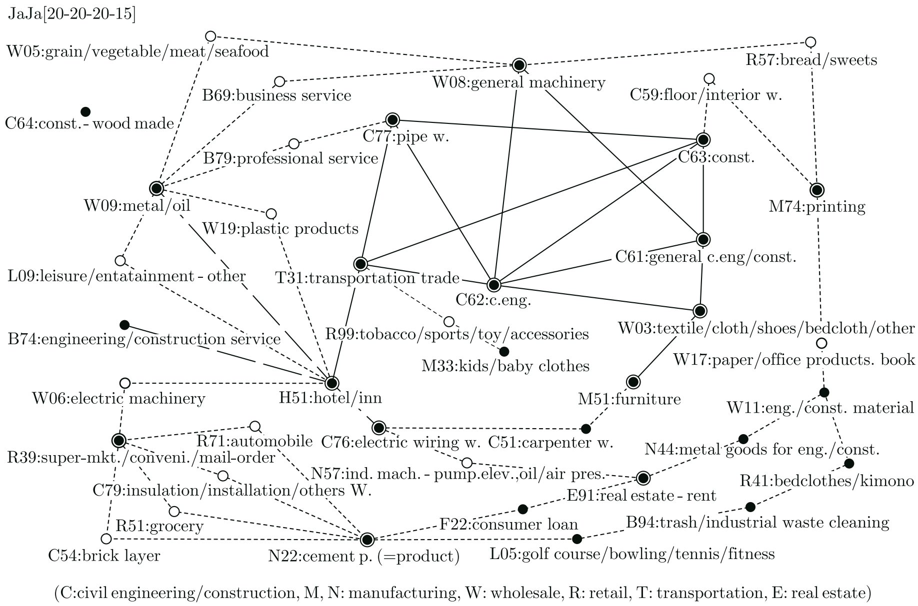


Fig. 1. KeyGraph by original method

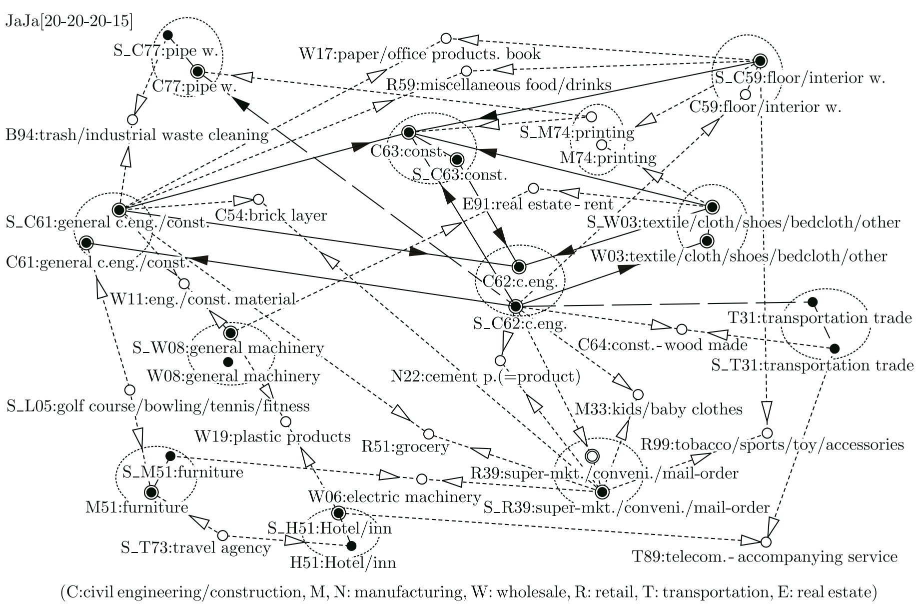


Fig. 2. KeyGraph by time order method ("S_" indicates starting default)

whether the co-occurrence is based on chain reaction. That is mainly because this graph lacks of information about time order among defaults.

(2) Time order method with directed KeyGraph (see Fig. 2 for result)
In Fig. 2, time orders among defaults in the industries are shown by arrows indicating causal relations among them. Nodes consist of starting defaults are indicated by "S_" put on industry codes. Time order is expressed by arrows from a node with "S_" to another node without "S_". Circled pair of two nodes, one with "S_", are in same industry. For example, an arrow from S_C61 goes to C63, showing defaults in C61 occurred before defaults in C63 in major case, not vice versa. C63 is linked to S_C63 and two nodes are circled. An arrow then goes from S_C63 to C62. It shows that defaults in C63 those occurred after defaults in C61 are, at the same time, defaults those occurred before those in C62. It indicates that defaults in C63 caused by those in C61 then triggered defaults in C62.

When we see Fig. 2 this way, we can better estimate causal relations among defaults. The number of arrows go out from C61 (= S_C61) and C62 (= S_C62) are greater than those go into C61 and C62, that indicates defaults in civil engineering/construction and civil engineering industry caused defaults in other industries in major case, not vice versa. The defaults in C61, C62 might then trigger defaults in variety of related industries like C54 (brick layer work), C77 (pipe work), T31 (transportation_trade), i.e. Arrows from S_W03 (textile/cloth/others wholesale) go to C62 and to C63, indicating defaults in W03, caused by depressed consumer spending for cloth, were source of defaults in C62 and in C63 other than decreased public construction work. Many arrows go out from R39, indicating defaults of super markets, caused by depressed consumer spending, triggered defaults of groceries, toy/accessory shops in the super markets and of electric machinery wholesalers who trade with the market.

5 Conclusion

In this article, we applied chance discovery method to estimate structure of industrial relations that are to transmit chain reaction of bankruptcy. We newly introduced time order method with directed KeyGraph to grasp/express time order among defaults and to better estimate causal relations among them.

The result of example analysis by new method with default data in Kyushu 2005 was promising. The structure grasped/expressed by new method was consistent with industrial knowledge. With further accumulation of analyses and improvement of method, a structure estimated by chance discovery method will sufficiently be a base for risk analysis and risk management of a credit portfolio.

The areas for further improvements are;

1. measurement of influence over a default probability of an industry/company of a default event to be transmitted through estimated industrial relations
2. techniques for extracting appropriate time range between start/end of defaults
3. modeling of estimated industrial relations in network structure for the use for risk analysis and risk management of a credit portfolio

References

1. Saunders, A.: Credit Risk Measurement (Japanese). Kinyu Zaisei Jijo Press (2001)
2. FISC: Working Report on Risk Management Model (1999)
3. Nakabayashi, A., Sasaki, M.: Models for credit risk measurement and its application to Japanese bank. FRI Review 2(2) (1998)
4. Torii, H.: Portfolio based credit risk management models – its effectiveness and limit. Weekly Kinyu Zaisei Jijo Magazine (June 1998)
5. Goda, S., Ohsawa, Y.: Chance Discovery in Credit Risk Management – Estimation of chain reaction bankruptcy structure by chance discovery method. In: Proceedings of the 2006 IEEE International Conference on Systems, Man, and Cybernetics, October 8–11, vol. 6 (2006)
6. Ohsawa, Y.: Discovering risky subduction zones of earthquake. In: Information Technology for Chance Discovery, pp. 314–325. Tokyo-Denki University Press (2005)
7. Ohsawa, Y.: Visualizing relations between chances and surrounding events. In: Information Technology for Chance Discovery, pp. 121–153. Tokyo-Denki University Press (2005)
8. Shigematsu, K., Ueno, M., Motomura, Y.: An Introduction to Bayesian Networks, Baifukan (2006)
9. Motomura, Y.: Modeling with large scale data by Bayesian networks. JSAI Technical Report SIG-CII-9707-04 (1997)
10. Motomura, Y.: Bayesian Network that Learns Conditional Probabilities by Neural Networks. In: Workshop on Information-Based Induction Sciences(IBIS) (1998)
11. Cooper, G., Herskovits, E.: A Bayesian methods for the induction of probabilistic Networks from data. Machine Learning 9, 309–347 (1992)
12. Mathematical Systems, Inc.: BayoNet 4.0 manual, http://www.msi.co.jp/BAYONET/files/manual.pdf

Risk Bias Externalization for Offshore Software Outsourcing by Conjoint Analysis

Zhongqi Sheng[1,2], Masayuki Nakano[1], Shingo Kubo[1], and Hiroshi Tsuji[1]

[1] Osaka Prefecture University, Graduate School of Engineering,
1-1 Gakuencho, Nakaku, Sakai, Japan, 559-8531
[2] Northeastern University, School of Mechanical Engineering,
No.3-11, Wenhua Road, Shenyang, Liaoning, China, 110004
{Sheng,Nakano,Kubo}@mis.cs.osakafu-u.ac.jp,Tsuji@cs.osakafu-u.ac.jp

Abstract. With the steady increase of volumes of software development, most Japanese companies are interested in offshore software outsourcing. In order to find out the know-how of experienced project managers and assess the risk bias brought by vendor countries and software types, this paper utilizes conjoint analysis method on questionnaire for project preference to externalize the tacit knowledge. After analyzing the *range*, *maximum*, *minimum* and *average* value of total utilities of three kinds of properties, we have the following findings: 1) the project property is the main item affecting success of outsourcing projects, which could lead to big success or severe failure, 2) the risk analysis result for vendors of India is different from that of China, which should be deeply investigated, 3) the risk value of middleware software is lower than that of the other two software types, which should be paid more attention to.

Keywords: Offshore Outsourcing, Risk Bias, Conjoint Analysis.

1 Introduction

Offshore software outsourcing refers to the outsourcing of software development work to facilities located in foreign countries [1]. The primary motivation for offshore software outsourcing is the low cost of software development in developing countries such as India, China and so on. The benefits of offshore software outsourcing also include compression of project development time, easy access to resource pool of highly skilled and experienced IT professionals and so on. In the era of globalization and specialization, outsourcing non-core activities to the third parties has become a universal tool, which helps companies to concentrate on their profit-generating activities. Offshore software outsourcing is playing an increasingly important role in the information technology strategies of major corporations.

According to the steady increase of volumes of software development, most Japanese companies are interested in offshore software outsourcing and many companies have successfully carried out offshore software developments [2]. However, due to the existing of language problems, opacity of software developments

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 255–268, 2008.

at the offshore site, insufficient level of control over offshore development team and so on, there are inevitably risks in offshore software outsourcing. Although those nominal risk items have been known, their magnitudes have not been measured. And the relationships among them are not known, which are important for taking trade-off in project decision. Therefore, it is an important task to measure risks and analyze their relationships for success/failure of offshore software outsourcing.

In order to extract the tacit knowledge of risk assessment from experienced project managers, we design one kind of questionnaire delivered to experienced project managers for risk research of offshore software outsourcing projects [3]. This paper uses conjoint analysis method on the collected questionnaire results for virtual projects to analyze the risk items and bias while SEM is used for real projects as discussed in [4]. Chapter 2 briefly introduces the basic concept and procedure of conjoint analysis, together with the concept of items for risk assessment in this paper. Chapter 3 presents the implement procedures supported by Joint Forum of Strategic Software Research in Japan. Next, chapter 4 introduces the analysis based on questionnaire of virtual projects from skilled managers. Chapter 5 discusses the analysis results and draws the conclusions.

2 Conjoint Analysis of Risk Assessment

2.1 Conjoint Analysis Method

Conjoint analysis is an advanced analytical technique used to determine the joint influence that attribute and level combinations of a product or service have on the purchase decisions[5]. A properly designed conjoint analysis study allows any combination of attributes and levels to be profiled to estimate market or choice share. Conjoint analysis includes a set of procedures that investigate the responses to products or services described with a set of independent attributes. The objective of conjoint analysis is to understand the contributions of these attributes to products or services and determine what combination of these attributes is most influential on respondent choice or decision making. A controlled set of potential products or services is shown to respondents and by analyzing how they make preferences between these products, the implicit valuation of the individual elements making up the product or service can be determined. These implicit valuations can be used to create market models that estimate market share, revenue and even profitability of new designs.

Conjoint analysis originated in mathematical psychology in the early 1970's and was developed by marketing professor Paul Green at the Wharton School of the University of Pennsylvania. Today it is used in many of the social sciences and applied sciences including marketing, product management, and operations research. Over the past decades conjoint analysis has become one of the most popular market research procedures for assessing how consumers with heterogeneous preferences trade off the various benefits they derive from product or service attributes. The analysis result usually provides critical input for many marketing decisions such as new product design, positioning and pricing[6][7].

Conjoint analysis is an analysis tool that allows a subset of the possible combinations of product attributes to be used to determine the relative importance of each attribute in the purchasing decision. Conjoint analysis is based on the fact that the relative values of attributes considered jointly can better be measured than when considered in isolation. In the case of conjoint analysis, the research object is taken as a composite of attributes. Every attribute has several levels and every project is consisted of a set of levels corresponding to these attributes. Conjoint analysis includes a systematic set of trade-offs between pairs of attributes, in which trade-offs would be analyzed to show the partial utilities of the separate level. By applying conjoint analysis, the partial utility of every level can be obtained, which shows its importance for the whole object. According to the partial utility values it would be possible to create newer and better composite of attributes, which will lead to better developed products or advertising campaigns.

2.2 Conjoint Analysis of Offshore Outsourcing Project

In this research, we apply conjoint analysis to find out the importance and utility of attributes and levels used to describe the offshore outsourcing projects. We ask respondents to vote on the feasibility of some virtual offshore software outsourcing projects. The virtual offshore outsourcing projects are a list of combinations of attributes and levels.Once this vote for the possibility of offshore outsourcing project is obtained, a calculation is carried out to find the utilities of different levels of each attribute that would show the respondent's preference. These virtual offshore outsourcing projects are created by orthogonal design method.The utilities can be determined using a subset of possible attribute combinations. So it is efficient in the sense that the survey does not need to be conducted using every possible combination of attributes. And from these results one can predict the desirability of the combinations that are not tested.

In the questionnaire of this research, we ask respondents to vote on a set of virtual offshore projects for the investigation of project property. These virtual offshore projects are created by orthogonal design method so every level can be investigated equally.Every project is described with five attributes which is labelled by p1, p2, p3, p4 and p5. Every attribute has two levels. So there are ten levels totally.

$$P = p_0 + \sum_{i=1}^{5} \sum_{j=1}^{2} p_{i,j} D \quad (1)$$

In equation (1), P is the performance of the project, which reflects the overall predilection of respondent to the project. $p_{i,j}$ is the partial utility of the jth level of ith attribute, which reflects the predilection of respondent to this level. D is a dummy variable. When the jth level of ith attribute is selected in this project, its value is 1 otherwise 0. Because every attribute has only two levels that are labelled by *high* and *low* and showed by 1 and -1 in calculation, the partial utilities of these two levels are opposite. So there are six unknown variables totally and six equations are necessary to solve out this question. That is to say we need to prepare six virtual projects to vote on.At last we design nine virtual projects in order to improve the analysis precision.The similar analysis is done for software property and vendor property.

2.3 Items for Risk Assessment

Here we take a series of virtual projects of offshore software outsourcing as the objects of conjoint analysis. Every virtual project is described by a group of attributes. Our basic premise is that experienced project managers know the risk as their tacit knowledge. The answer value given by respondents for the feasibility of every virtual project reveals the risk degree. Using conjoint analysis we can detect the partial utility of every attribute and their relative importance to the whole project.

Conjoint analysis calculates relative importance rate and partial utility of every attribute through regression method. In order to externalize the risk bias of repondents, we focus on four items including *average*,*range*,*maximum* and *minimum* on the base of conjoint analysis results.

Average value of risk is the const value given by conjoint analysis directly.

Range value of risk is the sum of range value of all partial utilities, which indicates the varying range of total utility.

$$Range = \sum |PartialUtility| * 2 \tag{2}$$

Maximum value of risk is the sum of *Average* value and *Range* value, which shows the maximum value of risk achieved possibly.

$$Maximum = Average + \sum |PartialUtility| \tag{3}$$

Minimum value of risk is the difference of *Average* value and *Range* value, which shows the minimum value of risk achieved possibly.

$$Minimum = Average - \sum |PartialUtility| \tag{4}$$

3 Questionnaire-Based Analysis Scheme

Here we adopt questionnaire-based analysis scheme for risk assessment of offshore software outsourcing, which is borrowed from the original work of US researchers [8][9]. In order to keep high response rate on questionnaire, we asked the industry members of JEITA (Japan Electronics and Information Technology Industries Association) as well as SSR (Joint Forum for Strategic Software Research) to collaborate with the authors to design the questionnaire [10]. The whole research steps include: 1) decision of the investigated attributes and levels, 2) design of questionnaire cards, 3) collection of sampling data, and 4) statistical analysis as well as risk assessment.

3.1 Decision of Attributes and Levels

Various issues affect the decision of client company for offshore software outsourcing. Referring to the previous questionnaire for US project managers, we confirm if it also works for Japan project managers at first. The authors visited five Japanese client companies and two vendor companies in October, 2005. We found that any attribute seems to have preference value for outsourcing and is not negligible for decision on outsourcing. We also found that each company has different strategy, especially on project control, for facing to vendors.

Table 1. Attributes and levels of offshore software outsourcing in questionnaire

Attributes		Two levels	Value	ID
Software Property	Software complexity and scale	Simple and small	High	S1
		Complex / large	Low	
	Software quality measurability	Easy to measure	High	S2
		Difficult to measure	Low	
	Requirement specifiablity	Easy to specify	High	S3
		Difficult to specify	Low	
	Requirement volatility	No change	High	S4
		Shall change	Low	
Vendor Property	Communication skill	Good	High	V1
		Bad	Low	
	Project management capability	Much reliable	High	V2
		Unreliable	Low	
	Vendor flexibility on specification changes	flexible	High	V3
		Not flexible	Low	
	Attrition rate	Small rate	High	V4
		Large rate	Low	
	Long term strategy	Yes	High	V5
		No	Low	
Project Property	Deadline urgency	Urgent	High	P1
		Not urgent	Low	
	Relative cost advantage	High advantage	High	P2
		Low advantage	Low	
	Client side technical expertise	Lack	High	P3
		Sufficient	Low	
	Strategic importance for future project	High	High	P4
		Low	Low	
	Ability to monitor vendor behavior	Easy to monitor	High	P5
		Difficult to monitor	Low	

Based on three kinds of theory including transaction cost theory, agency theory and knowledge-based theory and interview with skilled engineers for the feasibility on votes of projects evaluation, we prefer three properties for describing software development at last as listed in Table 1: software property with four attributes, vendor property with five attributes and project property with five attributes. Every attribute has two levels labeled by High and Low respectively. In Table 1, the right column ID is the abbreviation of attribute name used in the analysis procedure.

3.2 Design of Questionnaire Cards

Questionnaire cards are used to extract the tacit know-how knowledge of experienced project managers for risk assessment. The answer given by respondents for the feasibility of every project reveals the risk degree, which is used to identify the importance of attributes in the separate properties by conjoint analysis. Because the pre-analysis shows that it is difficult to image over five attributes at once, we classify fourteen attributes into four software attributes, five vendor attributes and five project attributes. Based on orthogonal planning, we prepare three sets of virtual projects. Table 2 is the example of the combination on vendor attributes for eight virtual projects.

Table 2. Combination on vendor attributes for virtual projects

No.	V1	V2	V3	V4	V5
1	Low	Low	High	High	Low
2	Low	Low	Low	Low	Low
3	High	Low	Low	High	High
4	High	High	Low	High	Low
5	High	High	High	Low	Low
6	Low	High	High	High	High
7	High	Low	High	Low	High
8	Low	High	Low	Low	High

An example question for vendor property is ”You will be presented with a series of eight virtual vendor profiles in Table 2. Based on your own experience and knowledge, please circle the appropriate numbers for the possibility of this project. How attractive would it be for your company to OUTSOURCE to this vendor?”. The similar questions are provided on software property and project property too. The evaluation for every virtual project is assigned by value where the choice comes from 1(low possibility for success) to 5 (high possibility for success).

3.3 Collection of Sampling Data

There are two approaches for sampling: random sampling and intentional sampling. While random sampling does not include bias, the return rate may be terrible because the contents in questionnaire are too confidential for respondents to disclose it. Then we use two channels for questionnaire delivery as written before: SSR and JEITA. Each company in SSR collected twenty responses and JEITA collected thirty responses. There are other volunteers who answer the questionnaire. Finally, we collect 175 responses. They are all Japanese client side people.

The response may be different from each other by respondents' background. For example, there may be difference in software types and vendor countries. So we include the following items in questionnaire for personal information of the respondent: number of years of IT experience, number of years of experience in current company, number of offshore projects involved in the outsourcing decision process, current position/role, vendor ISO/CMM ratings, type of software and vendor countries. These items can be used to assess and adjust the bias of respondents.

Table 3. The number of samples(NS) and the Pearson's R value(PRV)

Category		Project		Vendor		Software	
		NS	*PRV*	*NS*	*PRV*	*NS*	*PRV*
Totality	Totality	151	0.985	149	0.988	149	0.979
Vendor Country	India	34	0.984	34	0.989	34	0.988
	China	72	0.982	76	0.976	70	0.980
Software Type	Application	91	0.980	90	0.983	91	0.973
	Middleware	27	0.994	26	0.994	26	0.978
	Embedded	38	0.977	38	0.997	38	0.991
India	Application	10	0.980	10	0.985	10	0.952
	Middleware	17	0.991	17	0.989	17	0.992
	Embedded	6	0.955	6	0.989	6	0.971
China	Application	52	0.979	51	0.973	52	0.964
	Middleware	5	0.992	5	0.979	3	0.976
	Embedded	12	0.983	11	0.977	12	0.990

4 Risk Assessment of Virtual Projects

4.1 Sample Sizes and Fitness of Model

Based on questionnaire data, data analysis is done for the risk assessment on virtual projects generated by orthogonal planning. There are three sets of virtual projects. One set is described in attributes of software property, the second set is described in attributes of vendor property, and the third set is described in attributes of project property. First we calculate the partial utility of fourteen attributes of three properties using conjoint analysis method according the risk value for 26 kinds of virtual projects given by 175 respondents. We first analyze the total data. Then according to the outsourcing software type and vendor country, we make analysis of risk bias respectively. The number of samples and Pearson's R value given by conjoint analysis are listed in Table 3(Totality means all samples in this paper). All calculations are finished using the conjoint analysis module of SPSS 11.0. According to the analysis result, we know that the fitness of model is excellent for any category though the number of samples is not big and think that the analysis result is ideal and credible.

4.2 Outline of Analysis Result

Table 4, 5 and 6 show the result of conjoint analysis for three properties respectively where the samples are classified by totality, vendor country and software type. The partial utility value and average value is given by conjoint analysis and we calculate the range, maximum and minimum for risk assessment.

Table 4. Partial utility for project property

Attribute	Totality	Vendor Country		Software Type		
	Totality	India	China	Application	Middleware	Embedded
P1	-0.354	-0.467	-0.306	-0.331	-0.458	-0.299
P2	0.629	0.621	0.597	0.639	0.560	0.609
P3	0.333	0.408	0.267	0.328	0.375	0.339
P4	-0.001	-0.011	0.010	0.019	0.042	-0.109
P5	0.576	0.614	0.563	0.553	0.505	0.688
Average	2.878	2.842	2.854	2.894	2.792	2.918
Maximum	4.771	4.963	4.597	4.764	4.732	4.961
Minimum	0.985	0.721	1.111	1.025	0.852	0.875
Range	3.786	4.243	3.486	3.739	3.880	4.086

Table 5. Partial utility for vendor property

Attribute	Totality	Vendor Country		Software Type		
	Totality	India	China	Application	Middleware	Embedded
V1	0.350	0.342	0.411	0.364	0.240	0.446
V2	0.310	0.371	0.318	0.299	0.317	0.399
V3	0.276	0.327	0.261	0.280	0.279	0.270
V4	0.202	0.202	0.189	0.209	0.250	0.095
V5	0.199	0.143	0.211	0.190	0.173	0.196
Average	2.444	2.327	2.446	2.448	2.365	2.500
Maximum	3.781	3.713	3.836	3.789	3.625	3.905
Minimum	1.107	0.941	1.057	1.107	1.106	1.095
Range	2.673	2.772	2.779	2.681	2.519	2.811

For project property, the range value is bigger than 3.5, the maximum value is bigger than 4.5, the minimum value is smaller than 1 and the average value is near 2.8, which implies that project property is very weak, which could lead to big success or severe failure.

For vendor category, the range value varies between 2.5 and 3, the maximum value varies between 3.5 and 4, the minimum value is near 1 and the average value is near 2.5, which implies that vendor category is slightly weak.

For software category, the range value varies between 1.9 and 2.5, the maximum value varies between 3.7 and 4.2, the minimum value is above 1.4 and the average value is near 2.8, which shows that the respondents worry little about software property.

Table 6. Partial Utility for software property

Attribute	Totality	Vendor Country		Software Type		
	Totality	India	China	Application	Middleware	Embedded
S1	0.038	0.040	0.014	0.021	0.150	0.072
S2	0.281	0.305	0.282	0.254	0.290	0.434
S3	0.286	0.379	0.236	0.265	0.290	0.375
S4	0.464	0.445	0.414	0.457	0.450	0.408
Average	2.822	2.776	2.768	2.815	2.650	3.033
Maximum	3.891	3.945	3.714	3.813	3.830	4.322
Minimum	1.753	1.607	1.822	1.818	1.470	1.743
Range	2.138	2.338	1.893	1.994	2.360	2.579

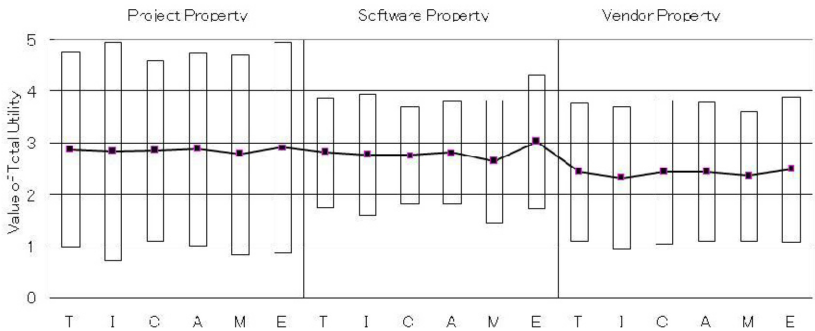


Fig. 1. Value of total utility for three properties

4.3 Analysis in View of Properties

In view of three properties, we analyze the range, maximum, minimum and average risk value for risk assessment and find some obvious characteristics.Analysis result implies that the respondents have clear orientation on three properties as plotted in Fig. 1(In the figure, T stands for totality, I stands for India,C stands for China, A stands for customer application, M stands for middleware, and E stands for embedded software).

The range risk value of project property is the biggest and that of software property is the smallest. The maximum risk value of project property is the biggest. The minimum risk value of software property is the biggest. The average risk value of vendor property is the smallest. We also found that the maximum risk value belongs to the respondents who experienced the offshore software outsourcing projects for embedded software development to India.

Among three categories, the range value of project property for risk assessment is the biggest, which shows that project management of offshore software outsourcing could lead to big success or severe failure and should be paid more attention to. The average risk value of software property is fairy high, which shows that the respondents think software property is not the main item affecting the success or failure of offshore projects.

4.4 Analysis in View of Vendor Countries

In view of vendor countries, we analyze the risk bias toward the outsourcing projects to vendors of two countries which are India and China for three software types which are customer application, middleware and embedded software. The analysis result of project property is listed in Table 7.

Table 7. Partial utility for project property according to vendor country and software type

Attribute	India			China		
	Application	Middleware	Embedded	Application	Middleware	Embedded
P1	-0.475	-0.507	-0.438	-0.293	-0.325	-0.292
P2	0.700	0.581	0.646	0.635	0.725	0.521
P3	0.425	0.419	0.313	0.274	0.425	0.208
P4	-0.050	0.022	0.021	0.039	-0.025	-0.104
P5	0.625	0.537	0.771	0.563	0.475	0.625
Average	2.800	2.757	2.979	2.894	2.575	2.854
Maximum	5.075	4.824	5.167	4.697	4.550	4.604
Minimum	0.525	0.691	0.792	1.091	0.600	1.104
Range	4.550	4.133	4.375	3.606	3.950	3.500

We find that the risk range value of India for project property and software property is bigger than that of China. We also find that the risk maximum value of India for project property and software property is bigger than that of China, and the risk minimum value of India for three properties is smaller than that of China as plotted in Fig. 2 (In the figure, T stands for totality, A stands

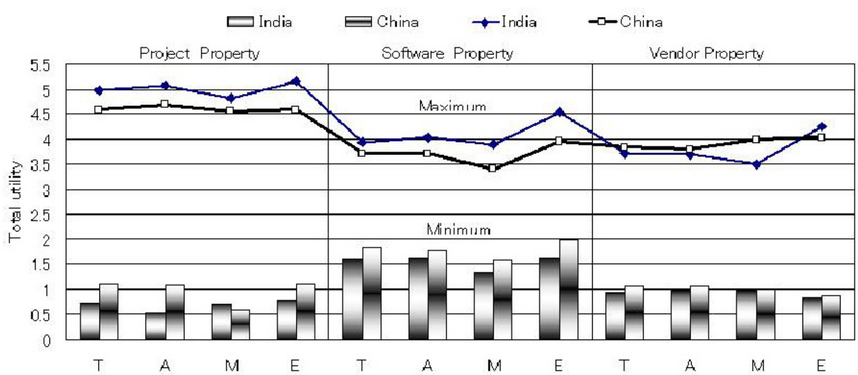


Fig. 2. Comparison of maximum and minimum in view of vendor country

for customer application, M stands for middleware, and E stands for embedded software). This shows that there is difference between India and China. Deep research is necessary for outsourcing to companies of India and China.

4.5 Analysis in View of Software Types

In view of software types, we analyze the risk bias toward the outsourcing projects to vendors of two countries which are India and China for three software types which are customer application, middleware and embedded software. The analysis result of vendor property is listed in Table 8.

Table 8. Partial utility for vendor property according to vendor country and software type

Attribute	Application		Middleware		Embedded	
	India	China	India	China	India	China
V1	0.413	0.387	0.235	0.600	0.542	0.443
V2	0.338	0.328	0.324	0.300	0.542	0.330
V3	0.313	0.265	0.324	0.150	0.375	0.330
V4	0.163	0.186	0.294	0.2500	0.042	0.125
V5	0.138	0.201	0.088	0.200	0.208	0.352
Average	2.338	2.436	2.235	2.500	2.542	2.443
Maximum	3.700	3.804	3.500	4.000	4.250	4.023
Minimum	0.975	1.069	0.971	1.000	0.833	0.864
Range	2.725	2.735	2.529	3.000	3.417	3.159

According to software type, there are some obvious characteristics. Among three kinds of software, the risk maximum value of embedded software is the biggest and the maximum value of middleware is the smallest as showed in Fig. 3. We also found the average value of middleware is the smallest.

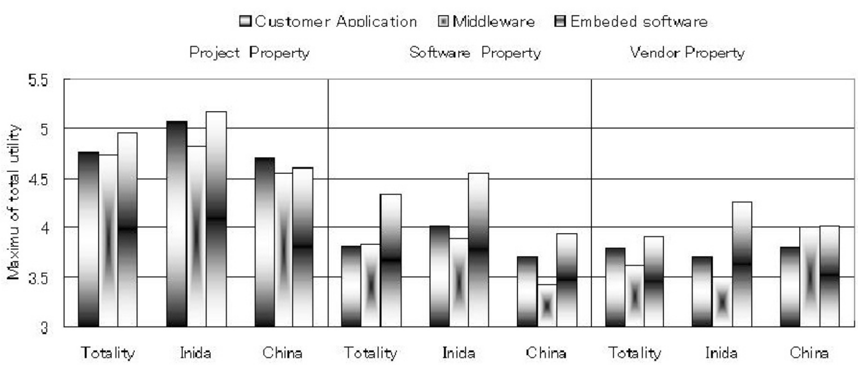


Fig. 3. Comparison of maximum in view of software type

We find that in project property for customer application software and middleware software the partial utility of *relative cost advantage* is the most important attribute among five attributes, but for embedded software the partial utility of *monitor vendor behavior* is the most important one. As for the average value of risk assessment, the difference between embedded software and middleware is big. This shows there is difference among three kinds of software.

4.6 Comparison with Importance Analysis

The importance rate of attributes for three kinds of property is also analyzed. Table 9 shows an example result of analysis for project category which includes the totality analysis result together with category according to vendor country and software type.

Table 9. Importance rate for project property

Attribute	Totality	Vendor Country		Software Type		
	Totality	India	China	Application	Middleware	Embedded
P1	18.58	22.03	17.59	17.34	24.59	17.18
P2	29.31	27.14	28.97	30.66	26.34	27.13
P3	16.03	18.18	13.91	15.57	18.63	16.33
P4	8.72	6.33	9.30	8.66	7.99	9.01
P5	26.69	26.32	28.83	26.65	22.45	30.35

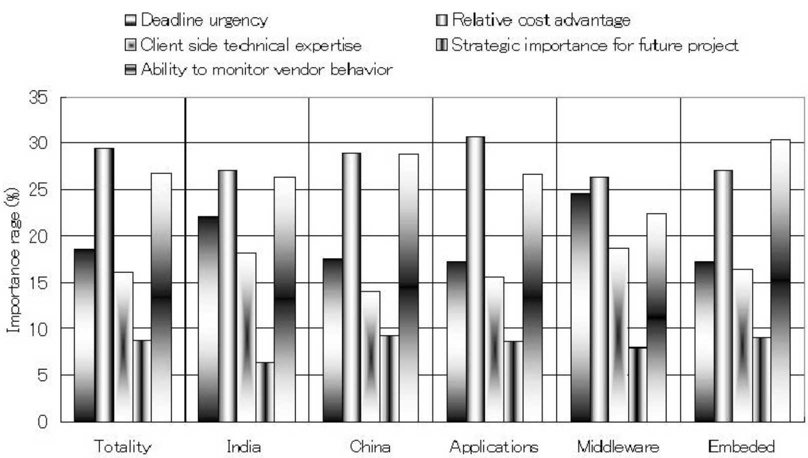


Fig. 4. Comparison of importance rate for project property

Compared with the calculate result of partial utility listed in table 4, the partial utility and importance rate of the *relative cost advantage* attribute is the biggest one among five attributes as plotted in Fig.4, which shows the *relative cost advantage* attribute is the most important attribute for project property.

The partial utility and importance rate of the *strategic importance for future project* attribute is the smallest one among five attributes and the value is fairly small, which show this attribute is paid little attention to by the respondents.

For software property, there is the same result as project property that the partial utility and importance rate of the *requirement volatility* attribute is the biggest one among five attributes, which should be paid more attention to. But for vendor category, there is difference for the calculate result between the importance rate and the partial utility. For importance rate, the *communication skill* attribute is the most important one among five attributes. For partial utility analysis, the *communication skill* attribute and *project management capability* attribute are the most important two attributes, whose order varies according to vendor country and software type.

5 Conclusions

This paper analyzes the *range*, *maximum*, *minimum* and *average* value of total utility for risk assessment of offshore software outsourcing projects, and assesses the risk bias brought by vendor countries and software types. The followings are our conclusions:

1) Project property is the main item among three properties, which could lead to big success or severe failure. The management of projects should be paid more attention to.

2) The risk analysis for vendors of India is different from that of China, which should be deeply investigated.

3) The risk value of middleware software is lower than that of the other two software types, which should be paid more attention to.

Acknowledgments. The authors would like to give sincere thanks to members of Joint Forum for Strategic Software Research (SSR) and Japan Electronics and Information Technology Industries Association (JEITA) who contributed to collect responses of questionnaire.

References

1. Aspray, W., Mayadas, F., Vardi, M.Y.(eds.): Globalization and Offshoring of Software, Report of the ACM Job Migration Task Force, Association for Computing Machinery (2006)
2. Software Business Committee, Report on Software Development Resource, Japan Electronics and Information Technology Industries Association, pp. 6–9 (2006)
3. Tsuji, H., et al.: Questionnaire-based Risk Assessment Scheme for Japanese Offshore Software Outsourcing. In: First International Conference on Software Engineering Approaches For Offshore and Outsourced Development, pp. 1–15 (Zurich 2007)

4. Wada, Y., Nakahigashi, D., Tsuji, H.: An Evaluation Method for Offshore Software Development by Structural Equation Modeling. In: First International Conference on Software Engineering Approaches For Offshore and Outsourced Development, pp. 1–14 (Zurich 2007)
5. Andrews, R.L., Ansari, A., Currim, I.S.: Hierarchical Bayes Versus Finite Mixture Conjoint Analysis Models: A Comparison of Fit, Prediction, and Partworth Recovery. Journal of Marketing Research XXXIX, 87–98 (2002)
6. Moskowitz, H.R., Silcher, M.: The applications of conjoint analysis and their possible uses in Sensometrics. Food Quality and Preference 17, 145–165 (2006)
7. Wittink, D.R., Cattin, P.: Commercial Use of Conjoint Analysis: An Update. Journal of Marketing 53(3), 91–96 (1989)
8. Tiwana, A.: Beyond the Black Box: Knowledge Overlaps in Software Outsourcing. IEEE SOFTWARE 21(5), 51–58 (2004)
9. Tiwana, A., et al.: Myths and Paradoxes in Japanese IT Outsourcing, forthcoming in Communications of the ACM (2006)
10. Tsuji, H., et al.: Risk Factors on Offshore Software Development by Conjoint Analysis (In Japanese) IPSJ Journal 47(2), 823–831 (2007)

Extracting Failure Knowledge with Associative Search

Masayuki Nakao[1], Kensuke Tsuchiya[1], Yoshiaki Harita[2], Kenji Iino[3],
Hiroshi Kinukawa[4], Satoshi Kawagoe[4], Yuji Koike[5], and Akihiko Takano[5]

[1] Department of Engineering Synthesis, The University of Tokyo,
Hongo 7-3-1, Bunkyo-ku, Tokyo, Japan 113-8656
[2] Flow Net Corp.
[3] Sydrose LP
[4] Tokyo Denki University
[5] National Institute of Informatics
nakao@hnl.t.u-tokyo.ac.jp

Abstract. We applied associative OR search on a Failure Knowledge Database with 1,242 failure accidents and 41 failure scenarios in the book “100 Scenarios of Failure” to find cases most analogous to risks that engineers were concerned with. Ninety engineers provided 203 input cases of risk concerns and the search for accidents most analogous to each input returned the most analogous accidents for 64% of the input cases of risk concerns within several minutes. Analogous scenario searches returned the most analogous scenarios for 63% of the input. Regular keyword AND searches take tens of minutes to narrow down the candidates to a few analogous cases, and thus associative search is a more effective tool for risk management.

Keywords: Associative Search, Failure Knowledge, Risk Management.

1 Introduction

Predicting future accidents from past failure knowledge can minimize the loss from risk. Finding the most effective cases from past knowledgebase for the current risk concern, however, is not an easy task. The authors participated in building past knowledgebase; “Failure Knowledge Database” (a project by Japan Science and Technology Agency (JST) that started in 2001 and published in 2006), and “100 Scenarios of Failure” (a book by Morikita Publishing Co., Ltd. 2005). The former collected 1,242 accident cases from fields of mechanical engineering, material science, chemical engineering, and civil engineering, and published them on the INTERNET (http://shippai.jst.go.jp). Nakao, Harita, and Iino contributed 485 cases from the mechanical engineering field. Harita and Iino collected almost 300 cases of severe accidents for the book “100 Failure Scenarios of Failure” and Nakao extracted 41 failure scenarios from them. The scenarios included technical causes like fracture, fatigue, or corrosion, as well as human or organizational factors like input error, removing safety tool, or lack of communication.

Finding knowledge, however, from these large accumulations of data for eliminating a risk concern that one is currently faced with is not easy. Nakao has tried giving assignments of finding useful knowledge from these databases to attendees of

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 269–276, 2007.

design seminars at corporations or universities, however, only 30 to 50% of them succeeded in extracting such knowledge within 1 hour. The reason was AND searches made typically with abstract words about failure that returned a large number of candidate cases, whereas when made with special terms related to specific products or process, the searches found hardly any matching cases.

For example, the top 10 keywords used in searching the Failure Knowledge Database up to that time were “defect”, “chemical”, “explosion”, “gas”, “leakage”, “piping”, “environment”, “fire”, “education”, and “corrosion” in this order, however, a search with any of these keywords will return a large number of cases; for example “defect” returned 986 cases, i.e., 79% of the entire database. Further adding the above abstract words for an AND search fails in significantly cutting down the number of hits because they are all common in chemical accidents. Then adding a specific word like “tetrafluoroethylene” returns only the case the searcher had entered in the database. Struggling with keywords to find cases analogous to a specific one quickly wastes time. A similar task is a search for patents that a new idea may infringe. Any engineer must have had such an experience of taking a whole day to a entire week in finding matches.

On the other hand, those in charge of plant safety often give speeches to bring up the safety awareness among the workers, and the speaker knows that introducing an actual case is way more effective than to talk about abstract safety rules. Effective cases for such speeches are those with analogous causes and sequences but with different products or fields. For example, servicemen that remove the safety fences because they interfere with maintenance work on assembly robots would also take off safety goggles when servicing laser equipment. Both cases share the same superordinate concept of removing safety tools. The computer, however, takes the words “robot” and “laser” as completely different words and thus its keyword search never finds the analogy between the two risks. Advancements in language processing are overcoming this difficulty by setting all the words from the input sentences as keywords and extracting the most analogous case by finding the case with the most number of hits with all the keywords. This, in other words, is an OR search with all words in a set of sentences. The computer a priori analyzes all its case data documents about the words that compose them, and once the searcher inputs a set of sentences, the computer adds up all the word hits and returns an output of the similar cases within a matter of seconds.

Takano et al [1] developed such an OR search software GETA (Generic Engine for Transposable Association). The word “Association” came from its encyclopedia search that brings up a word after another as if the computer is prompting the user about related knowledge, e.g., “Hiroshima”, “Atomic bomb”, “Enola Gay”, and so on. We assumed that if we apply this search to failure knowledgebase, it would automatically bring up cases analogous to the user input of risk concerns like “safety fence”, “removing safety tool”, “safety goggles” to aid the searcher in finding countermeasures to the risk concern.

GETA is designed as an effective tool for providing generic computation for association. It manipulates very large sparse matrices, which typically appear as index files for the large scale text retrieval. By providing the basic operations on this matrix, GETA enables the quantitative analysis of various proposed methods based on association, such as co-occurrence word analysis, topic word extraction or similarity measurement among documents. Using the parallel processing, its associative research is feasible for about 20 million documents within a few seconds. The GETA-based computation has been commercially used in the above-mentioned encyclopedia search service over the internet since 1998, and bio-Informatics DB search since 2000. GETA, which was mainly designed by Nishida and supported by Information Promotion Agency (IPA), was released as an open source software in 2002 (http://geta.ex.nii.ac.jp/).

IMAGINE is a software that executes the GETA-based retrieval with multi-database (DB), which was developed by Takano. So far, he has installed newspaper articles, encyclopedia, Wikipedia, book's table of contents, and so on in IMAGINE (http://imagine.bookmap.info/). Its user inputs a key document to IMAGINE and wants to search the most similar document from the multi-DB. GETA extracts summarized words from the given key document; and collects some documents from the target DB, which are relevant to the set of words in the key document; and finally ranks them in the order of higher similarity. The user notices mutual association between the key inputted document and the relevant outputted information, analyzing and discovering the proper judgment.

The authors applied GETA/IMAGINE to search the failure knowledgebase "Failure Knowledge Database" and "100 Scenarios of Failure" for knowledge analogous to the input case of risk concerns, and confirmed that it points to effective knowledge with a high rate of success.

2 Method of Experiment

We used IMAGINE illustrated in Figure 1. We added knowledge data from "Failure Knowledge Database" and "100 Scenarios of Failure" as our knowledgebase.

Japan Society of Mechanical Engineers and JST jointly conducted seminars for our experiment. We ran a total of 5 sessions from October of 2006 to February of 2007. The total number of participants was 90, mostly students and engineers in the mechanical engineering field. The participants input altogether 203 cases of risk concerns or accident reports of their own or of their companies. We had the participants type the input into an Excel worksheet and each case information was split into sections of "Incident", "Sequence", "Cause", "Countermeasures", "Knowledge comment" like the descriptions in Failure Knowledge Database.

The search engine GETA weighs the parts of speech differently by giving higher scores to hits with proper nouns and lower scores to prefixes and adverbs. This feature caused search results, when proper nouns were placed in the "Event" section, with high score with the same geographic location or the same company name whilst the scenarios were totally unrelated. Instead if we just input the text from the "Cause" or "Sequence" sections usually described with general words, the searches returned effective cases with analogous scenarios although the products or fields may differ. Also, using technical terms in the input text instead of general terms tended to find effective data. The reason for this behavior is because engineers edited the knowledgebase entries, thus, such words as "corrosion" or "velocity" saw more hits than "rust" or "speed".

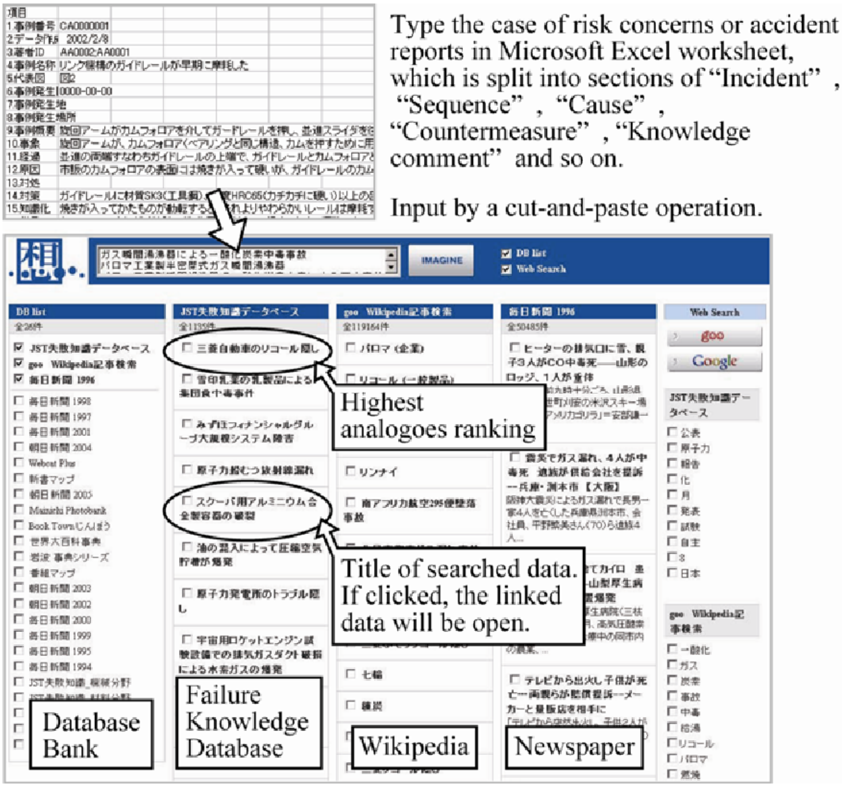


Fig. 1. Associative search engine with databases, "IMAGINE", which searches the analogous cases of the input case from the database bank

We did not use a thesaurus for our searches, however, if we added the words listed in Wikipedia entries to the OR searches, we often found a list of effective knowledge. For example, first looking up "welding" in Wikipedia, and adding its description to the input for our OR search produced results similar to using a thesaurus; it found further cases about "Welding Robots" and "Arc Welding".

Our experiment required a judgment criteria of whether the output data are really analogous to the input data, and for our study, Nakao and Harita read all the results to make judgments with the same criteria.

3 Experiment Results

3.1 Analogy Search on Failure Knowledge Database

For this experiment, an ideal result would be finding a case of an arm getting pulled into the snow inlet of a snow blower from an input case of fingers getting caught into the paper entrance of a paper shredder. These two accidents share the similarity of getting fingers or arms caught when trying to push the paper or snow into the machines when they are jammed.

A search with IMAGINE lists titles of 16 cases with high scores and the participants typically open the first 5 cases and glance through the description to find most analogous cases. As we explained above, they also tried trimming the input to only the "Cause" section or adding synonyms from the encyclopedia to acquire a larger list of candidates, and after they experienced several trial and error, they were able to pull up the most analogous case within several minutes.

Among the most analogous accident searched, 84% were found in the Failure Knowledge Database. For the remaining 16%, the input cases described minor accidents or incidences related to management or patents and they found analogous cases in the newspaper articles.

Figure 2 shows the search results of analogous accident cases. Among the 90 participants, 56 (62%) found analogous cases as shown in Figure 2(d), and of the 134 risk concerns, 86 (64%) had the most analogous accidents searched (Figure 2(a)). On the average, each line of text had 33 Japanese characters and when the input text had 24 or more lines, 80% of them found the most analogous accidents (Figure 2(b)), however, when less that 24 lines, only 49% did (Figure 2(c)). In other words, the probability of finding an analogous case is higher with more input information.

We conducted a survey with the participants after the seminars and 81% of them thought they successfully found analogous cases (Figure 2(e)), 88% thought associative search was more effective than keyword search, and 88% wanted to use it on their jobs. The authors, however, through their instructions found that those who can alter the input to longer text using words of similar concepts were those under the age of 30 years old who already had skills in searching with the computer.

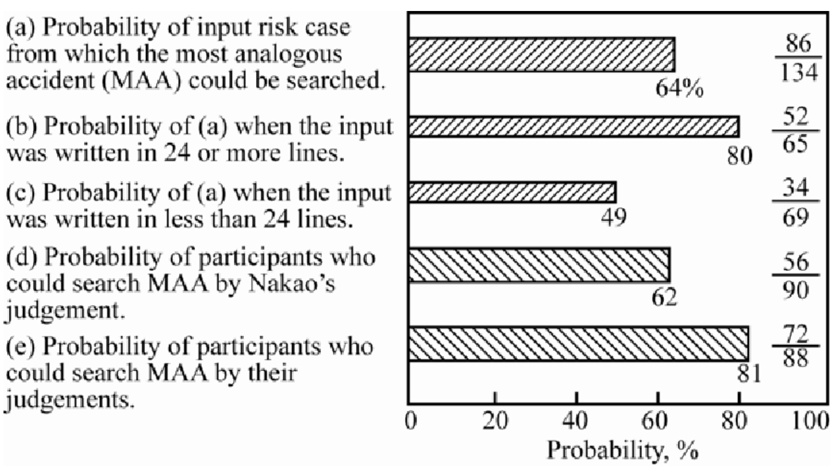


Fig. 2. Results of the seminar of "extracting failure knowledge with associative search"

The reason why more participants (81%) thought they were successful while the instructors judged fewer did (62%), was because the participants thought they found a match when the results of the accidents matched. A typical example is when an input of “carbon-monoxide poisoning due to a ruptured piping of a petroleum fan heater” returned “carbon-monoxide poisoning from a building fire” the participants thought it was a good match, however, even with the same results, the scenario was different and studying about building fires would not lead to eliminating an error in designing a petroleum fan heater.

3.2 Analogy Search on “100 Scenarios of Failure”

Consider an input case of getting fingers caught in the inlet of a paper shredder because of removing the safety cover or disabling the finger sensor to avoid frequent false alarms. An ideal match for this case is a failure scenario of “removing the safety tool”.

On the other hand, if the designer forgot to install safety covers or sensors, the input should return a scenario of “Incomplete safety measures,” however, such novice mistakes of forgetting necessary safety measures are not listed in “100 Scenarios of Failure.” The book collected well-known major failures that made it to the newspaper; minor errors that should be caught during design reviews or product tests are not in its failure scenarios. Such errors include, space interference, dimensional variation, thermal deformation, and careless mistakes.

Figure 3 shows the analogy ranking which IMAGINE decided, of the most analogous scenario that Nakao and Harita read and judged to be the most analogous match. There are some differences among individual judgment of “which scenario is most analogous to the input case of risk concerns”. There were differences in the opinions of Nakao and Harita in our study. The two instructors picked the same scenario for 67% of the input risk concerns, and for 18% of the input, close decisions had to be made between two scenarios, e.g., fatigue and resonance. Adding the two situations, 85% cases results in similar judgments by the two instructors, and there remained about 15% of noise effect caused by different instructors.

As the figure shows, IMAGINE ranked the most analogous scenario judged by Nakao the first for 71 cases (82 for Harita), i.e., it picked out the right match as much as about 35% (40%). The percentage of the right match in the second place and so on gradually decrease with a long tail in the probability density.

If we say that when the most analogous scenario is listed within the 5th place, then the search was successful, then IMAGINE succeeded in finding analogous cases for 63% (66%) of the input, and failed for the remainder.

For some cases, the most analogous scenario was listed in the 6th place or worse. The low ranking was due to noise by such nouns that frequently appear in accident reports as welding, tank, fire, manual, or sound. These nouns applied noise to the analogy values. When we took closer looks at the 40 cases that did not even make it to the 16th place, 10 of them were cases for which Nakao found analogous scenarios whereas the computer failed. For these cases the input sentences were short and written without technical terms, thus the search did not have the right clues. For the remaining 30, Nakao and the computer both failed to find analogous cases. They were cases of human error like design error or poor planning, that were not included in “100 Scenarios of Failure”, in other words, the right match did not exist in the database to begin with.

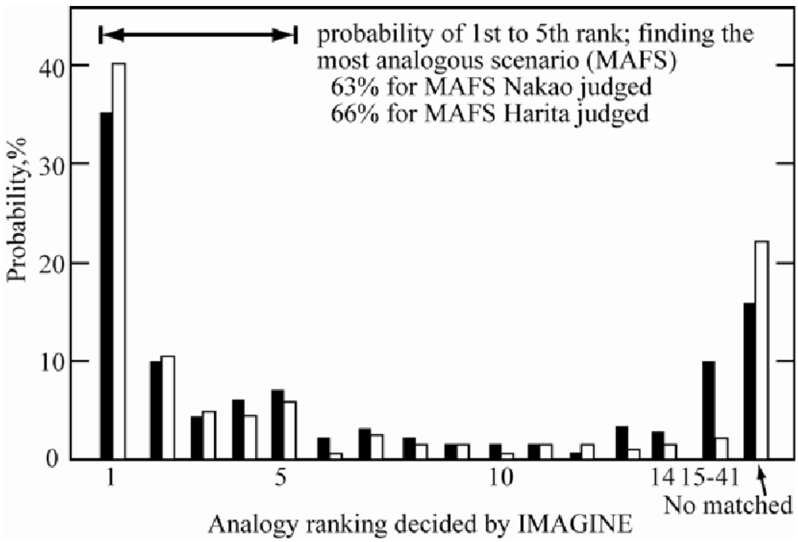


Fig. 3. Probability distribution of analogy ranking, which was decided by IMAGINE, of the most analogous failure scenario (MAFS) which was judged by Nakao (*black bar*) and Harita (*white bar*)

4 Discussions

Our results demonstrate that associative search extracts, although it is a qualitative argument, the righter results in shorter time compared to keyword search. The authors experience with larger databases with over 10,000 entries, is that a keyword search can also succeed in finding analogous cases, but with smaller knowledge bases, associative search has better success. The databases we studied were small and gave better results with associative search.

The chance of successfully finding an analogous match, however, was largely affected by uneven distribution of the data. Editors of the Failure Knowledge Database we used did not collect much cases in the electrical, semiconductor, medical, or food industries, and thus, such input cases often failed to find matches. Management or patent cases were also excluded from the knowledgebase and such input had to search the newspaper sections for analogous cases. The cases in “100 Scenarios of Failure” are also distributed unevenly. As we mentioned above design errors and human errors that caused minor injuries were not collected for the book. Also, such mistakes in planning that lead to no bodily damage or no criminal case are hard to recognize as failures and thus are difficult for editors to collect information about them.

5 Conclusions

We applied abstract associative OR search on the 1,242 accident cases in Failure Knowledge Database and on the 41 failure scenarios in “100 Scenarios of Failure” to find the most analogous case to a risk concern input in text. 90 engineers edited 203 cases of their own risk concerns and searched for the most analogous failure accident or scenario. The search succeeded in finding the analogous case: accident for 64%, scenario for 63% of the risk input within several minutes. Conventional keyword AND search takes several 10s of minutes to narrow down the most analogous case in the candidates. Associative search is effective for risk management.

We are grateful to JST who supported our study.

Reference

1. Takano, A.: Association Computation for Information Access. In: Grieser, G., Tanaka, Y., Yamamoto, A. (eds.) DS 2003. LNCS (LNAI), vol. 2843, pp. 33–44. Springer, Heidelberg (2003)

Data Mining Analysis of Relationship Between Blood Stream Infection and Clinical Background in Patients Undergoing Lactobacillus Therapy

Kimiko Matsuoka[1], Shigeki Yokoyama[2], Kunitomo Watanabe[3], and Shusaku Tsumoto[4]

[1] Osaka General Medical Center, Osaka 558-8558, Japan
cby41060@pop01.odn.ne.jp
[2] Koden Industry Co., Ltd, Tokyo 143-0015, Japan
shigey@st.rim.or.jp
[3] Gifu University, Gifu 501-1194, Japan
kuni@cc.gifu-u.ac.jp
[4] Shimane University, Izumo 693-8501, Japan
tsumoto@computer.org

Abstract. The aim of this study is to analyze the effects of lactobacillus therapy and the background risk factors on blood stream infection in patients by using data mining. The data were analyzed by data mining software, i.e. "*ICONS Miner*" (Koden Industry Co., Ltd.).The significant "If-then rules" were extracted from the decision tree between bacteria detection on blood samples and patients' treatments, such as lactobacillus therapy, antibiotics, various catheters, etc. The chi-square test, odds ratio and logistic regression were applied in order to analyze the effect of lactobacillus therapy to bacteria detection. From odds ratio of lactobacillus absence to lactobacillus presence, bacteria detection risk of lactobacillus absence was about 2 (95%CI: 1.57-2.99). The significant "If-then rules", chi-square test, odds ratio and logistic regression showed that lactobacillus therapy might be the significant factor for prevention of blood stream infection. Our study suggests that lactobacillus therapy may be effective in reducing the risk of blood stream infection. Data mining is useful for extracting background risk factors of blood stream infection from our clinical database.

Keywords: Lactobacillus Therapy, Data Mining, Medical Risk Management.

1 Introduction

Data mining is important for all medical organizations that utilize large data sets, any medical organization with large volumes of clinical data, huge patient databases, or clinical helpdesk service records can benefit from this field [1], [2].

To assess the incidence of blood stream infection and the effect of detection of this infection patient's care at Osaka General Medical Center, we analyzed the data on positive blood cultures [3], [4], [5]. The subject of the previous studies was the clinical background of the patients who underwent blood culture test. We analyzed the relationship between the patient's diseases etc. and the anaerobes by using data mining.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 277–288, 2007.

Usually, there are various bacteria in human body, such as, intestinal tract, oral cavity, and skin etc. And those bacteria form normal bacterial flora. While human maintains good health, the bacteria flora will keep balance and will act as a barrier against infection.

However, by external factor, such as medication, stress and diarrhea etc, the normal bacteria flora gets off balance, and loses a barrier against infection. Consequently, bacteria may invade into blood stream of human body. This bacteria invasion into blood stream is called bacterial translocation, which may cause blood stream infection [6], [7]. It was reported that bacterial translocation might be caused by antibiotics administration [8] and intravenous hyper-alimentation（IVH）[9].

In our medical center, lactobacillus therapy (probiotic product) has been used for patients' recovery from surgery and prevention of postoperative infection, since early 1990s. Currently, lactobacillus preparation is used in the most departments of our center.

This study was conducted to analyze the effects of lactobacillus therapy and the background risk factors of blood stream infection in patients by using data mining techniques.

From the blood culture samples of patients on blood stream infection, various bacteria have been usually detected. Therefore, there might be the relationships between various treatments (lactobacillus therapy, drug, catheter, etc.) and bacteria detection on positive blood cultures.

2 Purpose and Subjects

2.1 Purpose

For the prevention of blood stream infection, we analyzed the effects of lactobacillus therapy and the background risk factors of bacteria detection on blood cultures.

For the purpose of our study, we used the clinical data collected from the patients, such as laboratory results, isolated bacterium, anti-biotic agents, lactobacillus therapy, various catheters, departments, and underlying diseases, etc. Table 1 shows the attributes of the hospital infection control data.

2.2 Subjects

The population for this study consisted of 1291 patients with blood stream infection who were admitted to our center between January and December, 2002.

2.3 Analytical Methods

As the analytical methods, we used decision tree, chi-square test and logistic regression. Data mining software ("ICONS Miner", Koden Industry Co., Ltd.) was used in our study.

"If-then rules" were extracted from the decision trees. The chi-square test and logistic regression were applied in order to analyze the effect of lactobacillus therapy.

The subjects were divided into two groups by the absence or presence of lactobacillus therapy. Lactobacillus group patients were administrated lactobacillus preparation or yoghurt within 5 days from microbial detection in blood cultures, and control group patients never took those preparations.

Table 1. Attributes of infectious diseases database

Item	Attributes (63)
Patient's Profile	ID, Gender, Age
Department	Department, Ward, Diagnosis(3)
Order	Background Diseases, Sampling Date, Sample, No.
Symptom	Fever, Cathether(5), Traheotomy, Endotracheal intubation, Drainage(5)
Examination Data	CRP, WBC, Urin data, Liver/Kidney Function, Immunology
Therapy	Antibiotic agents(3), Steroid, Anti-cancer drug, Radiation Therapy, Lactobacillus Therapy
Culture	Colony count, Bacteria, Vitek biocode, β-lactamase
Susceptibility	Cephems, Penicillins, Aminoglycoside, Macrolides, Carbapenums, Chloramphenicol,Rifanpic, VCM, etc.

3 Results

3.1 Chi-Square Test and Odds Ratio

Chi-square test was applied to evaluate the association between lactobacillus therapy and blood stream infection (bacteria detection on blood cultures). Table 2 shows the cross table of the bacteria detection on blood samples and the lactobacillus therapy. In this cross table, its p-value was $0.000000159 < 0.01$. Therefore, the effect of lactobacillus presence to lactobacillus absence was considered statistically significant. Odds ratio was calculated as the relative risk of lactobacillus absence to lactobacillus presence.

Probability of bacteria detection on lactobacillus absence is $p = 247/914$. Probability of bacteria detection on lactobacillus presence is $q = 55/377$. Odds ratio (OR) of lactobacillus absence to lactobacillus presence is given by

$$OR = \frac{\frac{p}{1-p}}{\frac{q}{1-q}} \tag{1}$$

From (1), odds ratio of the lactobacillus absence, OR = 2.17. And the 95% CI (confidence interval) was between 1.57 and 2.99. As the bacteria detection risk of lactobacillus absence was about 2 (95%CI: 1.57-2.99) to lactobacillus presence, lactobacillus therapy might be significantly effective to prevent the bacteria detection on blood sample.

Thus, these results showed that lactobacillus therapy might have the effect to the prevention of blood stream infection.

Table 2. Contingency table of bacteria detection and lactobacillus therapy

	Lactobacillus therapy		**Total**
	N (Absence)	**Y (Presence)**	
Bacteria Y (Detection)	**247**	**55**	**302**
Bacteria N (No detection)	**667**	**322**	**989**
Total	**914**	**377**	**1291**

3.2 Decision Tree

The following decision tree was obtained as the relationship between the bacteria detection and the various factors, such as diarrhea, lactobacillus therapy, antibiotics, surgery, tracheotomy, CVP/IVH catheter, urethral catheter, drainage, other catheter, etc. (See Fig.1a, and Fig.1b).

Fig.1a shows the sub-tree of the decision tree on lactobacillus therapy = Y (Y means its presence.). And Fig.1b shows the sub-tree of the decision tree on lactobacillus therapy = N (N means its absence).

The target variable of the decision tree is bacteria(Y/N). The explanatory variables of the decision tree are diarrhea(Y/N), lactobacillus therapy(Y/N), antibiotics(Y/N), surgery(Y/N), tracheotomy(Y/N), CVP/IVH catheter(Y/N), urethral catheter(Y/N), drainage(Y/N), and catheter(Y/N).

The first node of the decision tree is lactobacillus therapy(Y/N). Therefore, lactobacillus therapy might be the most significant factor for prevention of blood stream infection.

In the sub-tree on lactobacillus therapy(Y/N) = Y (Fig.1a), the second branch is diarrhea(Y/N), and the third branch is catheter(Y/N).

On the other hand, in the sub-tree on lactobacillus therapy(Y/N) = N (Fig.1b), the second branch is tracheotomy(Y/N), and the third branch is diarrhea(Y/N) or surgery(Y/N).

The decision tree showed that bacteria(Y/N) have the strong relationship with lactobacillus therapy(Y/N), diarrhea(Y/N), catheter(Y/N) and tracheotomy(Y/N), etc.

3.3 If-Then Rules from the Decision Tree

The following significant "If-then rules" were extracted from the above decision tree between the bacteria detection(Y/N) and the various factors.

Decision tree start

```
Lactobacillus therapy (Y/N) = Y:
  Diarrhea (Y/N) = N:
    Catheter (Y/N) = N: Bacteria : N (74->71)
    Catheter (Y/N) = Y:
      Surgery (Y/N) = N: Bacteria : N (173->149)
      Surgery (Y/N) = Y:
        CVP/IVH (Y/N) = Y: Bacteria : N (12->12)
        CVP/IVH (Y/N) = N:
          Drainage (Y/N) = N: Bacteria : N (17->13)
          Drainage (Y/N) = Y:
            Urethral catheter(Y/N) = N: Bacteria : N (3->2)
            Urethral catheter(Y/N) = Y: Bacteria : Y (2->2)
  Diarrhea (Y/N) = Y:
    Catheter (Y/N) = Y: Bacteria : N (75->63)
    Catheter (Y/N) = N:
      Anti-biotics (Y/N) = N: Bacteria : N (7->7)
      Anti-biotics (Y/N) = Y: Bacteria : Y (14->9)
```

Fig. 1a. Sub-tree on lactobacillus therapy(Y/N) = Y

```
Lactobacillus therapy (Y/N) = N:
  Tracheotomy (Y/N) = N:
    Diarrhea (Y/N) = N:
      Anti-biotics (Y/N) = Y: Bacteria : N (428->356)
      Anti-biotics (Y/N) = N:
        Catheter (Y/N) = N: Bacteria : N (170->129)
        Catheter (Y/N) = Y:
          Surgery (Y/N) = N:
            Drainage (Y/N) = Y: Bacteria : N (9->7)
            Drainage (Y/N) = N:
              Urethral catheter(Y/N) = N: Bacteria : N (124->75)
              Urethral catheter(Y/N) = Y:
                CVP/IVH (Y/N) = N: Bacteria : Y (27->16)
                CVP/IVH (Y/N) = Y: Bacteria : N (18->11)
          Surgery (Y/N) = Y:
            Drainage (Y/N) = N: Bacteria : N (5->5)
            Drainage (Y/N) = Y: Bacteria : Y (6->4)
    Diarrhea (Y/N) = Y:
      Drainage (Y/N) = Y: Bacteria : N (14->13)
      Drainage (Y/N) = N:
        Urethral catheter(Y/N) = N:
          CVP/IVH (Y/N) = Y: Bacteria : Y (17->12)
          CVP/IVH (Y/N) = N:
            Anti-biotics (Y/N) = N: Bacteria : Y (15->8)
            Anti-biotics (Y/N) = Y: Bacteria : N (15->9)
        Urethral catheter(Y/N) = Y:
          Anti-biotics (Y/N) = N:
            Surgery (Y/N) = N: Bacteria : N (5->5)
            Surgery (Y/N) = Y: Bacteria : Y (3->2)
          Anti-biotics (Y/N) = Y:
            CVP/IVH (Y/N) = N: Bacteria : N (2->2)
            CVP/IVH (Y/N) = Y: Bacteria : Y (12->7)
  Tracheotomy (Y/N) = Y:
    Surgery (Y/N) = N: Bacteria : N (19->16)
    Surgery (Y/N) = Y:
      Diarrhea (Y/N) = Y: Bacteria : N (2->2)
      Diarrhea (Y/N) = N:
        Drainage (Y/N) = Y: Bacteria : Y (13->12)
        Drainage (Y/N) = N:
          Anti-biotics (Y/N) = Y: Bacteria : N (3->3)
          Anti-biotics (Y/N) = N:
            CVP/IVH (Y/N) = N: Bacteria : N (2->2)
            CVP/IVH (Y/N) = Y: Bacteria : Y (5->5)
```

Decision tree end

Fig. 1b. Sub-tree on lactobacillus therapy(Y/N) = N

Fig.1a shows the sub-tree on lactobacillus therapy presence. In case of lactobacillus therapy, it was considered that there might be trivial risk of bacteria detection from blood samples by the above odds ratio analysis. Therefore, the following significant "If-then rules" were extracted from the sub-tree in Fig.1a.

If-then rule -1:
If Lactobacillus therapy (Y/N) = Y and
Diarrhea (Y/N) = N and Catheter (Y/N) = Y and
Surgery (Y/N) = Y and CVP/IVH (Y/N) = Y,
then Bacteria = N. (1.00 = 12/12) (2)

If-then rule-1, (2) showed that lactobacillus therapy presence might prevent bacteria detection from blood sample when patient has not diarrhea and has central venous pressure (CVP) catheter and intravenous hyper-alimentation (IVH) catheter after the surgery.

If-then rule -2:
If Lactobacillus therapy(Y/N) = Y and
Diarrhea (Y/N) = Y and Catheter (Y/N) = Y,
then Bacteria = N. (0.84 = 63/75) (3)

By (3), it was considered that lactobacillus therapy presence might prevent bacteria detection from blood sample when patient has diarrhea and catheter inserted into the blood vessel. That is, even though patient has diarrhea, lactobacillus therapy might protect patient's normal bacterial flora.

If-then rule -3:
If Lactobacillus therapy (Y/N) = Y and
Diarrhea (Y/N) = Y and Catheter (Y/N) = N and
Antibiotics (Y/N) = Y,
then Bacteria = Y. (0.64 = 9/14) (4)

If-then rule-3, (4) showed that lactobacillus therapy presence might not prevent bacteria detection from blood sample when patient has diarrhea and has no catheter and anti-biotics. When patient might have diarrhea by anti-biotics, lactobacillus therapy could not protect patient's normal bacterial flora.

Fig.1b shows the sub-tree on lactobacillus therapy absence. In case of non lactobacillus therapy, it was considered that there might be some risk of bacteria detection from blood samples by the above odds ratio analysis.

Therefore, the following significant "If-then rules" were extracted from the sub-tree in Fig.1b.

If-then rule -4:
If Lactobacillus therapy (Y/N) = N and
Tracheotomy (Y/N) = Y and Surgery (Y/N) = Y and
Diarrhea (Y/N) = N and Drainage (Y/N) = Y,
then Bacteria = Y.
(Confidence:0.92 = 12/13) (5)

If-then rule-4, (5) showed that lactobacillus therapy absence might not prevent bacteria detection from blood sample when patient has tracheotomy, no diarrhea, central venous pressure (CVP) catheter, intravenous hyper-alimentation (IVH) catheter and drainage after the surgery.

If-then rule -5:
If Lactobacillus therapy (Y/N) = N and
Tracheotomy (Y/N) = Y and Surgery (Y/N) = Y and
Diarrhea (Y/N) = N and CVP/IVH (Y/N) = Y and
Anti-biotics (Y/N) = N,
then Bacteria = Y.
(Confidence1.00 = 5/5) (6)

By (6), it was considered that lactobacillus therapy absence might not prevent bacteria detection from blood sample when patient has tracheotomy, no diarrhea, central venous pressure (CVP) catheter, intravenous hyper-alimentation (IVH) catheter and no anti

If-then rule -6:
If Lactobacillus therapy (Y/N) = N and
Tracheotomy (Y/N) = N and Diarrhea (Y/N) = N and
Antibiotics (Y/N) = Y,
then Bacteria = N.
(Confidence: 0.83 = 428/356) (7)

If-then rule-6, (7) showed that bacteria detection from blood sample might be prevented by anti-biotics when patient has lactobacillus therapy absence, no tracheotomy and no diarrhea.

From these rules, there might be the strong relationship between treatment (tracheotomy, surgery, etc.) and bacteria detection from blood samples in case of lactobacillus therapy absence.

3.4 Logistic Regression

Using the above same target variable and the same explanatory variables, logistic regression was applied to analyze the relationship between bacteria detection and lactobacillus therapy, anti-biotics, etc.

Target variable (W) is bacteria detection(Y/N) and explanatory variables (X_1, X_2, , , X_9) are lactobacillus therapy(Y/N), anti-biotics(Y/N), , , diarrhea(Y/N).(see Table 3).

This logistic regression model is given by the next equation:

$$\log\left(\frac{W}{1-W}\right) = B_1X_1 + B_2X_2 + \cdots + B_8X_8 + B_9X_9 + B_0 \qquad (8)$$

Table 3 shows the analytical result of this logistic regression equation. In Table 3, lactobacillus therapy(Y/N) has the biggest absolute value among the standard partial regression coefficients. Therefore, lactobacillus therapy might be the most significant factor for prevention of blood stream infection.

Table 3. Logistic regression results

	Explanatory Variables	p-Value	Standardized Partial Regression Coefficient	
X_1	**Lactobacillus therapy (Y/N)**	**0.0000002**	B_1	**-0.386**
X_2	**Anti-biotics(Y/N)**	**0.0000012**	B_2	**-0.321**
X_3	**CVP/IVH catheter(Y/N)**	**0.718**	B_3	**0.029**
X_4	**Urethral catheter(Y/N)**	**0.842**	B_4	**0.016**
X_5	**Drainage (Y/N)**	**0.288**	B_5	**0.079**
X_6	**Catheter (Y/N)**	**0.054**	B_6	**0.164**
X_7	**Tracheotomy (Y/N)**	**0.034**	B_7	**0.148**
X_8	**Surgery (Y/N)**	**0.641**	B_8	**-0.036**
X_9	**Diarrhea (Y/N)**	**0.000217**	B_9	**0.245**

The top five rankings of these absolute values of the standardized partial regression coefficient in Table 3 include lactobacillus therapy(Y/N), antibiotics(Y/N), diarrhea (Y/N), catheter(Y/N), and tracheotomy(Y/N).

Therefore, the result of the logistic regression showed that bacteria detection have strong relationship with lactobacillus therapy, anti-biotics, diarrhea, catheter, and tracheotomy.

These explanatory variables from this logistic regression analysis are very similar in ranking to the high order branch variables of the above decision tree. However, logistic regression analysis could not indicate the analytical results, which were constructed by the combination of explanatory variables as If-then rules from decision tree.

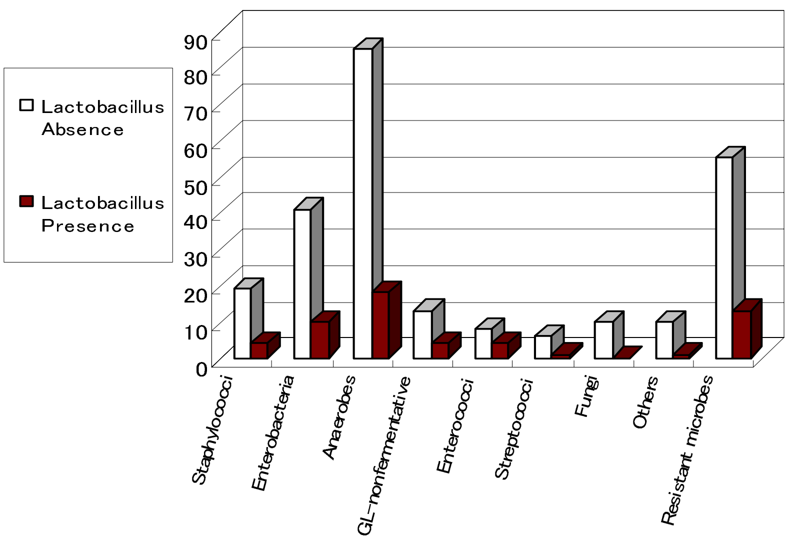


Fig. 2. Blood microbial counts in lactobacillus therapy. The counts in lactobacillus presence group are smaller than those counts in absence group.

3.5 Adjusted Residual Analysis

Various bacteria, such as *Staphylococci, Enterobacteria, Anaerobes, Gl-nonfermentative, Enterococci, Streptococci, Fungi*, and Resistant microbes, etc., were detected from blood samples. Fig.2 showed these bacteria counts in lactobacillus absence and the bacteria counts in lactobacillus presence. The counts in lactobacillus presence group are smaller than those counts in the absence group.

Table 4. Adjusted residue of blood microbisal counts in Lactobacillus therapy

	Lactobacillus Absence	**Lactobacillus Presence**
Staphylococci	**1.257**	**-1.257**
Enterobacteria	**1.538**	**-1.538**
Anaerobes	**2.728**	**-2.728**
GL-nonfermentative	**0.518**	**-0.518**
Enterococci	**-0.316**	**0.316**
Streptococci	**0.870**	**-0.870**
Fungi	**2.039**	**-2.039**
Others	**1.473**	**-1.473**
Resistant microbes	**1.879**	**-1.879**
No detection	**-4.799**	**4.799**
p = 0.00151 < 0.01		

Adjusted residue was applied in order to analyze the above blood microbial counts in lactobacillus therapy. Table 4 indicates adjusted residue for the blood microbial counts in lactobacillus absence and lactobacillus presence. Its p-value = 0.00151 < 0.01 statistically significant.

The top three rankings of these absolute values of the adjusted residue in Table 4 include Anaerobes, Fungi, and Resistant microbes.

Therefore, lactobacillus therapy might be especially effective for the reduction of blood stream infection by Anaerobes, Fungi, and Resistant microbes.

4 Discussion

We had an empirical rule that lactobacillus therapy (probiotic product) is effective in patient prognosis. Currently, lactobacillus preparation is used in the most departments of our center.

This analysis was conducted to extract background risk factors of blood stream infection in a year data of 2002, by chi-square test, decision tree, If-then rules and logistic regression.

Anti-biotics preparation has antibiotic properties, but it tends to get off balance of the normal bacteria flora and to cause diarrhea. On the other hand, lactobacillus therapy regulates the functions of the intestines and has no side-effects [10].

From the results of chi-square test (Table II), its p-value was 0.000000159 < 0.01. The odds ratio of lactobacillus absence to lactobacillus presence showed that bacteria detection risk of lactobacillus absence was about 2 (95%CI: 1.57-2.99). Therefore, lactobacillus therapy might be significantly effective to prevent the bacteria detection on blood sample.

On the other hand, the first node of the decision tree was lactobacillus therapy(Y/N). Therefore, lactobacillus therapy might be the most significant factor for prevention of blood stream infection. Various significant If-then rules were extracted from the decision tree.

From (If-then rule-1), lactobacillus therapy presence might prevent bacterial translocation when patient has not diarrhea and has central venous pressure (CVP) catheter and intravenous hyper-alimentation (IVH) catheter after the surgery. From (If-then rule-2), it was considered that lactobacillus therapy presence might protect patient's normal bacteria flora and might prevent bacterial translocation from the intestinal tract even though patient has diarrhea.

Furthermore, (If-then rule-4) and (If-then rule-5) showed that tracheotomy might caused bacteria detection from blood sample when patient has intravenous hyper-alimentation (IVH) catheter on lactobacillus therapy absence.

As the above mentioned, it was reported that bacterial translocation might be caused by antibiotics administration [8] and intravenous hyper-alimentation (IVH) [9]. Patient, who has tracheotomy, could not almost swallow down. Further-more, when the patient has also intravenous hyperalimentation (IVH) catheter, bact-eria in patient's oral cavity might increased abnormally and the patient's intestinal tract might lost its functions. Therefore, bacterial translocation from oral cavity or intestinal tract might be yielded.

The results of the logistic regression (Table 3) showed that bacteria detection has strong relationship with lactobacillus therapy, anti-biotics, diarrhea, catheter, and tracheotomy etc.

From the results of adjusted residue (Table 4), lactobacillus therapy might be especially effective for the reduction of blood stream infection by Anaerobes, Fungi, and Resistant microbes.

As shown in these results, decision trees, rule induction and logistic regression give us more useful results for decision making and risk management. Finding particular rules, patterns, and trends in the vast database is important, not only to the extraction and to the verification of the evidence-based medicine (EBM) for infectious diseases, but also to the prevention of incidents and the promotion of risk management. Especially, in the hospital infection control, the prevention of the hospital infection is mostly important. It is necessary and indispensable for the prevention of the hospital infection to analyze the background factors of blood stream infection.

Thus, the above empirical study shows that decision tree and rule induction method were found to be an effective method by our analytical study of background factors in blood stream infection.

In the age of the information technology, data on the clinical inspections have been stored and processed by the hospital information systems. Finding particular rules, patterns, and trends in the vast database is important, not only for extraction and verification of the evidence-based medicine (EBM) for infectious diseases, but also for prevention of incidents and the promotion of risk management.

In the hospital infection control, the prevention of the hospital infection is mostly important. It is necessary and indispensable for the prevention of the hospital infection to analyze the background factors of blood stream infection, etc. For hospital infection control, data mining was found to be an effective method by our analytical study of background factors in blood stream infection.

We consider it necessary to develop more concrete rules for the prevention of blood stream infection in datasets stored between 2003 and 2006 and to take the more specific measures for the prevention of hospital infections based on these analytical results.

5 Conclusion

In this paper, C4.5(decision tree and rule induction), chi-square test, odds ratio, logistic regression, and adjusted residual analysis were applied in order to extract certain patterns from our hospital clinical microbiology database, whose aim is to analyze the effects of lactobacillus therapy and the background risk factors on blood stream infection in patients by using data mining.

The significant “If-then rules” were extracted from the decision tree between bacteria detection on blood samples and patients’ treatments, such as lactobacillus therapy, antibiotics, and various catheters. Then, chi-square test, odds ratio and logistic regression were applied in order to analyze the effect of lactobacillus therapy to bacteria detection. From odds ratio of lactobacillus absence to lactobacillus presence, bacteria detection risk of lactobacillus absence was about 2 (95%CI: 1.57-2.99).

The significant “If-then rules”, chi-square test, odds ratio and logistic regression showed that lactobacillus therapy might be the significant factor for prevention of blood stream infection.

Experimental results show that lactobacillus therapy may be effective in reducing the risk of blood stream infection. Especially, rule induction method is useful for extracting background risk factors of blood stream infection from our clinical database.

Acknowledgments. The authors would like to express their thanks to Masateru Harao, Kouichi Hirata, Taeko Izumi, Noriyuki Okada, Chikara Nakahama, and finally to Satoshi Ichiyam.

References

1. Adriaans, P., et al.: Data Mining. Addison-Wesley, Reading (1996)
2. Quinlan, J.R.: C4.5 Programs for Machine Learning. Morgan Kaufmann, San Francisco (1993)
3. Matsuoka, K., Watanabe, K., et al.: Incidence and Clinical Significance of Anaerobic Bacteremia in a Community Hospital in Japan. In: The congress of the confederation of Anaerobe Societies, France (1998)

4. Matsuoka, K., Yokoyama, S., Tsumoto, S., et al.: Study on the relationship of Patients' Diseases and the Occurrence of Anaerobes by Using Data Mining Techniques. In: International Congress of the Confederation of Anaerobe Societies, ANAEROBE 2000, vol. IXa-P2(4), p. 186 (2000)
5. Yokoyama, S., et al.: Study on the association between the patients' clinical background and the Anaerobes by data mining in infectious diseases database. Biomedical Soft Computing and Human Sciences 7, 69–75 (2001)
6. Berg, R.D., Garligton, A.W.: Translocation of certain indigenous bacteria from the gastrointestinal tract to the mesenteric lymph nodes and other organs in a gnotobiotic mouse model. Infect Immun 23, 403–411 (1979)
7. Alexander, J.W., Boyce, S.T., et al.: The process of microbial translocation. Ann Surgery 212, 496–512 (1990)
8. Berg, R.D.: Promotion of the translocation of enteric bacteria from the gastrointestinal tract of mice by oral treatment with penicillin, lindamycin, or metronidazole. Ann Surgery 212, 496–512 (1990)
9. Alverdy, J.C., Aoys, E., Moss, G.S., et al.: Total parenteral nutrition promotes bacterial translocation from the gut. Surgery 104, 185–190 (1988)
10. Bruce, A.W., Reid, G.: Intravaginal instillation of Lactobacilli for prevention of recurrent urinary tract infections. Can. J. Microbiology 34, 339–343 (1990)

Discovery of Risky Cases in Chronic Diseases: An Approach Using Trajectory Grouping

Shoji Hirano and Shusaku Tsumoto

Department of Medical Informatics, Shimane University, School of Medicine
89-1 Enya-cho, Izumo, Shimane 693-8501, Japan
hirano@ieee.org, tsumoto@computer.org

Abstract. This paper presents an approach to finding risky cases in chronic diseases using a trajectory grouping technique. Grouping of trajectories on hospital laboratory examinations is still a challenging task as it requires comparison of data with mutidimensionalty and temporal irregulariry. Our method first maps a set of time series containing different types of laboratory tests into directed trajectories representing the time course of patient states. Then the trajectories for individual patients are compared in multiscale and grouped into similar cases. Experimental results on the chronic hepatitis data demonstrated that the method could find the groups of discending trajectories that well corresponded to the cases of higher fibrotic stages.

1 Introduction

Recent advances in medical information technology enable us to automatically collect huge amount of temporal data on clinical laboratory examinations. Cross-patient analysis of such temporal data has attracted much interests because it might reveal underlying relationships between the time course of examination results and diseases, which might be commonly observed on many patients. However, despite of its importance, large-scale analysis of time-series medical databases has rarely been performed due to the following problems: (1) sampling intervals and lengths of data can be both irregular, as they depend on the condition of each patient. (2) a time series can include various types of events such as acute changes and chronic changes. When comparing the time series, one is required to appropriately determine the correspondence of data points to be compared taking into account the above issues. Additionally, the dimensionality of data can be usually high due to the variety of medical examinations. These fearures prevent us from using conventional time series analysis methods.

This paper presents a novel cluster analysis method for multidimensional time-series data on medical laboratory tests. Our method represents time series of test results as trajectories in multidimensional space, and compares their structural similarity by using the multiscale comparison technique [1]. It enables us to find the part-to-part correspondences between two trajectories, taking into account the relationships between different tests. The resultant dissimilarity can be further used as input for clustering algorithms for finding the groups of similar cases. In the experiments we demonstrate the usefulness of our approach through the grouping tasks of artificially generated digit stroke trajectories and medical test trajectories on chronic hepatitis patients.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 289–302, 2008.

The main contributions of our method are twofold. First, it proposes a new approach to comparing trajectories of medical data by shape comparison techniques, not by standard time-series analysis techniques. Second, it introduces a two-step method for deriving dissimilarity between trajectories; multiscale matching is firstly applied in order to find structurally similar parts, and after that value-based dissimilarity is derived for each of the matched pairs and accumulated as the final dissimilarity between trajectories. This scheme makes the dissimilarity more informative as it takes not only value-based features but also structural features into account.

The remainder of this paper is organized as follows. In Section 2 we describe the methodoology, including preprocessing of the data. In Section 3 we show experimental results on a synthetic data (digit strokes) and chronic hepatitis data (albumin-platelet trajectories and cholinesterase-platelet trajectories). Finally, Section 4 is a conclusion of this paper.

2 Method

2.1 Preprocessing

Time-series examination data is often represented as a tuple of examination date and results. Interval of examinations is usually irregular, as it depends on the condition of a patient. However, in the process of multiscale matching, it is neccessary to represent time-series as a set of data points with a constant interval in order to represent the time span by the number of data points. Therefore, we employed linear interpolation and constructed new equi-interval data.

2.2 Multiscale Description of Trajectories by the Modified Bessel Function

Let us consider examination data for one person, consisting of I different time-series examinations. Let us denote the time series of i-th examination by $ex_i(t)$, where $i \in I$. Then the trajectory of examination results, $c(t)$ is denoted by

$$c(t) = \{ex_1(t), ex_2(t), \ldots, ex_I(t)\}$$

Next, let us denote an observation scale by σ and denote a Gaussian function with scale parameter σ^2 by $g(t, \sigma)$. Then the time-series of the i-th examination at scale σ, $EX_i(t, \sigma)$ is derived by convoluting $ex_i(t)$ with $g(t, \sigma)$ as follows.

$$EX_i(t, \sigma) = ex_i(t) \otimes g(t, \sigma) = \int_{-\infty}^{+\infty} \frac{ex_i(u)}{\sigma\sqrt{2\pi}} e^{\frac{-(t-u)^2}{2\sigma^2}} du$$

Applying the above convolution to all examinations, we obtain the trajectory of examination results at scale σ, $C(t, \sigma)$, as

$$C(t, \sigma) = \{EX_1(t, \sigma), EX_2(t, \sigma), \ldots, EX_I(t, \sigma)\}$$

By changing the scale factor σ, we can represent the trajectory of examination results at various observation scales. Figure 1 illustrates an example of multiscale representation of trajectories where $I = 2$. Increase of σ induces the decrease of convolution

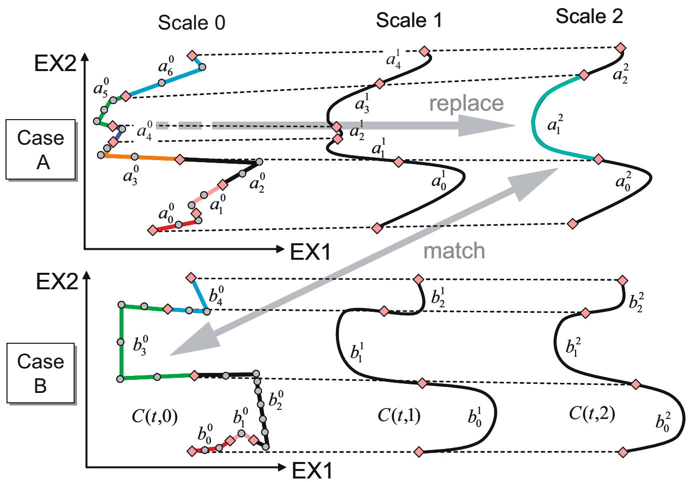


Fig. 1. Multiscale representation and matching scheme

weights for neighbors. Therefore, more flat trajectories with less inflection points will be observed at higher scales.

Curvature of the trajectory at time point t is defined by, for $I = 2$,

$$K(t, \sigma) = \frac{EX_1' EX_2'' + EX_1'' EX_2'}{(EX_1'^2 + EX_2'^2)^{3/2}}$$

where EX_i' and EX_i'' denotes the first- and second-order derivatives of $EX_i(t, \sigma)$ respectively. The m-th order derivative of $EX_i(t, \sigma)$, $EX_i^{(m)}(t, \sigma)$, is defined by

$$EX_i^{(m)}(t, \sigma) = \frac{\partial^m EX_i(t, \sigma)}{\partial t^m} = ex_i(t) \otimes g^{(m)}(t, \sigma)$$

It should be noted that many of the real-world time-series data, including medical data, can be discrete in time domain. Thus, a sampled Gaussian kernel is generally used for calculation of $EX_i(t, \sigma)$, changing an integral to summation. However, Lindeberg [2] pointed out that, a sampled Gaussian may lose some of the properties that a continuous Gaussian has, for example, non-creation of local extrema with the increase of scale. Additionally, in a sampled Gaussian kernel, the center value can be relatively large and imbalanced when the scale is very small. Ref. [2] suggests the use of kernel based on the modified Bessel function, as it is derived by incorporating the discrete property. Since this influences the description ability about detailed structure of trajectories, we employed the Lindeberg's kernel and derive $EX_i(t, \sigma)$ as follows.

$$EX_i(t, \sigma) = \sum_{n=-\infty}^{\infty} e^{-\sigma} I_n(\sigma) ex_i(t - n)$$

where $I_n(\sigma)$ denotes the modified Bessel function of order n. The first- and second-order derivatives of $EX_i(t,\sigma)$ are obtained as follows.

$$EX_i^{'}(t,\sigma) = \sum_{n=-\infty}^{\infty} -\frac{n}{\sigma} e^{-\sigma} I_n(\sigma) ex_i(t-n)$$

$$EX_i^{''}(t,\sigma) = \sum_{n=-\infty}^{\infty} \frac{1}{\sigma}(\frac{n^2}{\sigma} - 1) e^{-\sigma} I_n(\sigma) ex_i(t-n)$$

2.3 Segment Hierarchy Trace and Matching

For each trajectory represented by multiscale description, we find the places of inflection points according to the sign of curvature. Then we divide each trajectory into a set of convex/concave segments, where both ends of a segment correspond to adjacent inflection points. Let A be a trajectory at scale k composed of $M^{(k)}$ segments. Then A is represented by $\mathbf{A}^{(k)} = \{a_i^{(k)} \mid i = 1, 2, \cdots, M^{(k)}\}$, where $a_i^{(k)}$ denotes i-th segment at scale k. Similarly, another trajectory B at scale h is represented by $\mathbf{B}^{(h)} = \{b_j^{(h)} \mid j = 1, 2, \cdots, N^{(h)}\}$.

Next, we chase the cross-scale correspondence of inflection points from top scales to bottom scale. It defines the hierarchy of segments and enables us to guarantee the connectivity of segments represented at different scales. Details of the algorithm for checking segment hierarchy is available on ref. [1]. In order to apply the algorithm for closed curve to open trajectory, we modified it to allow replacement of odd number of segments at sequence ends, since cyclic property of a set of inflection points can be lost.

The main procedure of multiscale matching is to search the best set of segment pairs that satisfies both of the following conditions:

1. Complete Match: By concatenating all segments, the original trajectory must be completely formed without any gaps or overlaps.
2. Minimal Difference: The sum of segment dissimilarities over all segment pairs should be minimized.

The search is performed throughout all scales. For example, in Figure 1, three contiguous segments $a_3^{(0)} - a_5^{(0)}$ at the lowest scale of case A can be integrated into one segment $a_1^{(2)}$ at upper scale 2, and the replaced segment well matches to one segment $b_3^{(0)}$ of case B at the lowest scale. Thus the set of the three segments $a_3^{(0)} - a_5^{(0)}$ and one segment $b_3^{(0)}$ will be considered as a candidate for corresponding segments. On the other hand, segments such as $a_6^{(0)}$ and $b_4^{(0)}$ are similar even at the bottom scale without any replacement. Therefore they will be also a candidate for corresponding segments. In this way, if segments exhibit short-term similarity, they are matched at a lower scale, and if they present long-term similarity, they are matched at a higher scale.

2.4 Local Segment Difference

In order to evaluate the structural (dis-)similarity of segments, we first describe the structural feature of a segment by using shape parameters defined below.

1. Gradient at starting point: $g(a_m^{(k)})$
2. Rotation angle: $\theta(a_m^{(k)})$
3. Velocity: $v(a_m^{(k)})$

Figure 2 illustrates these parameters. Gradient represents the direction of the trajectory at the beginning of the segment. Rotation angle represents the amount of change of direction along the segment. Velocity represents the speed of change in the segment, which is calculated by dividing segment length by the number of points in the segment.

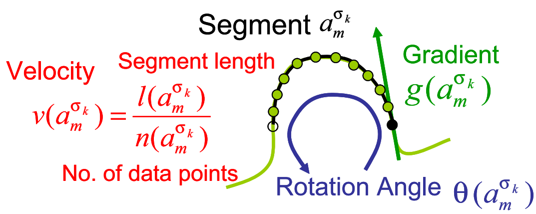


Fig. 2. Segment Parameters

Next, we define the local dissimilarity of two segments, $a_m^{(k)}$ and $b_n^{(h)}$, as follows.

$$d(a_m^{(k)}, b_n^{(h)}) = \sqrt{\left(g(a_m^{(k)}) - g(b_n^{(h)})\right)^2 + \left(\theta(a_m^{(k)}) - \theta(b_n^{(h)})\right)^2}$$
$$+ \left|v(a_m^{(k)}) - v(b_n^{(h)})\right| + \gamma \left\{cost(a_m^{(k)}) + cost(b_n^{(h)})\right\}$$

where $cost()$ denotes a cost function used for suppressing excessive replacement of segments, and γ is the weight of costs. We define the cost function using local segment dissimilarity as follows. For segment $a_m^{(k)}$ that replaces p segments $a_r^{(0)} - a_{r+p-1}^{(0)}$ at the bottom scale,

$$cost(a_m^{(k)}) = \sum_{q=r}^{r+p-1} d(a_q^{(0)}, a_{q+1}^{(0)})$$

2.5 Sequence Dissimilarity

After determining the best set of segment pairs, we newly calculate value-based dissimilarity for each pair of matched segments. The local segment dissimilarity defined in the previous section reflects the structural difference of segments, but does not reflect the difference of original sequence values; therefore, we calculate the value-based dissimilarity that can be further used as a metric for proximity in clustering.

Suppose we obtained L pairs of matched segments after multiscale matching of trajectories A and B. The value-based dissimilarity between A and B, $D_{val}(A, B)$, is defined as follows.

$$D_{val}(A, B) = \sum_{l=1}^{L} d_{val}(\alpha_l, \beta_l)$$

where α_l denotes a set of contiguous segments of A at the lowest scale that constitutes the l-th matched segment pair ($l \in L$), and β_l denotes that of B. For example, suppose that segments $a_3^{(0)} \sim a_5^{(0)}$ of A and segment $b_3^{(0)}$ of B in Figure 1 constitute the l-th matched pair. Then, $\alpha_l = a_3^{(0)} \sim a_5^{(0)}$ and $\beta_l = b_3^{(0)}$, respectively. $d_{val}(\alpha_l, \beta_l)$ is the difference between α_l and β_l in terms of data values at the peak and both ends of the segments. For the i-th examination ($i \in I$), $d_{val_i}(\alpha_l, \beta_l)$ is defined as

$$d_{val_i}(\alpha_l, \beta_l) = peak_i(\alpha_l) - peak_i(\beta_l) + \frac{1}{2}\{left_i(\alpha_l) - left_i(\beta_l)\} + \frac{1}{2}\{right_i(\alpha_l) - right_i(\beta_l)\}$$

where $peak_i(\alpha_l)$, $left_i(\alpha_l)$, and $right_i(\alpha_l)$ denote data values of the i-th examination at the peak, left end and right end of segment α_l, respectively. If α_l or β_l is composed of plural segments, the centroid of the peak points of those segments is used as the peak of α_l. Finally, d_{val_i} is integrated over all examinations as follows.

$$d_{val}(\alpha_l, \beta_l) = \frac{1}{I}\sqrt{\sum_i d_{val_i}(\alpha_l, \beta_l)}$$

3 Experimental Results

The usefulness of our method was evaluated through the clustering experiments using two different types of datasets. The first dataset contained artificially generated trajectories representing the nine digits. The second dataset contained real medical data; time series hospital laboratory data on chronic hepatitis patients.

3.1 Clustering of Synthetic Trajectories

We have created a total of 90 two-dimensional trajectories based on the shapes of nine digits according to the following procedure.

1. Create base trajectories for numbers one to nine by manually tracing the center line of the display font (we used Arial font for simplicity). Each trajectory is represented as a pair of time series: $c(t) = (x(t), y(t))$, where $x(t)$ and y(t) denote horizontal and vertical positions of a point at time t respectively. Time proceeds according to the standard stroke order in Japan, e.g., 1 (one) is written from top to bottom. We modified stroke order of two digits, 4 (four) and 5 (five), so that they can be written by one stroke. Figure 3 illustrates an example of trajectory for digit 3 (three).
2. Add noise to the position of each point. We used Gaussian noise that follows $N(0, 1)$. The noise is added independently to $x(t)$ and $y(t)$; therefore the local shapes of trajectories are disturbed quite largely. Figure 4 provides an example of the trajectory after adding noise. By repeating this process, we generated 10 noisy trajectories for each of the nine base trajectories. Consequently, we obtained a dataset containing a total of 90 noisy trajectories.

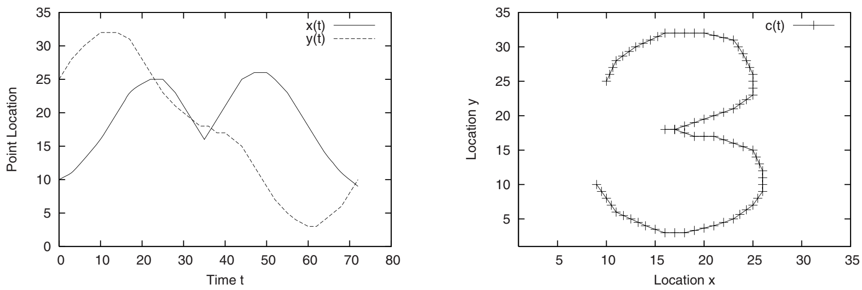


Fig. 3. An example of trajectory for digit 3. Left: time series $x(t)$ and $y(t)$. Right: trajectory $c(t) = (x(t), y(t))$. Time t starts from the point at (10,25). The ranges of x and y were set to be between 0 and 35 in this experiment.

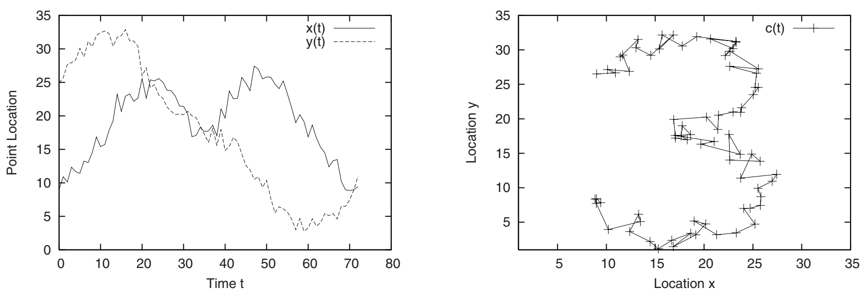


Fig. 4. An example of trajectory after adding noise to Figure 3. for digit 3. Left: time series $x(t)$ and $y(t)$. Right: trajectory $c(t) = (x(t), y(t))$.

By using this dataset, we performed cluster analysis and evaluated whether the proposed method could produce adequate dissimilarity that can be used to correctly group up trajectories representing the same digit. The procedure was as follows. For each pair of trajectories in the dataset, we calculated dissimilarity by applying the proposed multiscale comparison method. This generated a 90×90 dissimilarity matrix. Based on this matrix, we applied conventional hierarchical clustering method [3] and obtained a dendrogram. The parameters for multiscale comparison were determined through the pilot examination as follows: starting scale=0.5, scale interval=0.05, number of scales= 200, weight for segment replacement cost=0.5. We used group average as a linkage criterion for hierarchical clustering.

Figure 5 shows the dendrogram for the digit stroke trajectory dataset. As the dataset contained nine types of digits, we determined a cutting point on the dendrogram so that it creates nine clusters. The notions $c_1, c_2, \ldots, c_9$ represent cluster number respectively. The notions $6, 8, \ldots, 1$ at the bottom of the dendrogram represent correct groups, namely, digits that the trajectories represent.

Each of the clusters except for c_3, c_7 and c_8 contained 10 trajectories correctly representing the same digit. Cluster 3 contained a total of 20 trajectories, half representing

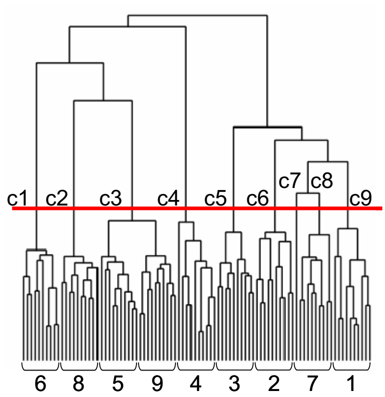


Fig. 5. Dendrogram for noisy digit dataset. The vertical line represents cutting point for 9 clusters solution. Notions c1 - c9 represents cluster number.

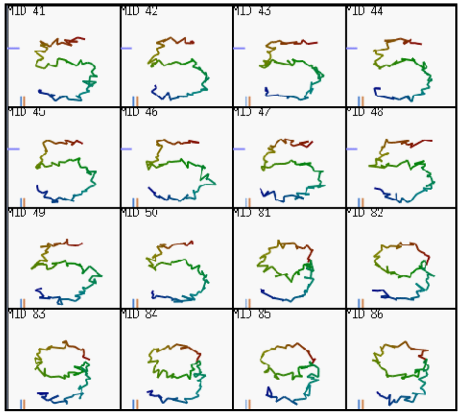


Fig. 6. Trajectories in cluster 3

digit 5 (five) and the remaining half representing 9 (nine). Figure 6 provides a part of trajectories in c_3. The stroke of 9 (nine) starts from the junction at center right. The stroke of 5 (five) should start from upper left, however, we modified its stroke order to start from upper right so that it could be written in one stroke. As a result, the base trajectories for these two digits look very similar. They were grouped into the same cluster under the nine cluster solution, however, the dendrogram shows they could be separated correctly at relatively high values of dissimilarity.

Trajectories representing the digit 7 (seven) were separated into two clusters, c_7 with 1 case, and c_8 with remaining 9 cases, though they could be merged at the next step. This seemed to happen because the method failed to find the good match for the one case in c_7 under the given parameters.

Although the results did not reach the perfect clustering, the above observations demonstrate that the dissimilarity measures induced by the proposed matching method could well reflect the fundamental difference of the trajectories.

3.2 Clustering of Trajectories on Laboratory Data

Next, we applied our method to the chronic hepatitis dataset which was a common dataset in ECML/PKDD discovery challenge 2002-2004 [4]. The dataset contained time series laboratory examinations data collected from 771 patients of chronic hepatitis B and C. In this work, we focused on analyzing the temporal relationships between platelet count (PLT), albumin (ALB) and cholinesterase (CHE), that were generally used to examine the status of liver function. Our goals were set to: (1) find groups of trajectories that exhibit interesting patterns, and (2) analyze the relationships between these patterns and the stage of liver fibrosis.

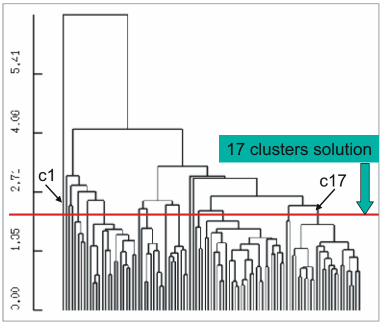


Fig. 7. Dendrogram for ALB-PLT trajectories in Type C without IFN dataset

Table 1. Cluster constitutions of ALB-PLT trajectories, stratified by fibrotic stages. Small clusters of $N < 2$ were omitted.

Cluster	# of Cases / Fibrotic stage				Total
	F0,F1	F2	F3	F4	
5	0	1	0	3	4
7	3	2	2	9	16
9	6	2	0	0	8
11	7	0	0	0	7
14	2	1	0	0	3
15	17	2	7	1	27
16	1	0	1	0	2
17	20	2	1	0	23

We selected a total of 488 cases which had valid examination results for all of PLT, ALB, CHE and liver biopsy. Constitution of the subjects classified by virus types and administration of interferon (IFN) was as follows. Type B: 193 cases, Type C with IFN: 296 cases, Type C without IFN: 99 cases. In the following sections, we mainly describe the results about Type C without IFN cases, which contained the natural courses of Type C viral hepatitis.

Experiments were conducted as follows. This procedure was applied separately for ALB-PLT trajectories and CHE-PLT trajectories.

1. Select a pair of cases (patients) and calculate the dissimilarity by using the proposed method. Apply this procedure for all pairs of cases, and construct a dissimilarity matrix.
2. Create a dendrogram by using conventional hierarchical clustering [3] and the dissimilarity matrix. Then perform cluster analysis.

Parameters for multiscale matching were empirically determined as follows: starting scale = 0.5, scale interval = 0,5, number of scales = 100, weight for segment replacement cost = 1.0. We used group average as a linkage criterion for hierarchical clustering. The experiments were performed on a small PC cluster consisted of 8 DELL PowerEdge 1750 (Intel Xeon 2.4GHz 2way) workstations. It took about three minutes to make the dissimilarity matrix for all cases.

Results on ALB-PLT trajectories. Figure 7 shows the dendrogram generated from the dataset on Type C without IFN cases. The dendrogram suggested splitting of the data into two or three clusters; however, in order to carefully examine the data structure, we avoided excessive merge of clusters and determined to split it into 17 clusters where dissimilarity increased relatively largely at early stage. For each of the 8 clusters that contained ≥ 2 cases, we classified cases according to the fibrotic stage. Table 1 shows the summary. The leftmost column shows cluster number. The next column shows the number of cases whose fibrotic stages were F0 or F1. The subsequent three columns show the number of F2, F3, and F4 cases respectively. The rightmost column shows the total number of cases in each cluster.

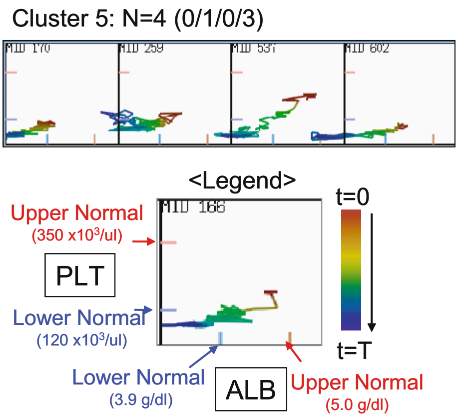


Fig. 8. Trajectories in Cluster 5

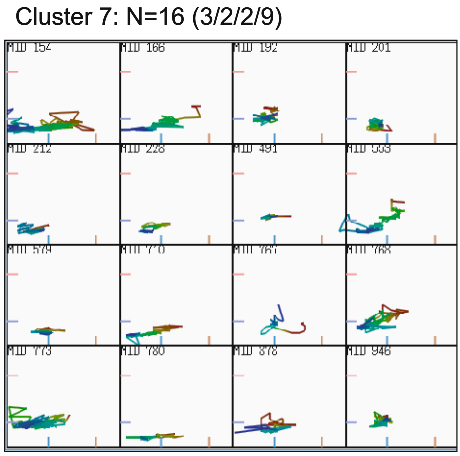


Fig. 9. Trajectories in Cluster 7

From Table 1, it could be recognized that the clusters can be globally classified into one of the two categories: one containing progressed cases of liver fibrosis (clusters 5 and 7) and another containing un-progressed cases (clusters 9, 11, 14, 15, 16 and 17). This can be confirmed from the dendrogram in Figure 7, where these two types of clusters appeared at the second devision from the root. This implied that the difference about ALB and PLT might be related to the fibrotic stages.

In order to recognize the detailed characteristics of 8 clusters, we observed the feature of grouped trajectories. Figures 8-11 show the examples of grouped ALB-PLT trajectories. Each quadrate region contains a trajectory of ALB-PLT values for a patient. If the number of cases in a cluster was larger than 16, the first 16 cases w.r.t. ID number were selected for visualization. The bottom part of Figure 8 provides the legend. The horizontal axis represents ALB value, and the vertical axis represents PLT value. Lower end of the normal range (ALB:$3.9g/dl$, PLT:$120 \times 10^3/ul$) and Upper end of the normal range (ALB:$5.0g/dl$, PLT:$350 \times 10^3/ul$) were marked with blue and red short lines on each axis respectively. Time phase on each trajectory was represented by color phase: red represents the start of examination, and it changes toward blue as time proceeds.

Figure 8 shows cases grouped into cluster 5 which contained remarkably many F4 cases (3/4). The skewed trajectory of ALT and PLT clearly demonstrated that both values decreased from the normal range to the lower range as time proceeded, due to the dysfunction of the liver. Cluster 7, shown in Figure 9, also contained similarly large number of progressed cases (F4:9/16, F3:2/16) and exhibited the similar characteristics, though it was relatively weaker than in cluster 5.

On the contrary, clusters that contained many un-progressed cases exhibited different characteristics. Figure 10 shows the trajectories grouped into cluster 17, where the number of F0/F1 cases was large (20/23). Most of the trajectories moved within the normal range, and no clear feature about time-direction dependency was observed. Figure 11 (top) shows the trajectories in cluster 11, where all of 7 cases were F0/F1. They moved within the normal range, but the PLT range was higher than in cluster 17.

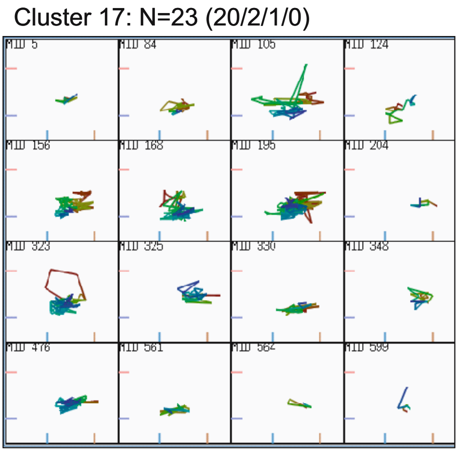


Fig. 10. Trajectories in Cluster 17

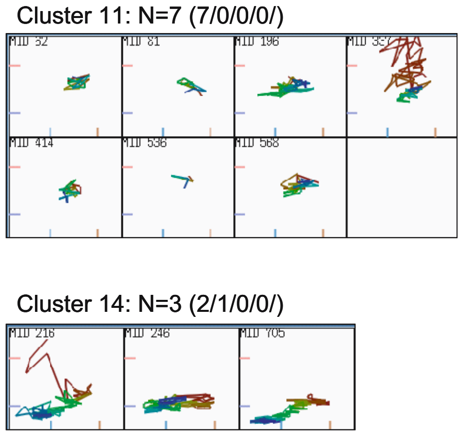


Fig. 11. Trajectories in Cluster 11 and 14

Figure 11 (bottom) shows the trajectories in cluster 14, where trajectories exhibited skewed shapes similarly to cluster 5. But this cluster consisted of F0/F1 and F2 cases, whereas cluster 5 contained mainly progressed cases. The reason why these cases were separated into different clusters should be investigated further, but it seemed that the difference of progress speed of liver fibrosis, represented as a velocity term, might be a candidate cause.

Results on CHE-PLT trajectories. Figure 12 shows the dendrogram generated from CHE-PLT trajectories of 99 Type C without IFN cases. Similarly to the case of ALB-PLT trajectories, we split the data into 15 clusters where dissimilarity increased largely at early stage. Table 2 provides cluster constitution stratified by fibrotic stage. In Table 2, we could observe a clear feature about the distribution of fibrotic stages over clusters. Clusters such as 3, 4, 6, 7 and 8 contained relatively large number of F3/F4 cases, whereas clusters such as 9, 11, 12, 13, 14, 15 contained no F3/F4 cases. These two types of clusters were divided at the second branch on the dendrogram; therefore it implied that, with respect to the similarity of trajectories, the data can be globally split into two categories, one contains the progressed cases and another contained un-progressed cases.

Now let us examine the features of trajectories grouped into each cluster. Figure 13 shows CHE-PLT trajectories grouped into cluster 3. The bottom part of the figure provides the legend. The horizontal axis corresponds to CHE, and the vertical axis corresponds to PLT. This cluster contained four cases: one F3 and three F4. The trajectories settled around the lower bounds of the normal range for PLT ($120 \times 10^3/ul$), and below the lower bounds of CHE (180 IU/l), with global direction toward lower values. This meant that, in these cases, CHE deviated from normal range earlier than PLT.

Figure 14 shows trajectories grouped into cluster 4, which contained nine F3/F4 cases and three other cases. Trajectories in this cluster exhibited interesting characteristics. First, they had very clear descending shapes; in contrast to trajectories in other

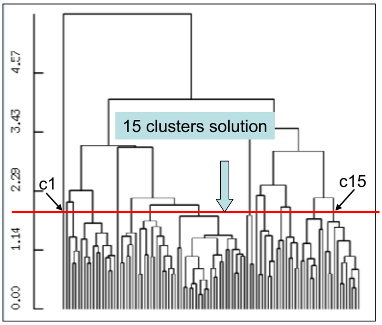


Fig. 12. Dendrogram for Type C without IFN dataset (CHE-PLT trajectories)

Table 2. Cluster constitutions of CHE-PLT trajectories, stratified by fibrotic stages. Small clusters of $N < 2$ were omitted.

Cluster	# of Cases / Fibrotic stage				Total
	F0,F1	F2	F3	F4	
3	0	0	1	3	4
4	2	1	2	7	12
6	3	0	1	2	6
7	5	2	3	3	13
8	9	8	4	2	23
9	1	2	0	0	3
11	4	2	0	0	6
12	2	0	1	0	3
13	5	0	0	0	5
14	8	0	0	0	8
15	12	0	0	0	12

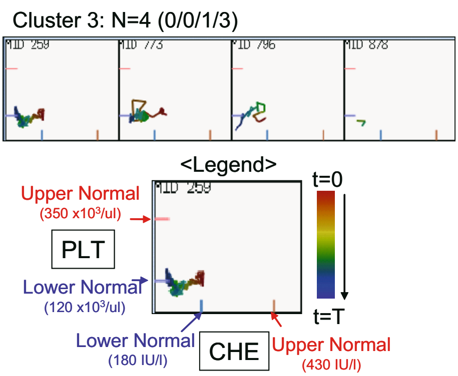


Fig. 13. Trajectories in Cluster 3

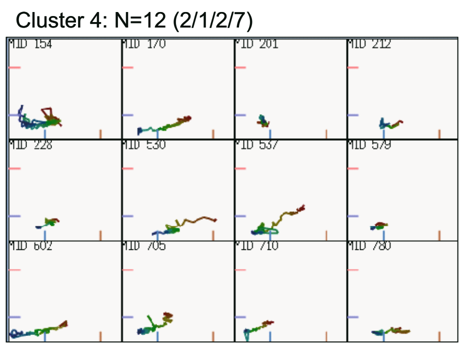


Fig. 14. Trajectories in Cluster 4

clusters in which trajectories changed directions frequently and largely, they moved toward the left corner with little directional changes. Second, most of the trajectories settled below the normal bound of PLT whereas their CHE values ranged within normal range at early phase. This meant that, in these cases, CHE deviated from normal range later than PLT.

Figure 15 shows trajectories grouped into cluster 6, which contained three F3/F4 cases and three other cases. Trajectories in this cluster exhibited descending shapes similarly to the cases in cluster 4. The average levels of PLT were higher than those in cluster 4, and did not largely deviated from the normal range. CHE remained within the normal range for most of the observations.

Figure 16 shows trajectories grouped into cluster 15, which contained twelve F0/F1 cases and no other cases. In contrast to the high stage cases mentioned above,

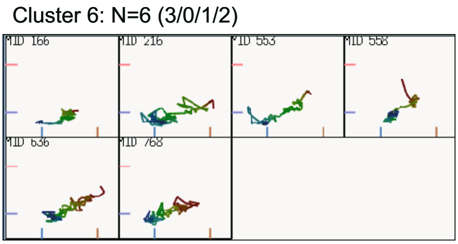


Fig. 15. Trajectories in Cluster 6

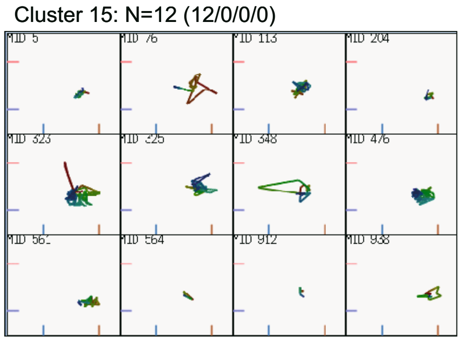


Fig. 16. Trajectories in Cluster 15

trajectories settled within the normal ranges for both CHE and PLT and did not exhibit any remarkable features about their directions.

These results suggested the followings about the CHE-PLT trajectories on type C without IFN cases used in this experiment: (1) They could be globally divided into two categories, one containing high-stage cases and another containing low-stage cases, (2) trajectories in some high-stage clusters exhibited very clear descending shapes. (3) in a group containing descending trajectories, PLT deviated from normal range faster than CHE, however, in another group containing descending trajectories, PLT deviated from normal range later than CHE.

4 Conclusions

In this paper we have presented a method for grouping trajectories on hospital examinations and its application to finding risky cases in chronic diseases. Our method employed a two-stage approach. Firstly, it compared two trajectories based on their structural similarity, and determines the best correspondence of partial trajectories. After that, it calculated the value-based dissimilarity for the all pairs of matched segments, and outputs the total sum as dissimilarity of the two trajectories.

Clustering experiments on the chronic hepatitis dataset yielded promising results. First, the clusters constructed with respect to the similarity of trajectories well matched with the distribution of fibrotic stages, especially with the distribution of high-stage cases and low-stage cases, for both ALB-PLT trajectories and CHE-PLT trajectories. It would make it possible to conduct further study about the non-invasive estimation of fibrotic stages based on the patterns of ALB-PLT/CHE-PLT courses. Next, we could find some groups of trajectories that exhibited descending patterns; for example, clusters 5 and 14 in ALB-PLT data and clusters 3, 4 and 6 in CHE-PLT data. They might be used for prediction of risky cases, by setting the clusters as new classes and inducing rules on the entire attributes that characterize these cases.

Our future work include the clinical validation of the results, treatment of the high dimensional data, and refinement of segment parameters.

References

1. Ueda, N., Suzuki, S.: A Matching Algorithm of Deformed Planar Curves Using Multiscale Convex/Concave Structures. IEICE Transactions on Information and Systems J73-D-II(7), 992–1000 (1990)
2. Lindeberg, T.: Scale-Space for Discrete Signals. IEEE Trans. PAMI 12(3), 234–254 (1990)
3. Everitt, B.S., Landau, S., Leese, M.: Cluster Analysis, 4th edn. Arnold Publishers (2001)
4. http://lisp.vse.cz/challenge/
5. Tsumoto, S., Hirano, S., Takabayashi, K.: Development of the Active Mining System in Medicine Based on Rough Sets. Journal of Japan Society of Artificial Intelligence 20(2), 203–210 (2005)

Part IV

Learning with Logics and Logics for Learning

The Fifth Workshop on Learning with Logics and Logics for Learning (LLLL2007)

Akihiro Yamamoto[1] and Kouichi Hirata[2]

[1] Graduate School of Informatics, Kyoto University, Japan
akihiro@i.kyoto-u.ac.jp
[2] Department of Artificial Intelligence, Kyushu Institute of Technology, Japan
hirata@ai.kyutech.ac.jp

1 The History of LLLL

The workshop on Learning with Logics and Logics for Learning (LLLL) was started in January 2002 in Sapporo, Japan, in order to encourage the interchange of computational logic and machine learning. After held twice as a domestic workshop, it was re-started in 2005 as an collocated international workshop with the Annual Conference of Japanese Society for Artificial Intelligence (JSAI).

In the past four workshops, we accepted 55 papers in total. We could classify them into two types. The first type is to introduce computational logic into machine learning, of which elements are Boolean algebra, clausal theories and structured data such as first-order terms. The second type is to provide and analyze semantics of logic and mathematics with machine learning, for example, clarifying the relation between computational algebra and machine learning.

2 The LLLL2007 Workshop

We planned to hold the fifth workshop on LLLL (LLLL2007) in the same manner as the last two workshops (LLLL2005/6), that is, as an international workshop collocated with the 21st Annual Conference of JSAI (JSAI2007).

We first organized the program committee consisting of 14 researchers concerning with the interchange of logic and learning[1]. Taking the stream of submitted papers into account, we announced a call for papers on the following topics (but not exclusive): learning and knowledge discovery using logics, algorithmic aspects of learning based on logics, logics for machine learning and knowledge discovery, logics using machine learning, machine learning as a foundation of mathematics/mathematical procedures, amalgamation of logic-based learning and statistical/information theoretical learning, and learning and knowledge discovery from relational data, structured/semi-structured data, and real-valued data. Every submitted paper to the workshop was reviewed by two PC members, and 8 papers were accepted.

[1] More information for LLLL has been available at the workshop homepage in the following URL: http://www.i.kyoto-u.ac.jp/~akihiro/LLLL.html

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 305–306, 2008.

The workshop was held on June 18 and 19 in 2007 at World Convention Center Summit, Miyazaki, Japan, including 2 invited talks by Dr. Ljupčo Todorovski (University of Ljubljana and Jožef Stefan Institute) and Prof. Taisuke Sato (Tokyo Institute of Technology)[2]. All of the talks attracted audiences very much and provoked a lot of discussions and questions.

3 Post-workshop Proceedings

The post-workshop proceedings of LLLL2007 were planned to be a volume of LNAI, Springer-Verlag (this volume), with papers selected from other collocated workshops and the main conference. After the workshop, we asked the authors if they would like to revise and submit their paper to the post-workshop proceedings. Then, by reviewing every paper (without withdrawn papers) by three PC members, two PC members previously assigned and another PC member, we have selected the following 3 papers to publish the post-workshop proceedings.

Arimura and Uno give a deductive foundation of data mining from sequential data. Then, they design an algorithm MaxFlex for enumerating all maximal flexible patterns from input sequences of characters in polynomial delay and polynomial space.

Takamatsu *et al.* discuss how the basis of a polynomial ideal works for identification in the limit from positive data, aiming at giving a new role in mathematical inference to machine learning. Then, they design an algorithm for finding the characteristic set of a given union of two polynomial ideals.

Yamamoto *et al.* analyze a CF-induction which is a mechanism of deriving hypotheses from given examples and a background theory under clausal logic. Then, they introduce a new sound and complete deductive operator, γ-operator, for clausal logic and apply it to the generalization process in the CF-induction.

Acknowledgments

We have finished the workshop LLLL2007 successfully. We would like represent our thankfulness to the invited speakers. We are grateful for their great support received from the program committee members: Yohji Akama, Marta Arias, Hiroki Arimura, Tamás Horváth, Katsumi Inoue, Eric Martin, Shin-ichi Minato, Luc de Raedt, Taisuke Sato, Takayoshi Shoudai, György Turán and Hiroaki Watanabe. In particular, we would express our great thankfulness to Prof. Hiroki Arimura, the chair of SIG-FPAI, for his encouragement and financial support with the Grant-in-Aid for Specially Promoted Research. We also thank Prof. Akihiro Inokuchi for his organization of the international workshops collocated with JSAI2007 and Prof. Ken Satoh for his arrangement to publish the LNAI volume of this post-workshop proceedings. Finally, we thank all speakers and all audiences attended the workshop.

[2] The working note of the workshop was published from JSAI (ISBN 4-915905-30-6 C3004(JSAI)).

Mining Maximal Flexible Patterns in a Sequence*

Hiroki Arimura[1] and Takeaki Uno[2]

[1] Graduate School of Information Science and Technology, Hokkaido University
Kita 14 Nishi 9, Sapporo 060-0814, Japan
arim@ist.hokudai.ac.jp
[2] National Institute of Informatics
2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
uno@nii.jp

Abstract. We consider the problem of enumerating all maximal flexible patterns in an input sequence database for the class of flexible patterns, where a *maximal pattern* (also called a closed pattern) is the most specific pattern among the equivalence class of patterns having the same list of occurrences in the input. Since our notion of maximal patterns is based on position occurrences, it is weaker than the traditional notion of maximal patterns based on document occurrences. Based on the framework of reverse search, we present an efficient depth-first search algorithm MaxFlex for enumerating all maximal flexible patterns in a given sequence database without duplicates in $O(||\mathcal{T}|| \times |\Sigma|)$ time per pattern and $O(||\mathcal{T}||)$ space, where $||\mathcal{T}||$ is the size of the input sequence database $\mathcal{T}$ and $|\Sigma|$ is the size of the alphabet on which the sequences are defined. This means that the enumeration problem for maximal flexible patterns is shown to be solvable in polynomial delay and polynomial space.

1 Introduction

The rapid growth of fast networks and large-scale storage technologies has led to the emergence of a new kind of massive data called *semi-structured data* emerged, which is a collection of weakly structured electronic data modeled by combinatorial structures, such as sequences, trees, and graphs. Hence, demand has arisen for efficient knowledge discovery algorithms for such semi-structured data.

In this paper, we consider the *maximal pattern discovery problem* for the class of flexible patterns in a sequence database [6,7,11,12], which is also called the *closed sequence mining problem* [11,12]. A *flexible pattern* is a sequence of constant strings separated by special gap symbols '$*$' such as AB*B*ABC, which means that the substring AB appears first in the input sequence, followed by B and

* This research was partly supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Specially Promoted Research, 17002008, 2007 on "semi-structured data mining", and 18017015, 2007 on "developing high-speed high-quality algorithms for analyzing huge genome database".

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 307–317, 2008.

then ABC. A pattern is *maximal w.r.t. position occurrences* in a sequence database if there is no properly more specific pattern that has the same set of occurrences in the input sequences. Thus, the *maximal flexible pattern discovery problem* is to enumerate all maximal flexible patterns in a given sequence database without duplicates.

For any minimum frequency threshold parameter $\sigma \geq 0$, the set of all frequent maximal patterns $\mathcal{M}_\sigma$ in input sequences contains complete information on the set of all frequent patterns $\mathcal{F}_\sigma$, and furthermore, $\mathcal{M}_\sigma$ is typically much smaller than $\mathcal{F}_\sigma$ if σ is small. Thus, the solution for maximal pattern discovery has the merit of increasing both efficiency and comprehensiveness of frequent pattern mining. On the other hand, the (frequent) maximal pattern discovery problem has high computational complexity compared with frequent pattern discovery. Thus, we need a lightweight and fast mining algorithm to solve the maximal pattern problem. In terms of algorithm theory, the efficiency of such enumeration algorithms is evaluated according to the worst-case computation time per solution. Particularly, if the algorithm works in polynomial space in terms of the input size, and the maximum time required to output the next pattern after outputting the previous one, called the delay, is of polynomial order of the input size, the algorithm said to be good.

As related works, Wang and Han [12] gave an efficient maximal pattern discovery algorithm BIDE for the class $\mathcal{SP}$ of sequential episodes [5] or subsequence patterns [12], where an episode is a pattern of the form $a_1 * \cdots * a_n$ $(a_i \in \Sigma, 1 \leq i \leq n)$. Arimura and Uno [4,6] gave a polynomial delay and polynomial time algorithm MaxMotif for maximal pattern discovery for the class $\mathcal{RP}$ of rigid patterns (or motifs with wildcards) [7], where a rigid pattern is of the form $w_1 \circ \cdots \circ w_n$ $(w_i \in \Sigma^*, 1 \leq i \leq n)$ for a single symbol wildcard $\circ$. However, no polynomial space and polynomial delay algorithm, or even no output polynomial time algorithm, has been known for the maximal pattern discovery problem for the class $\mathcal{FP}$ of flexible patterns. Here maximal patterns in $\mathcal{FP}$ are the patterns which are maximal among the patterns appearing at the same locations (or positions) of the given sequence database.

As a main result of this paper, for the class $\mathcal{FP}$ of flexible patterns, we present an efficient depth-first search algorithm MaxFlex that enumerates all maximal patterns $P \in \mathcal{FP}$ in a given sequence database $\mathcal{T}$ without duplicates in $O(|\Sigma| \times ||\mathcal{T}||)$ time per maximal pattern using $O(||\mathcal{T}||)$ space, where $||\mathcal{T}||$ is the size of $\mathcal{T}$, and $|\Sigma|$ is the size of the alphabet. A key to the algorithm is a depth-first search tree built on maximal patterns based on the reverse search framework [1]. Besides this, we discuss how to implement an efficient location list computation and maximality test. As a corollary, we show that the maximal pattern discovery problem for the class $\mathcal{FP}$ of flexible patterns is polynomial space and polynomial delay solvable. This result properly generalizes the output-polynomial complexity of the class $\mathcal{SP}$ of subsequence patterns [12].

The organization of this paper is as follows. In Section 2, we introduce the class $\mathcal{FP}$ of flexible patterns and define our data mining problem. We show an adjacency relation between maximal flexible patterns that implicitly induces a tree-shaped search route in Section 3, and describe algorithm MaxFlex that performs a depth-first search on the search route in Section 4. We conclude this paper in Section 5.

2 Preliminaries

Let Σ be an alphabet of symbols. We denote the set of all possibly empty strings and the set of all non-empty finite strings over Σ by Σ^* and $\Sigma^+ = \Sigma^* - \{\varepsilon\}$, respectively. Let $s = a_1 \cdots a_n \in \Sigma^*$ be a string over Σ of length n. We denote the length of s by $|s|$, i.e., $|s| = n$. The string with no symbol is called an *empty string*, and it is denoted by ε. For any $1 \le i \le j \le n$, we denote the *i-th symbol* of s by $s[i] = a_i$. For i and j such that $1 \le i \le j \le |s|$, the string $a_i \cdots a_j$ is called a *substring* of s, and denoted by $s[i..j]$. We say that a substring v *occurs in s at position i* iff $v = s[i..j]$. For two strings $s = a_1 \cdots a_n$ and $t = b_1 \cdots b_n$, the *concatenation* of t to s is the string $s = a_1 \cdots a_n b_1 \cdots b_n$, and denoted by $s \bullet t$ or simply st. A string u is called a *prefix* of s if $s = uv$ holds for some string v, and is called a *suffix* of s if $s = vu$ holds for some v. For a set $\mathcal{S}$ of strings, we denote the *cardinality* of S by $|S|$. The sum of the string lengths in $\mathcal{S}$ is called the *total size* of $\mathcal{S}$ and denoted by $||\mathcal{S}||$.

2.1 Patterns and Their Location Lists

We introduce the class $\mathcal{FP}$ of flexible patterns [6], also known as erasing regular patterns. Let $\Sigma = \{a, b, c, \ldots\}$ be a finite alphabet of *constant symbols*. A *gap* or a *variable-length don't care*, (*VLDC*) is a special symbol $* \notin \Sigma$, which represents an arbitrarily long possibly-empty finite string in Σ^*. A *constant string* is a string composed only of constant symbols. A *flexible pattern* (*pattern*, for short) over Σ is a sequence $P = w_0 * w_1 * \cdots * w_d$ of non-empty constant strings separated by gap symbols, where each constant string $w_i \in \Sigma^+$, $(0 \le i \le d, d \ge 0)$ is called a *segment* of P. w_0 is called the *first segment* of P and denoted by $seg_0(P)$. $w_1 * \cdots * w_d$ is called the *segment suffix* of P, and denoted by $sfx_0(P)$. A flexible pattern is called an *erasing regular pattern* in the field of machine learning (Shinohara [10]) and a *VLDC pattern* in the field of pattern matching. The size of P is $|P| = \sum_i^d |w_i|$. For every $d \ge 0$, a d-pattern is a pattern with $d+1$ segments. We denote the class of flexible patterns over Σ by $\mathcal{FP}$. Clearly, $\Sigma^+ \subseteq \mathcal{FP}$.

Example 1. Let $\Sigma = \{A, B, C\}$. Then, $P_0 = AB$, $P_1 = A{*}B$, $P_2 = AB{*}A{*}ABC$ are flexible patterns.

Let $P = w_0 * \cdots * w_d \in \mathcal{FP}, w_i \in \Sigma^+$ be a d-pattern $(d \ge 0)$. A *substitution* for P is a d-tuple $\theta = (u_1, \ldots, u_d) \in \mathcal{FP}^d$ of non-empty constant strings. We define the application of θ to P, denoted by $P\theta$, as the string $P\theta = w_0 u_1 w_1 u_2 \cdots u_d w_d \in \mathcal{FP}$, where the i-th occurrence of the variable $*$ is replaced with the i-th string u_i for every $i = 1, \ldots, d$. The string $P\theta$ is said to be an *instance* of P by substitution θ.

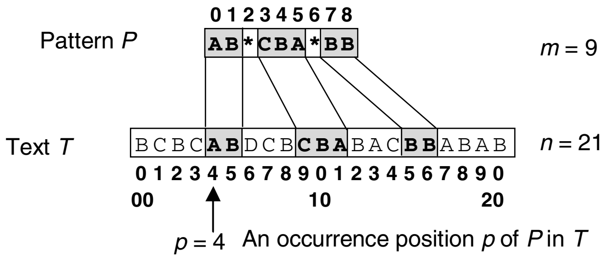


Fig. 1. A flexible pattern P and its position in a constant string T

We define a binary relation $\sqsubseteq$ over $\mathcal{P}$, called the *specifity relation*, as follows [2,6,10]. A *position* in a flexible pattern P is a number $p \in \{1, \ldots, |P|\}$. Here each position i means the position of ith constant symbol.

For flexible patterns $P, Q \in \mathcal{FP}$, we say P *occurs in* Q if there exists a substitution θ such that $\alpha(P\theta)\beta = Q$ holds for some $\alpha, \beta \in \mathcal{FP}$. We say P occurs in Q at position $|\alpha| + 1$. Here $|\alpha|$ also means the number of constant symbols in α. The *location list* for P in Q, denoted by $LO(P, Q)$ is the set of all possible positions of pattern P in Q.

If P occurs in Q we say that either Q is *more specific than* P or P is *more general than* Q, and write $P \sqsubseteq Q$. If both $P \sqsubseteq Q$ and $Q \not\sqsubseteq P$ hold, we say Q is *properly more specific than* P and write $P \sqsubset Q$. Note that $P = Q$ holds if and only if $P \sqsubseteq Q$ and $Q \sqsubseteq P$ hold.

Lemma 1. *$(\mathcal{FP}, \sqsubseteq)$ is a partial ordering with the smallest element ε.*

Example 2. For patterns $P_3 = A*BA*B$ and $P_4 = BAB*BA*AB*B$, we can see that $P_3 \sqsubseteq P_4$ by the embedding $B\underline{A}B*\underline{BA}*A\underline{B}*B$, where the image of P_4 is indicated by underlines. Formally, $P_3 \sqsubseteq P_4$ holds because P_3 has an instance $P_3\theta = AB*BA*AB$ as a substring of P_4 for the substitution $\theta = (B*, *A)$. Clearly, $P_3 \sqsubset P_4$ since $P_3 \sqsubseteq P_4$ but $P_4 \not\sqsubseteq P_3$.

A *sequence database* $\mathcal{T} = \{T_1, \ldots, T_m\}$ is a set of constant strings $T_i \in \Sigma^*$. We denote the number of strings by m. The sum of the sizes of $T_1, \ldots, T_m$ is called the *size* of $\mathcal{T}$ and denoted by $||\mathcal{T}||$, i.e., $||\mathcal{T}|| = \sum_{i=1}^{m} |T_i|$. In what follows, we fix the sequence database unless stated otherwise. For a flexible pattern P and a set of flexible patterns $\mathcal{T}$, which can be a sequence database, the *location set* for P in $\mathcal{T}$, denoted by $LO(P, \mathcal{T})$ is the set of location lists $LO(P, T_i)$ in all $T_i \in \mathcal{T}$.

For flexible patterns P and Q, the largest position in $LO(P, Q)$ is called the *rightmost position* of P in Q, and denoted by $p_{max}(P, Q)$. If $LO(P, Q)$ is empty, we define $p_{max}(P, Q)$ as $-\infty$. If P is an empty sequence, we define $p_{max}(P, Q)$ as $|Q| + 1$.

Lemma 2. *For any $P, T \in \mathcal{FP}$, $LO(P, T) = \{p \mid p \in LO(seg_0(P), T), p + |seg_0(P)| \leq p_{max}(sfx_0(P), T)\}$.*

The *frequency* $frq(P, \mathcal{T})$ of a flexible pattern P in a sequence database $\mathcal{T}$ is $|LO(P, \mathcal{T})| = |\{LO(P, T_i) \mid T_i \in \mathcal{T}, LO(P, T_i) \neq \emptyset\}|$. A *minimum support threshold* is a non-negative integer $0 < \sigma \leq n$. A flexible pattern P is σ*-frequent* in $\mathcal{T}$ if its frequency in $\mathcal{T}$ is no less than σ, i.e., $frq(P, \mathcal{T}) \geq \sigma$.

2.2 Maximal Pattern Enumeration Problem

Definition. 1 A flexible pattern P is *maximal* in $\mathcal{T}$ if there is no proper specialization Q of P that has the same location list, that is, there is no $Q \in \mathcal{FP}$ such that $P \sqsubset Q$ and $LO(P, \mathcal{T}) = LO(Q, \mathcal{T})$.

We see that a pattern $P \in \mathcal{FP}$ is maximal iff P is a maximal element w.r.t. $\sqsubseteq$ in the equivalence class $[P]_{\mathcal{T}} = \{Q \in \mathcal{FP} | P \equiv_{\mathcal{T}} Q\}$ under the equivalence relation $\equiv_{\mathcal{T}}$ over $\mathcal{FP}$ defined by $P \equiv_{\mathcal{T}} Q \Leftrightarrow LO(P, \mathcal{T}) = LO(Q, \mathcal{T})$.

Lemma 3. *The maximal patterns in each equivalence class $[P]_{\mathcal{T}}$ are not unique in general.*

We denote the family of the maximal patterns in $\mathcal{T}$ by $\mathcal{M}$, and the family of σ-frequent flexible patterns by $\mathcal{F}_\sigma$. Let $\mathcal{M}_\sigma = \mathcal{F}_\sigma \cap \mathcal{M}$ be the family of σ-frequent maximal patterns. It is easy to see that the number of frequent flexible patterns in $\mathcal{T}$ can be exponential in $||\mathcal{T}||$ and the same is true for $\mathcal{M}_\sigma$.

Lemma 4. *There is an infinite sequence $\mathcal{T}_1, \mathcal{T}_2, \ldots$ of sequence databases such that the number of maximal flexible patterns in $\mathcal{T}_i$ is exponential in $||\mathcal{T}_i||$.*

Now, we state our data mining problem as follows.

Position Maximal Flexible Pattern Enumeration Problem:

Input: sequence database $\mathcal{T}$ over an alphabet Σ, a minimum support threshold σ

Output: all maximal σ-frequent flexible patterns in $\mathcal{M}_\sigma$ in $\mathcal{T}$ without duplicates

Our goal is to develop an efficient enumeration algorithm for this problem.

3 Tree-Shaped Search Route for Maximal Flexible Patterns

In this section, we introduce a tree-shaped search route $\mathcal{R}$ spanning all elements in $\mathcal{M}$. In section 4, we give a memory efficient algorithm for enumerating all maximal flexible patterns based on the depth-first search over $\mathcal{R}$. Our strategy is as follows: First we define a binary relation between maximal patterns, called the *parent function*, which indicates a directed edge from a child to its parent. The parent function induces an adjacency relation on $\mathcal{M}$, whose form is a spanning tree.

We start with several technical lemmas.

Definition. 2 A flexible pattern Q is said to be a *prefix specialization* of another flexible pattern P if $1 \in LO(P, Q)$. If Q is a specialization of P but not a prefix specialization, Q is said to be a *non-prefix specialization* of P.

The following two lemmas are essential for flexible patterns. The first one says that a limited kind of monotonicity holds for flexible patterns. The proof is obvious from the transitivity of the specialization relation.

Lemma 5. *For any $P, Q, T \in \mathcal{FP}$ if P is a prefix specialization of Q, then $LO(P, T) \supseteq LO(Q, T)$.*

The second lemma gives us a technical property that a non-prefix specialization of P always has a location list different from that of P. Thus, attaching a new symbol to the left of a flexible pattern never preserves its location list.

Lemma 6. *Let $T, P, Q \in \mathcal{FP}$ such that $P \sqsubset Q \sqsubseteq T$. Then, if Q is a non-prefix specialization of P then $LO(P, T) \neq LO(Q, T)$.*

Proof. Since Q is a specialization of P but not a prefix specialization, $LO(P, Q)$ includes a position $p > 1$. This means that P occurs in $T[p_{max}(Q, T)+p-1..|T|]$, thus $p_{max}(P, T) \neq p_{max}(Q, T)$. This implies $LO(P, T) \neq LO(Q, T)$. □

Corollary 1. *A flexible pattern $Q \in \mathcal{FP}$ is maximal if and only if none of its prefix specializations has a location list equal to Q.*

Definition. 3 The *parent* $\mathcal{P}(P)$ of a flexible pattern P is the flexible pattern obtained from P by removing its first symbol and $*$ if the following operator is $*$.

Lemma 7. *For any non-empty flexible pattern $Q \in \mathcal{FP}$, its parent is always defined and unique.*

Lemma 8. *For any non-empty flexible pattern P and constant symbol a, $LO(a \bullet P, T) = \{p \in LO(a, T) \mid p+1 \in LO(P, T)\}$, and $LO(a * P, T) = \{p \in LO(a, T) \mid p < p_{max}(P, T)\}$.*

Corollary 2. *For any flexible pattern P, $frq(P, \mathcal{T}) \leq frq(\mathcal{P}(P), \mathcal{T})$.*

The proof of the above lemma is omitted, but it is not difficult. A *root pattern* in T is a maximal pattern P such that $LO(P, \mathcal{T}) = LO(\varepsilon, \mathcal{T})$. The root pattern is either ε or a symbol a if all symbols in any sequence $T \in \mathcal{T}$ are a. Now, we are ready for the main result of this section.

Theorem 1 (reverse search property of $\mathcal{M}$). *Let $Q \in \mathcal{M}$ be a maximal flexible pattern in $\mathcal{T}$ that is not a root pattern. Then, $\mathcal{P}(Q)$ is also a maximal flexible pattern in $\mathcal{T}$, that is, if $Q \in \mathcal{M}$ then $\mathcal{P}(Q) \in \mathcal{M}$ holds.*

Proof. Let Q be a maximal flexible pattern that is not a root pattern. To prove the theorem by contradiction, we suppose that there is a proper specialization P' of $\mathcal{P}(Q)$ such that $LO(\mathcal{P}(Q), \mathcal{T}) = LO(P', \mathcal{T})$. If P' is not a prefix specialization of $\mathcal{P}(Q)$, then from Lemma 6, we see that $LO(P, \mathcal{T}) \neq LO(P', \mathcal{T})$. Thus, we consider the case that P' is a prefix specialization of $\mathcal{P}(Q)$.

From the definition of the parent, $Q = a \circledcirc \mathcal{P}(Q)$ for some $a \in \Sigma$ and $\circledcirc \in \{\bullet, *\}$. Let $Q' = a \circledcirc P' \in \mathcal{FP}$. Since P' is a proper prefix specialization of

$\mathcal{P}(Q)$, Q' is also a proper prefix specialization of Q. Let T be a sequence in $\mathcal{T}$ and p be a position $p \in LO(Q,T)$. If $\circledcirc = \bullet$, then $p+1 \in LO(\mathcal{P}(Q),T)$ and thus also $p+1 \in LO(P',T)$. Thus we have $p \in LO(Q',T)$. If $\circledcirc = *$, then $p < p_{max}(\mathcal{P}(Q),T)$ thus also $p < p_{max}(P',T)$. Thus, we have $p \in LO(Q',T)$. In both cases, we have $LO(Q,T) \subseteq LO(Q',T)$, thus $LO(Q,T) = LO(Q',T)$. This immediately implies that $LO(Q,\mathcal{T}) = LO(Q',\mathcal{T})$, contradiction. □

Definition. 4 A *search route* for $\mathcal{M}$ w.r.t. $\mathcal{P}$ is a directed graph $\mathcal{R} = (\mathcal{M}, \mathcal{P}, \perp)$ with root, where $\mathcal{M}$ is the set of nodes, i.e., the set of all maximal flexible patterns in $\mathcal{T}$, $\mathcal{P}$ is the set of *reverse edge* such that $(P,Q) \in \mathcal{P}$ iff $P = \mathcal{P}(Q)$ holds, and $\perp \in \mathcal{M}$ is the root pattern in $\mathcal{T}$.

Since each non-root node has its parent in $\mathcal{M}$ by Theorem 1 and $|\mathcal{P}(P)| < |P|$, the search route $\mathcal{R}$ is actually a directed tree with reverse edges. Therefore, we have the following corollary.

Corollary 3. *For any sequence database $\mathcal{T}$, $\mathcal{R} = (\mathcal{M}, \mathcal{E}, \perp)$ forms a rooted spanning tree with the root $\perp$.*

We have the following lemma on the shape of $\mathcal{T}$.

Lemma 9. *Let $P \in \mathcal{M}$ be any maximal pattern in $\mathcal{T}$ and $m = |P|$. Then,*

(i) the depth of P in $\mathcal{T}$ (the length of the unique path from the root to P) is at most m.
(ii) the branching of P in $\mathcal{T}$ (the number of the children for P) is at most $2|\Sigma|$.

4 A Polynomial Time and Polynomial Delay Algorithm

In Figure 2 shows a polynomial space and polynomial delay enumeration algorithm MaxFlex for maximal flexible patterns. This algorithm starts from the bottom pattern $\perp$ and searches from smaller to larger all maximal patterns in a depth-first search manner over the search route $\mathcal{R}$.

However, since the search route $\mathcal{R}$ is defined by the reverse edges from children to their parents, it is not an easy task to traverse the edges. We firstly explain how to compute all children of given parent pattern $P \in \mathcal{M}$. The following lemma ensures that any child can be obtained by attaching a new symbol with an operator to the left of P.

Lemma 10. *For any maximal flexible patterns $P, Q \in \mathcal{M}$, $P = \mathcal{P}(Q)$ if and only if Q is maximal, and $Q = a \circledcirc P$ holds for some constant symbol $a \in \Sigma$ and $\circledcirc \in \{\bullet, *\}$.*

Since $\mathcal{P}(Q)$ is defined for any non-empty pattern Q, we know from Lemma 10 that any flexible pattern can be obtained from $\perp$ by a finite number of applications of the operator $\bullet$ or $*$. Therefore, Theorem 1 ensures that we can correctly prune all descendants of the current pattern P if it is no longer maximal in

```
Algorithm MaxFlex(Σ, T, σ):
input: sequence database T on an alphabet Σ s.t. any T ∈ T is in Σ*, minimum support threshold σ
output: All maximal patterns in M_σ
compute the root pattern ⊥ //the maximal pattern in T equivalent to ε
ExpandMaxFlex(⊥, LO(⊥, T), T, σ);

Procedure ExpandMaxFlex(P, LO(P, T), T, σ):
input: maximal pattern P, location list LO(P, T), sequence database T, minimum support threshold σ
output: all maximal patterns in M_σ that are descendants of P
if frq(P, T) < σ then return
if P is not maximal in T then return
output P
foreach pair of a ∈ Σ and ⊚ ∈ {•, *} do begin
    ExpandMaxFlex(a ⊚ P, LO(a ⊚ P, T), T)
end
```

Fig. 2. An algorithm MaxFlex for enumerating all maximal flexible patterns in a sequence database

depth-first search of $\mathcal{M}$. Moreover, if $frq(P, \mathcal{T}) < \sigma$, no descendant of P is frequent. Thus, we can also prune the descendants.

Secondly, we discuss the computation time of the algorithm. The bottleneck of the computation is the check of the maximality of the current pattern P, thus we need an efficient way to test it.

The notion of the refinement operator was introduced by Shapiro [8]. The refinement operator for flexible patterns in $\mathcal{FP}$, under the name of erasing regular patterns, was introduced by Shinohara [10]. The following version is due to [2,3].

Definition. 5 Let $P \in \mathcal{FP}, P = w_0 * \cdots * w_d$ be any flexible pattern. Then, we define the set $\rho(P) \subseteq \mathcal{FP}$ as the set of all patterns $Q \in \mathcal{FP}$, called *basic refinements* of P if Q is obtained from P by one of the following operations:

(1) replace w_i by $w_i * a$ for some $a \in \Sigma$ and $0 \leq i \leq d$.
(2) replace $w_i * w_{i+1}$ by $w_i w_{i+1}$ for some $0 \leq i \leq d - 1$.

Lemma 11. *A flexible pattern P is maximal in $\mathcal{T}$ if and only if there is no basic refinement $Q \in \rho(P)$ such that $LO(P, \mathcal{T}) = LO(Q, \mathcal{T})$.*

Suppose that P and T are flexible patterns and the segments of a flexible pattern P are $w_0, \ldots, w_d$. Let $left(P, T, i), 0 \leq i \leq d$ be the minimum position p such that $T[1..p]$ is a specification of $w_0 * \cdots * w_i$. If T is not a specification of P, $left(P, T, i)$ is defined by $+\infty$. If P is an empty sequence, we define $left(P, T, i) = 0$.

For any segment w and position p in T, let $succ(w, T, p)$ be the smallest position q such that $q \geq p$ and $q \in LO(w, T)$. For any string w, computing $succ(w, T, p)$ for all pairs of $T \in \mathcal{T}$ and $p \in LO(w, T)$ can be done in $O(||\mathcal{T}||)$ time.

Lemma 12. *There is a basic refinement obtained by replacing w_i by $w_i * a$ having the same location list as P if and only if there is a common symbol in $T[left(P, T, i) + 1..p_{max}(w_{i+1} * \cdots * w_d, T) - 1]$ for all $T \in \mathcal{T}$.*

Proof. The if part of the statement is obvious. We check the "only if" part, by showing that there is a common symbol a. Suppose that a basic refinement obtained by replacing w_i by $w_i * a$ has the same location list as P. Then for any $T \in \mathcal{T}$, there is a position q such that $T[q] = a$, $T[1..q-1]$ is a specification of $w_0 * \cdots * w_i$, and $T[q+1..|T|]$ is a specification of $w_{i+1} * \cdots * w_d$. Then, from the definition of p_{max} and $left$, $left(P, T, i) < q < p_{max}(w_{i+1} * \cdots * w_d, T)$. This states the statement of the lemma. □

Similarly, we obtain the following lemma.

Lemma 13. *There is a basic refinement obtained by replacing* $w_i * w_{i+1}$ *by* $w_i w_{i+1}$ *having the same location list as* P *if and only if* $succ(w_i w_{i+1}, T, left(P, T, i-1)) + |w_i w_{i+1}| \leq p_{max}(w_{i+2} * \cdots * w_d, T)$ *for any* $T \in \mathcal{T}$.

Lemma 14. *Using* $left$*,* $succ$*, and* p_{max}*, we can determine whether there is a basic refinement having the same location list as* P *in* $O(||\mathcal{T}||)$ *time.*

Proof. The statement is clear for basic refinements obtained by replacing $w_i * w_{i+1}$ by $w_i w_{i+1}$. Thus, we consider basic refinements obtained by replacing w_i by $w_i * a$.

Let $Count(a, i), a \in \Sigma, 0 \leq i \leq d$ be the number of sequences $T \in \mathcal{T}$ such that $T[left(P, T, i) + 1..p_{max}(w_{i+1} * \cdots * w_d, T) - 1]$ includes a. What we have to do is to check whether $Count(a, i) = m$ holds for some $a \in \Sigma$ or not. For any i, $Count(a, i)$ for all $a \in \Sigma$ can be computed in $O(||\mathcal{T}||)$ time.

For $i > 0$, let $diff(P, T, i)$ be the set of symbols with signs, such as $+a, -a, +b, -b$, such that $+a$ is included in $diff(P, T, i)$ iff $T[left(P, T, i-1) + 1..p_{max}(w_i * \cdots * w_d, T) - 1]$ does not include a but $T[left(P, T, i) + 1..p_{max}(w_{i+1} * \cdots * w_d, T) - 1]$ includes a, and $-a$ is included in $diff(P, T, i)$ iff $T[left(P, T, i-1) + 1..p_{max}(w_i * \cdots * w_d, T) - 1]$ includes a but $T[left(P, T, i) + 1..p_{max}(w_{i+1} * \cdots * w_d, T) - 1]$ does not include a. Using $diff(P, T, i)$, $Count(a, i+1)$ for all $a \in \Sigma$ can be obtained from $Count(a, i-1)$ of all $a \in \Sigma$ in $O(\sum_{T \in \mathcal{T}} |diff(P, T, i)|)$ time. Since each $diff(P, i, T)$ can be computed in $O((left(P, T, i) - left(P, T, i-1)) + (p_{max}(w_{i+1} * \cdots * w_d, T) - p_{max}(w_i * \cdots * w_d, T)))$ time, computing $diff(P, i, T)$ for all pairs of $i, 1 \leq i \leq d$ and $T \in \mathcal{T}$ takes $O(||\mathcal{T}||)$ time. This concludes the lemma. □

Lemma 15. *Using* $succ$*,* $left(P, T, i)$ *for all* $T \in \mathcal{T}$ *can be computed in* $O(||\mathcal{T}||)$ *time.*

By combining the above lemmas, we get the main result of this paper, which says that MaxFlex is a memory and time efficient algorithm.

Theorem 2. *Let* Σ *be an alphabet,* $\mathcal{T}$ *a sequence database, and* σ *a minimum support threshold. Then, the algorithm* MaxFlex *in Fig. 2 enumerates all maximal flexible patterns* $P \in \mathcal{M}_\sigma$ *of* $\mathcal{T}$ *without duplicates in* $O(|\Sigma| \times ||\mathcal{T}||)$ *time per maximal flexible pattern within* $O(||\mathcal{T}||d)$ *space, where* d *is the maximum number of gaps in a flexible pattern* $P \in \mathcal{M}_\sigma$*.*

Proof. The correctness of the algorithm is clear, thus we discuss the complexity. From Lemmas 14 and 15, the computation time for checking the maximality of a flexible pattern can be done in $O(||\mathcal{T}||)$ time, by using p_{max} and *succ*. For the task, *succ* is needed only for segments w and consecutive segments ww' in the current pattern P. Thus, the memory complexity is $O(||\mathcal{T}||d)$.

We next show how much time we need to compute those for a flexible pattern $a \circledcirc P$ by using those for P. Actually, for any $i > 0$ and position p, $succ(w_i, T, p)$, $succ(w_i w_j, T, p)$, and $p_{max}(w_i * \cdots * w_d, T)$ are common to P and P', hence we have to compute those only for $i = 0$. Thus, it can be done in $O(||\mathcal{T}||)$ time. We have at most $2|\Sigma|$ candidates for the children of a maximal flexible pattern, thus the statement holds. □

The *delay* of an enumeration algorithm is the maximum computation time between a pair of consecutive outputs.

Corollary 4. *The maximal pattern enumeration problem for the class $\mathcal{FP}$ of flexible patterns w.r.t. position maximality is solvable in polynomial delay and polynomial space in the input size.*

5 Conclusion

In this paper, we considered the maximal pattern discovery problem for the class $\mathcal{FP}$ of flexible patterns [6], which are also called erasing regular patterns in machine learning. The motivation of this study is the potential application to the optimal pattern discovery problem in machine learning and knowledge discovery. Our main result was a polynomial space and polynomial delay algorithm for enumerating all maximal patterns appearing in a given string without duplicates in terms of position-maximality defined through the equivalence relation between the location sets. Extending this work to document-based maximal patterns and to more complex classes of flexible patterns are interesting future problems.

Acknowledgments

The authors would like to thank Tetsuji Kuboyama, Akira Ishino, Kimihito Ito, Shinichi Shimozono, Takuya Kida, Shin-ichi Minato, Ayumi Shinohara, Masayuki Takeda, Kouichi Hirata, Akihiro Yamamoto, Thomas Zeugmann, and Ken Satoh, for their valuable discussions and comments. The authors also thank the anonymous referees for their valuable comments that greatly improved the quality of this paper.

References

1. Avis, D., Fukuda, K.: Reverse Search for Enumeration. Discrete Appl. Math. 65, 21–46 (1996)
2. Arimura, H., Fujino, R., Shinohara, T.: Protein motif discovery from positive examples by minimal multiple generalization over regular patterns. In: Proc. GIW 1994, pp. 39–48 (1994)

3. Arimura, H., Shinohara, T., Otsuki, S.: Finding minimal generalizations for unions of pattern languages and its application to inductive inference from positive data. In: Enjalbert, P., Mayr, E.W., Wagner, K.W. (eds.) STACS 1994. LNCS, vol. 775, pp. 649–660. Springer, Heidelberg (1994)
4. Arimura, H., Uno, T.: A polynomial space and polynomial delay algorithm for enumeration of maximal motifs in a sequence. In: Deng, X., Du, D.-Z. (eds.) ISAAC 2005. LNCS, vol. 3827, Springer, Heidelberg (2005)
5. Mannila, H., Toivonen, H., Verkamo, A.I.: Discovery of frequent episodes in event sequences. Data Min. Knowl. Discov. 1(3), 259–289 (1997)
6. Parida, L., Rigoutsos, I., et al.: Pattern discovery on character sets and real-valued data: Linear-bound on irredandant motifs and efficient polynomial time algorithms. In: Proc. SODA 2000, SIAM-ACM (2000)
7. Pisanti, N., et al.: A basis of tiling motifs for generating repeated patterns and its complexity of higher quorum. In: Rovan, B., Vojtáš, P. (eds.) MFCS 2003. LNCS, vol. 2747, Springer, Heidelberg (2003)
8. Shapiro, E.Y.: Algorithmic Program Debugging. MIT Press, Cambridge (1982)
9. Shimozono, S., Arimura, H., Arikawa, S.: Efficient discovery of optimal word-association patterns in large text databases. New Generation Comput. 18(1), 49–60 (2000)
10. Shinohara, T.: Polynomial time inference of extended regular pattern Languages. In: Proc. RIMS Symp. on Software Sci. & Eng., pp. 115–127 (1982)
11. Yan, X., Han, J., Afshar, R.: CloSpan: Mining closed sequential patterns in large databases. In: Proc. SDM 2003, SIAM (2003)
12. Wang, J., Han, J.: BIDE: efficient mining of frequent closed sequences. In: Proc. ICDE 2004 (2004)

Computing Characteristic Sets of Bounded Unions of Polynomial Ideals

Itsuo Takamatsu[1], Masanori Kobayashi[2], Hiroo Tokunaga[2], and Akihiro Yamamoto[3]

[1] Department of Mathematics
Tokyo Metropolitan University
1-1 Minami-Ohsawa, Hachiohji-shi 192-0397 Japan⋆
[2] Department of Mathematics and Information Science
Graduate School of Science and Engineering,
Tokyo Metropolitan University
1-1 Minami-Ohsawa, Hachiohji-shi 192-0397 Japan
kobayashi-masanori@tmu.ac.jp, tokunaga@tmu.ac.jp
[3] Graduate School of Informatics, Kyoto University
Yoshida Honmachi, Sakyo-ku, Kyoto 606-850 Japan
akihiro@i.kyoto-u.ac.jp

Abstract. The surprising fact that Hilbert's basis theorem in algebra shows identifiabilty of ideals of polynomials in the limit from positive data is derived by the correspondence between ideals and languages in the context of machine learning. This correspondence also reveals the difference between the two and raises new problems to be solved in both of algebra and machine learning. In this article we solve the problem of providing a concrete form of the characteristic set of a union of two polynomial ideals. Our previous work showed that the finite basis of every polynomial ideal is its characteristic set, which ensures that the class of ideals of polynomials is identifiable from positive data. Union or set-theoretic sum is a basic set operation, and it could be conjectured that there is some effective method which produces a characteristic set of a union of two polynomial ideals if both of the basis of ideals are given. Unfortunately, we cannot find a previous work which gives a general method for how to find characteristic sets of unions of languages even though the languages are in a class identifiable from positive data. We give methods for computing a characteristic set of the union of two polynomial ideals.

1 Introduction

One of the surprising facts in machine learning is that Hilbert's basis theorem in algebra shows that ideals in a Noetherian ring are identifiable in the limit from positive data [12]. This suggests that the interchange of algebra and machine learning could be investigated and would lead to rich contribution to both of the areas. As a starting of such interchange, we introduce in the present article

⋆ Current address: A503 175-9, Kozukue, Kouhoku, Yokohama, Japan.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 318–329, 2008.

a characteristics set of bounded unions into the class of ideals of polynomials, both of which have been used in investigating learnablity of formal languages from positive data. The results presented in this article will contribute both of the two areas as explained below.

1.1 Learnability of Polynomial Ideals

In this article we assume that the coefficient of every monomial is a rational number or an integer, and the variables in a polynomial are from a finite set $\{X_1, X_2, \ldots, X_n\}$. An ideal is a set of polynomials which is closed under addition of two elements in it and multiplication of an element in it and any polynomial. A basis of an ideal is one of its subsets, the elements of which generate all elements in the ideal. Hilbert's basis theorem ensures that every ideal of polynomials has a finite basis. Buchberger [2] showed a special type of basis, called a Groebner basis, exists for every ideal.

It is curious that quite similar results hold for learning formal languages. Angluin [1] showed that a uniformly recursive class of languages is identifiable in the limit from positive data if and only if a finite tell-tale of each language in the class is uniformly enumerable. In fact, when we regard every ideal as a formal language and the class of all ideals of polynomials as the class for learning, the finite basis of every ideal works both as its grammar and as its finite tell-tale, and therefore the class is identifiable from positive data [12]. This fact has promoted providing new semantics to logical formulae with learning [6].

In our previous work [10] we analyzed the correspondence by using two previous results on identification of languages in the limit from positive data: A uniformly recursive class C of languages is identifiable in the limit from positive data if every language in C has a characteristic set, but the converse does not hold in general [8]. Every language in C has a characteristic set if C has the finite elasticity property, but the converse does not hold in general [14,11]. We showed that the class of polynomial ideals has the finite elasticity property, and therefore every ideal has a characteristic set. We also showed that every basis of a polynomial ideal is a characteristic set.

1.2 Contributions of the Presented Results

The correspondence between ideals and languages in the context of machine learning reveals their difference and raises new problems to be solved in both of the areas of algebra and machine learning. We focus on the following fact: The finite elasticity property of a uniformly recursive class C of language ensures identifiability and is hereditary w.r.t. the union operation of languages, that is, the class of unions of two languages in C is identifiable [14,11]. Since unions of two ideals also have the finite elasticity property, every union has a characteristic set. In the context of algebra, a characteristic set of an ideal acts as its basis, so we can expect to extend the basis theorem to unions of ideals. However, unions (set-theoretic sum) of ideals have seldom interested mathematicians because the union of two ideals is not identical to the sum of the two. In the present article we give several contributions to the union of ideals on the point of language learning. We give a concrete form of the characteristic set of a union of two polynomial ideals.

Our results are in both of the areas of machine learning and algebra.

In machine learning, we are successful in providing an algorithm for computing a characteristic set of the union of two languages, where every language is a polynomial ideal. Since union is a basic operation for sets, it could be conjectured that there is some effective method which produces a characteristic set of a union of two polynomial ideals if both of the basis of ideals are given. Unfortunately, we cannot find a previous work which gives a general method for how to find characteristic sets of unions of languages even though the language class is identifiable from positive data. The method presented here is in a more general form so that it can be applied to computing a characteristic set of the union of at most N polynomial ideals.

In algebra, a concrete form of the characteristic set of a union of two ideals suggests that unions of two ideals might be useful. Since a finite basis of an ideal is a characteristic set, our previous result shows the existence of a "finite basis" for every union of two polynomial ideals and the result in this article provides how to compute it. One reason why the union of ideals has not been so interested in algebra so far is that it is not an ideal and no geometric meaning has been found for the union and another reason is that it is not shown the existence of a finite basis for the union. We removed the second reason and expect new mathematical results by using finite basis to unions of ideals.

This article is organized as follows: In Section 2 we give an overview of the relation between identification in the limit from positive data and ideals of a commutative ring. In Section 3 we present a method for computing a characteristic set of the union of at most N polynomial ideals whose coefficients are rational numbers. We discuss in Section 4 a method for computing a characteristic set of the union of two polynomial ideals whose coefficients are integers, and we give two examples of characteristic sets in Section 5. We conclude the article in the last section.

2 Identification in the Limit from Positive Data and Commutative Rings

2.1 Identification in the Limit from Positive Data

We first explain some conditions for identifiability from positive data briefly. For details, we refer to [9].

In the theory of learning of formal languages, a language L is a subset of a countable recursive set U, which is $\{0,1\}^*$ in many cases, such that L is expressed by $L = L(G)$, where G is a finite expression normally called a *grammar*. When we denote the set of all grammars by $\mathcal{H}$, a family of all languages is given by $\mathcal{L} = \{L(G) \mid G \in \mathcal{H}\}$. We call $\mathcal{H}$ a *hypothesis space* and every grammar in $\mathcal{H}$ a *hypothesis*. Throughout this article, we assume that there exists an algorithm as follows: For any $G \in \mathcal{H}$ and an arbitrary $w \in U$, one can determine by this algorithm if w is in $L(G)$ or not. When such an algorithm exists, we say that $\mathcal{L}$ is *uniformly recursive*.

A **positive presentation** (or **positive data**) of $L(G)$ is an infinite sequence σ: e_1, e_2, e_3, ... of $L(G)$ such that every element in $L(G)$ appears at least once. We consider an effective *learning machine* M which perpetually repeats requesting an input $e_n \in U$ and output a hypothesis $G_n \in \mathcal{H}$ for $n = 1, 2, 3, \ldots$. We say that M *identifies* $L(G)$ *in the limit from positive data* if it satisfies the following:

> For any positive data $\sigma : e_1, e_2, e_3, \ldots$ of $L(G)$ there exists a natural number N such that $L(G_n) = L(G)$ for any $n \geq N$, where G_n $(n = 1, 2, 3, \ldots)$ is a sequence of guesses output by M as above.

This identification is called *behaviorally correct* (BC) for distinguishing it from the following condition called *explanatory*(EX):

> For any positive data $\sigma : e_1, e_2, e_3, \ldots$ of $L(G)$ there exists an N such that $G_n = G_N$ and $L(G_N) = L(G)$ for any $n \geq N$.

Moreover, we say that M identifies $\mathcal{L}$ from positive data, if M identifies any $L \in \mathcal{L}$ from positive data. Under the assumption that $\mathcal{L}$ is uniformly recursive, the two definitions above are equivalent.

Angluin [1] gave an necessary and sufficient condition for a family of languages $\mathcal{L}$ to be identifiable from positive data by using a notion of *finite tell-tale sets*. In the present article, we consider a notion similar to it. A finite subset C of $L(G) \in \mathcal{L}$ is called a *characteristic set* of $L(G)$ if it satisfies the following condition:

$$\forall L(G') \in \mathcal{L},\, C \subseteq L(G') \Rightarrow L(G) \subseteq L(G').$$

Theorem 1 ([8]). A family of languages $\mathcal{L}$ is identifiable in the limit from positive data if every element $L \in \mathcal{L}$ has a characteristic set.

We next introduce another notion for identifiability from positive data. We say that a family of languages $\mathcal{L}$ has *infinite elasticity* if there exist an infinite sequence of words $w_0, w_1, \ldots, w_n, \ldots$ in U and an infinite sequence of languages $L(G_1), \ldots, L(G_n), \ldots$ satisfying the following condition:

- For $\forall n$, $\{w_0, w_1, \ldots, w_{n-1}\} \subseteq L(G_n)$ and
- $w_n \notin L(G_n)$.

We say that $\mathcal{L}$ has *finite elasticity* if it does not have infinite elasticity. Wright et al. [14,11] showed the following:

Theorem 2 ([14,11]). A family of languages $\mathcal{L}$ is identifiable in the limit from positive data if it has finite elasticity.

Theorem 3 ([14,11]). If $\mathcal{L}$ has finite elasticity, then every $L(G) \in \mathcal{L}$ has a characteristic set.

The condition given by Wright et al. is important when we consider identifiability of the family of the union of languages.

Theorem 4 ([14,11]). *If both $\mathcal{L}$ and $\mathcal{M}$ have finite elasticity, then $\mathcal{L} \cup \mathcal{M}$ has finite elasticity.*

2.2 Ideals of a Commutative Ring and Identification in the Limit from Positive Data

A **commutative ring** R is a set on which two operations, an addition $a+b$ and a multiplication $a \cdot b$ for $a, b \in R$, satisfy the following conditions:

- R is an abelian group with respect to $+$.
- R is a commutative monoid with respect to $\cdot$.
- $+$ and $\cdot$ satisfy distributivity multiplication

In this article, we simply call a commutative ring a **ring**. Also we assume that R is countable and recursive, and $+$ and $\cdot$ are computable. Examples of such rings are the set $\boldsymbol{Q}$ of all rational numbers and the set $\boldsymbol{Z}$ of all integers with usual addition and multiplication of numbers. The set of all polynomials whose variables are from $\{X_1, X_2, \ldots, X_n\}$ and coefficients are in a ring K is denoted by $K[X_1, X_2, \ldots, X_n]$. The set $K[X_1, X_2, \ldots, X_n]$ is also a ring. Both of $\boldsymbol{Q}[X_1, X_2, \ldots, X_n]$ and $\boldsymbol{Z}[X_1, X_2, \ldots, X_n]$ satisfy the conditions above.

A subset I of R is called an *ideal* if it satisfies the following conditions:

- For $\forall f, g \in I \mathrm{C} f - g \in I$.
- For $\forall f \in I$ and $\forall h \in R$, $hf \in I$.

For $f_1, f_2, \ldots, f_r \in R$, we define a subset of R to be

$$\langle f_1, f_2, \ldots, f_r \rangle = \left\{ \sum_{i=1}^{r} h_i f_i \,\middle|\, h_i \in R \right\}$$

and we call $\langle f_1, f_2, \ldots, f_r \rangle$ an *ideal generated by* $\{f_1, f_2, \ldots, f_r\}$. The set $\{f_1, f_2, \ldots, f_r\}$ is called a *basis*. An ideal I of R is said to be finitely generated if there exist finitely many elements $f_1, f_2, \ldots, f_r$ of R which generates I.

We regard a basis as a grammar, by which a finitely generated ideal is considered as a formal language. That is, the hypothesis space is the family of all finite subsets of R, and the family of all languages is the set of finite generated ideals:

$$\mathcal{I}' = \{\langle f_1, f_2, \ldots, f_r \rangle \mid f_1, f_2, \ldots, f_r \in R\}.$$

Thus we can consider identifiability of "finitely generated ideals" in the limit from positive data. Note that a basis of an ideal is not only a grammar, but it is also regarded as a characteristic set as follows:

Proposition 1. *For a finitely generated ideal $I = \langle f_1, f_2, \ldots, f_r \rangle$, $\{f_1, f_2, \ldots, f_r\}$ is a characteristic set of I in $\mathcal{I}'$.*

Hence $\mathcal{I}'$ is BC-identifiable in the limit from positive data.

For a ring R, not every ideal is finitely generated. For the polynomial ring $\boldsymbol{Q}[X_1, X_2, \ldots, X_n]$, however, the following fact is well-known:

Theorem 5 (Hilbert). *Any ideal of $\boldsymbol{Q}[X_1, X_2, \ldots, X_n]$ is finitely generatedD*

Buchberger[2] gave an algorithm for computing an ideal of a special type, called a *reduced Groebner basis*, for any ideal of $\boldsymbol{Q}[X_1, X_2, \ldots, X_n]$. This algorithm combined with Theorem 5 shows that the class of all ideals of $\boldsymbol{Q}[X_1, X_2, \ldots, X_n]$ is uniformly recursive, and therefore the class is EX-identifiable in the limit from positive data.

In order to generalize Theorem 5 from a view point of the identification in the limit from positive data, we assume that the family $\mathcal{I}$ of all ideals of R is described by grammars in the hypothesis space $\mathcal{H}$:

$$\mathcal{I} = \{I(G) \mid G \in \mathcal{H}\}. \tag{1}$$

Note that the hypothesis space $\mathcal{H}$ for $\mathcal{I}$ is not necessarily defined by subsets of R as that for $\mathcal{I}'$.

One can characterize a finite generation of any ideal of R by using the ascending chain condition (Noether condition). Namely, for any infinite ascending chain of ideals $I_1, I_2, \ldots, I_n, \ldots,$ of R, i.e.,

$$I_1 \subseteq I_2 \subseteq \cdots \subseteq I_n \subseteq \cdots$$

there exists N such that $I_N = I_{N+1} = \cdots$. A ring R is called a *Noetherian ring* if it satisfies the Noether condition. For a Noetherian ring, the following fact is well-known:

Theorem 6. *Any ideal of R is finitely generated $\Leftrightarrow$ R is Noetherian.*

Stephan and Ventsov [12] showed:

Theorem 7 ([12]). *R is Noetherian $\Leftrightarrow$ $\mathcal{I}$ is identifiable in the limit from positive data.*

We can refine the results almost from the definitions:

Theorem 8 ([10]). *$\mathcal{I}$ is identifiable in the limit from positive data $\Leftrightarrow$ $\mathcal{I}$ has finite elasticity.*

3 A Method for the Construction of a Characteristic Set of $\bigcup^N \mathcal{I}$ in $\boldsymbol{Q}[X_1, X_2, \ldots, X_n]$

We first note that for ideals the sum and the set-theoretic sum (i.e. union) are different. The sum of ideals $I_1, I_2, \ldots, I_k$ is the minimal ideal containing the set-theoretic sum $\cup_{j=1}^k I_j$, or equivalently, the ideal generated by $\cup_{j=1}^k I_j$.

In algebra, we usually use sums of ideals. Here we discuss the set-theoretic sum from the viewpoint of identification in the limit from positive data. For a fixed positive integer N we put

$$\cup^N \mathcal{I} = \{\cup_{j=1}^k I_j \mid I_1, I_2, \ldots, I_k \in \mathcal{I} \text{ and } k \leq N\}.$$

From Theorems 4, 7 and 8, for a Noetherian ring R, $\bigcup^N \mathcal{I}$ has the finite elasticity. Moreover from Theorem 3, each element of $\bigcup^N \mathcal{I}$ has a characteristic set. Proposition 1 shows that for a ideal I, a characteristic set and a base are the same notion. On the other hand, since Theorem 4 uses Ramsey's theorem and the pigeon-hole principle, we have not given yet a concrete construction of a characteristic set of the set-theoretic sum. In the following, we describe a way to construct a characteristic set of $\cup_{j=1}^k I_j$ in $R = \boldsymbol{Q}[X_1, X_2, \ldots, X_n]$ when we know a characteristic set for each I_j in $I_1, I_2, \ldots, I_k \in \mathcal{I}$ $(k \leq N)$.

First we do some linear algebra.

Lemma 1. *For given M, n $(M \geq n)$, there exist M vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M$ in the vector space $\boldsymbol{Q}^n$, such that any n vectors $\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \ldots, \boldsymbol{x}_{i_n}$ are linearly independent.*

Proof. Take an $n \times M$-matrix

$$A = \begin{pmatrix} X_{11} & X_{12} & \ldots & X_{1M} \\ X_{21} & X_{22} & \ldots & X_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \ldots & X_{nM} \end{pmatrix}$$

whose entries are variables. There are exactly $l = \binom{M}{n}$ n-th minors of A, which we denote by

$$f_{i_1, i_2, \ldots, i_n} = \det(\boldsymbol{X}_{i_1}, \boldsymbol{X}_{i_2}, \ldots, \boldsymbol{X}_{i_n})$$

where $i_1 < i_2 < \cdots < i_n$ and

$$\boldsymbol{X}_{i_j} = \begin{pmatrix} X_{1i_j} \\ \vdots \\ X_{ni_j} \end{pmatrix} \qquad (j = 1, 2, \ldots, n).$$

To prove the lemma, it is enough to show that there exists a point in $\boldsymbol{Q}^{nM}$ where

$$F = \prod_{i_1 < i_2 < \cdots < i_n} f_{i_1, i_2, \ldots, i_n}$$

does not vanish.

In general, from the factorization theorem of one variable, there exists only finitely many rational roots for nonzero polynomial of one variable. Since the cardinality of $\boldsymbol{Q}$ is infinite, there exists a point $a \in \boldsymbol{Q}$ such that the value at a is nonzero.

We proceed by induction on the number of variables. Assume that for any nonzero polynomial of $k-1$ variables, there exists a point in $\boldsymbol{Q}^{k-1}$ where the value of the polynomial is nonzero. For a nonzero polynomial of k variables, we write it as a polynomial of the last variable whose coefficients are polynomials of the first $k-1$ variables. By the hypothesis of induction, there exists a set of values of the first $k-1$ variables such that the highest term with respect to the last variable does not vanish. Thus again by the factorization theorem of one variable, there exists a point in $\boldsymbol{Q}^k$ where the value of the polynomial is nonzero. So for the case $k = nM$. □

With that fact in mind, we construct a characteristic set of $\cup_{j=1}^{k} I_j$ $(k \leq N)$. Suppose each I_i is represented as

$$I_i = \langle f_1^{(i)}, f_2^{(i)}, \dots, f_{n_i}^{(i)} \rangle,$$
$$f_k^{(i)} \in \boldsymbol{Q}[X_1, X_2, \dots, X_n] \qquad (k = 1, 2, \dots, n_i).$$

We re-choose a basis of I_i as follows.

By lemma 1 above, there exist M_i vectors $\boldsymbol{c}_1, \dots, \boldsymbol{c}_{M_i}$ in $\boldsymbol{Q}^{n_i}$ which satisfy

1. $M_i = N(n_i - 1) + 1$.
2. Any n_i vectors $\boldsymbol{c}_{k_1}, \dots, \boldsymbol{c}_{k_{n_i}}$ are linearly independent.

Using the vector

$$\boldsymbol{c}_k = \begin{pmatrix} c_{1k} \\ \vdots \\ c_{n_i k} \end{pmatrix} \qquad (k = 1, \dots, M_i),$$

we define the following polynomial:

$$\begin{cases} h_{k_1}^{(i)} = c_{1k_1} f_1^{(i)} + c_{1k_2} f_2^{(i)} + \cdots + c_{1k_{n_i}} f_{n_i}^{(i)} \\ \qquad \vdots \\ h_{k_{n_i}}^{(i)} = c_{n_i k_1} f_1^{(i)} + c_{n_i k_2} f_2^{(i)} + \cdots + c_{n_i k_{n_i}} f_{n_i}^{(i)} \end{cases}$$

As matrix expression this becomes

$$\begin{pmatrix} h_{k_1}^{(i)} \\ \vdots \\ h_{k_{n_i}}^{(i)} \end{pmatrix} = \begin{pmatrix} c_{1k_1} & \cdots & c_{1k_{n_i}} \\ \vdots & \ddots & \vdots \\ c_{n_i k_1} & \cdots & c_{n_i k_{n_i}} \end{pmatrix} \begin{pmatrix} f_1^{(i)} \\ \vdots \\ f_{n_i}^{(i)} \end{pmatrix}.$$

The coefficient matrix C_i of the formula above is a regular matrix of degree n_i. Thus, $\{h_{k_1}^{(i)}, h_{k_2}^{(i)}, \dots, h_{k_{n_i}}^{(i)}\}$ is a basis of I_i. We get M_i polynomials $h_1^{(i)}, h_2^{(i)}, \dots, h_{M_i}^{(i)}$.

Theorem 9. *$S = \bigcup_{i=1}^{k} \{h_1^{(i)}, h_2^{(i)}, \dots, h_{M_i}^{(i)}\}$ is a characteristic set of $\bigcup_{i=1}^{k} I_i$.*

Proof. First we note that S is finite since it is a finite sum of finite sets. Next suppose that for $\bigcup_{j=1}^{k} J_j \in \bigcup^{N} \mathcal{I}$,

$$S \subseteq \bigcup_{j=1}^{k} J_j \in \bigcup^{N} \mathcal{I}$$

holds. Then for each i, we have

$$H_i = \{h_1^{(i)}, h_2^{(i)}, \ldots, h_{M_i}^{(i)}\} \subseteq \bigcup_{j=1}^{k} J_j \in \bigcup^{N} \mathcal{I}.$$

Here H_i decomposes to at most N sets, that is, each element of H_i belongs to some ideal J_j. So if we decompose H_i as the sum of $H_{i,l}$ $(l = 1, 2, \ldots, N)$, and assume moreover that

$(*)$ for any $H_{i,l}$, $\sharp H_{i,l} \leq n_i - 1$,

then we have

$$M_i = \sum_{l=1}^{N} \sharp H_{i,l} \leq N(n_i - 1),$$

which contradicts to the definition of M_i. Thus $(*)$ does not hold. There exists some l which satisfies

$$n_i \leq n_i' = \sharp H_{i,l},\ \ H_{i,l} \subseteq J_l.$$

Since any n_i elements of $H_{i,l}$ becomes a basis of I_i,

$$I_i \subseteq J_l$$

holds. Since this holds for each I_i, we have

$$\bigcup_{i=1}^{k} I_i \subseteq \bigcup_{j=1}^{k} J_j.$$

This shows that S is a characteristic set of $\bigcup_{i=1}^{k} I_i$. □

4 Constructing a Characteristic Set of Elements of $\mathcal{I} \cup \mathcal{I}$ for $\boldsymbol{Z}[X_1, X_2, \ldots, X_n]$

In the previous section we gave a construction of a characteristic set for elements of $\bigcup^{N} \mathcal{I}$ in $\boldsymbol{Q}[X_1, X_2, \ldots, X_n]$. In this section we consider a similar problem for $\boldsymbol{Z}[X_1, X_2, \ldots, X_n]$, which is also a Noetherian ring.

If we can choose the vector in Lemma 1 so that (i) all components are integers, (ii) all n-th minors are ± 1, then we can construct a characteristic set in a completely similar manner as in the previous section. However it is difficult yet to prove the existence of an 'integer' solution for the polynomial F. So here we consider a characteristic set of elements of $\mathcal{I} \cup \mathcal{I}$ as a first step.

Lemma 2. *There exists a finite subset* $S = \{\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \ldots\ , \boldsymbol{a}_M\}$ *of* $\boldsymbol{Z}^n$ *which satisfies that for any bi-partition of* S*,*

$$S = S_1 \cup S_2, \qquad S_1 \cap S_2\ = \phi,$$

and S_1 or S_2 generates $\boldsymbol{Z}^n$. In fact, it is enough to choose $M = \frac{1}{2}n(n+1)$, and $\boldsymbol{a}_i$ $(i = 1, 2, \ldots, M)$ as the column vectors of the following $n \times M$ integer matrix A_n.

$$A_n = (\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \ldots\ , \boldsymbol{a}_M)$$

$$= \begin{pmatrix} 1\,1\,1 \cdots 1\,0\,0 \cdots 0 \cdots 0 \\ 0\,1\,0 \cdots 0\,1\,1 \cdots 1 \cdots 0 \\ 0\,0\,1 \cdots 0\,0\,1 \cdots 0 \cdots 0 \\ \vdots\,\vdots\,\vdots\,\ddots\,\vdots\,\vdots\,\vdots\,\ddots\,\vdots \cdots \vdots \\ 0\,0\,0 \cdots 1\,0\,0 \cdots 1 \cdots 1 \end{pmatrix}.$$

The proof is by induction on n.

Using Lemma 2, it is easy to see that a characteristic set of elements of $\mathcal{I} \cup \mathcal{I}$ is given by an argument similar to the one used in the previous section.

5 Examples

In this section, we give two explicit examples of characteristic sets of set unions.

Example 1. Let $R = \boldsymbol{Q}[X_1, X_2]$ and let I_1, I_2 and I_3 be ideals given by

$$\begin{aligned} I_1 &= \langle X_1^2, X_2^2 \rangle, \\ I_2 &= \langle X_1^2 + X_2, X_1^3 + X_2^4 \rangle, \text{ and} \\ I_3 &= \langle X_1^2 + X_2^2 - 1, X_2^2 - X_1^3 \rangle. \end{aligned}$$

Consider a 2×4 matrix M as follows:

$$M = \begin{pmatrix} 1\,0\,1\,1 \\ 0\,1\,1\,2 \end{pmatrix}.$$

As every 2-minor of M is non-zero, two of any 4 column vectors of M are linearly independent. Hence by Lemma 1 and Theorem 9,

$$\begin{aligned} &\{X_1^2, X_2^2, X_1^2 + X_2^2, X_1^2 + 2X_2^2\} \\ \bigcup\ &\{X_1^2 + X_2, X_1^3 + X_2^4, (X_1^2 + X_2) + (X_1^3 + X_2^4), (X_1^2 + X_2) + 2(X_1^3 + X_2^4)\} \\ \bigcup\ &\{X_1^2 + X_2^2 - 1, X_2^2 - X_1^3, (X_1^2 + X_2^2 - 1) + (X_2^2 - X_1^3), (X_1^2 + X_2^2 - 1) + 2(X_2^2 - X_1^3)\} \end{aligned}$$

is a characteristic set of $I_1 \cup I_2 \cup I_3$ in $\bigcup^3 \mathcal{I}$.

Example 2. Let $R = \boldsymbol{Z}[X_1, X_2]$ and let J_1 and J_2 be ideals given by

$$\begin{aligned} J_1 &= I_1 \cap \boldsymbol{Z}[X_1, X_2], \text{ and} \\ J_2 &= I_2 \cap \boldsymbol{Z}[X_1, X_2], \end{aligned}$$

where I_i $(i = 1, 2)$ as in Example 1, respectively. Consider a 2×3 matrix N as follows:

$$N = \begin{pmatrix} 1\,0\,1 \\ 0\,1\,1 \end{pmatrix}.$$

As every 2-minor of N is either 1 or -1, one can apply the argument in section 4 and obtains a characteristic set

$$\{X_1^2, X_2^2, X_1^2 + X_2^2\}$$
$$\bigcup \{X_1^2 + X_2, X_1^3 + X_2^4, (X_1^2 + X_2) + (X_1^3 + X_2^4)\}$$

of $J_1 \cup J_2$ in $\cup^2 \mathcal{I}_{\boldsymbol{Z}}$, where $\mathcal{I}_{\boldsymbol{Z}}$ denotes the set of ideals in $\boldsymbol{Z}[X_1, X_2]$.

We end this section with explaining a difference between $\boldsymbol{Q}$ and $\boldsymbol{Z}$ coefficients through the above examples.

We first note that $\{X_1^2, X_1^2 + 2X_2^2\}$ is not a basis of J_1. In fact, suppose that J_1 is generated by X_1^2 and $X_1^2 + 2X_2^2$. Then as $X_2^2 \in J_1$, we have

$$X_2^2 = f_1(X_1, X_2)X_1^2 + f_2(X_1, X_2)(X_1^2 + 2X_2^2)$$

for some $f_1, f_2 \in \boldsymbol{Z}[X_1, X_2]$. But this is impossible because each monomial in the right hand side which is divisible by X_2^2 satisfies that either

(i) it is also divisible by X_1^2, or
(ii) its coefficient is an even integer.

Hence $\langle X_1^2, X_1^2 + 2X_2^2\rangle$ is properly contained in J_1. This means that one can not use the matrix M in Example 1 in order to compute a characteristic set for a language in $\cup^3 \mathcal{I}_{\boldsymbol{Z}}$. For example, let $\langle X_2^2\rangle, \langle X_1^2 + X_2^2\rangle$ and $\langle X_1^2, X_1^2 + 2X_2^2\rangle$ be ideals in $\boldsymbol{Z}[X_1, X_2]$. Then

$$\{X_2^2, X_1^2 + X_2^2, X_1^2, X_1^2 + 2X_2^2\} \subset \langle X_2^2\rangle \cup \langle X_1^2 + X_2^2\rangle \cup \langle X_1^2, X_1^2 + 2X_2^2\rangle,$$

but the right hand side is properly contained in $\langle X_2^2\rangle \cup \langle X_1^2 + X_2^2\rangle \cup \langle X_1^2, X_2^2\rangle$, or equivalently, in $\langle X_1^2, X_2^2\rangle$.

6 Conclusion

In this article we have shown how to describe a characteristic set of a finite set-theoretic sum of ideals in a polynomial ring. When we identify the finite set-theoretic sum of formal languages in the limit, we formulate the property of finite elasticity as a property of the whole family of languages and show the finite elasticity for a concrete family of languages by contradiction (reductio ad absurdum), in which case generally no characteristic sets or finite tell-tales are shown. The proof of this article uses a contradiction, but it is reduced to the pigeon-hole principle. The pigeon-hole principle is used to show that a family of languages with finite elasticity can be identified in the limit from positive data. In this sense we think the description of a characteristic set of set-theoretic sum of ideals considered in this article natural.

The polynomial ring is defined by use of algebraic operations between elements. It has a characteristic that a basis, which is a set of elements, has a function as a grammar. In the learning of formal languages, pattern languages have a similar property as ideals. So we expect that they have similar results as in this article.

Acknowledgment

This research is partially supported by Grant-in-Aid for Scientific Research (B) 19300046 from JSPS. The third author is partially supported by K18-XI-234 from Kayamori Foundation of Informational Science Advancement.

References

1. Angluin, D.: Inductive Inference of Formal Languages from Positive Data. Information and Control 45, 117–135 (1980)
2. Buchberger, B.: Groebner Bases: An Algorithmic Method in Polynomial Ideal Theory. In: Bose, N.K. (ed.) Multidimensional System Theory, pp. 184–232. D. Reidel Publishing Company, Dordrecht (1985)
3. Cox, D., Little, J., O'Shea, D.: Ideals, Varieties, and Algorithms. Springer, Heidelberg (1992) (revised version, 1997)
4. de Brecht, M., Kobayashi, M., Tokunaga, H., Yamamoto, A.: Inferability of Closed Set Systems From Positive Data. In: Washio, T., Satoh, K., Takeda, H., Inokuchi, A. (eds.) JSAI 2006. LNCS (LNAI), vol. 4384, pp. 265–275. Springer, Heidelberg (2007)
5. Gold, D.: Language Identfication in the Limit. Information and Control 10, 447–474 (1967)
6. Hayashi, S.: Mathematics based on Incremental Learning –Excluded middle and Inductive inference. Theoretical Computer Science 350(1) (2006)
7. Jain, S., Osherson, D., Royer, J.S., Sharma, A.: Systems That Learn – An introduction to Learning Theory, 2nd edn. MIT Press, Cambridge (1999)
8. Kobayashi, S.: Approximate Identification, Finite Elasticity and Lattice Structure of Hypothesis Space, Technical Report, CSIM 96-04, Dept. of Compt. Sci. and Inform. Math., Univ. of Electro-Communications (1996)
9. Sakakibara, Y., Kobayashi, S., Yokomori, T.: Computational Learning Theory (in Japanese), Baifukan (2001)
10. Kobayashi, M., Tokunaga, H., Yamamoto, A.: Ideals of Polynomial Rings and Learning from Positive Data (in Japanese). In: Proc. 8th IBIS, pp. 129–134 (2005)
11. Motoki, T., Shinohara, T., Wright, K.: The Correct Definition of Finite Elasticity: Corrigendum to Identification of Unions. In: Proceedings of COLT 1991, pp. 587–626. Morgan-Kaufman, San Francisco (1991)
12. Stephan, F., Ventsov, Y.: Learning Algebraic Structures from Text. Theoretical Computer Science 268, 221–273 (2001)
13. van der Waerden, B.L.: Moderne Algebra, vol. II. Spiringer, Heidelberg (1940)
14. Wright, K.: Identification of Unions of Languages Drawn from an Identifiable Class. In: Proceedings of COLT 1989, pp. 328–388. Morgan Kaufmann, San Francisco (1989)

Towards a Logical Reconstruction of CF-Induction

Yoshitaka Yamamoto[1], Oliver Ray[2], and Katsumi Inoue[1,3]

[1] Department of Informatics, Graduate University for Advanced Studies
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
yy@nii.ac.jp
[2] Department of Computer Science, University of Bristol
Bristol BS8 1UB, United Kingdom
oray@cs.bris.ac.uk
[3] National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
ki@nii.ac.jp

Abstract. CF-induction is a sound and complete hypothesis finding procedure for full clausal logic which uses the principle of inverse entailment to compute a hypothesis that logically explains a set of examples with respect to a prior background theory. Currently, CF-induction computes hypotheses by applying combinations of several complex generalisation operators to an intermediate theory called a bridge formula. In this paper we propose an alternative approach whereby hypotheses are derived from a bridge formula using a single deductive operator and a single inductive operator. We show that our simplified procedure preserves the soundness and completeness of CF-induction.

Keywords: inverse entailment, CF-induction, generalisation operator.

1 Introduction

Given a background theory B and observations E, the task of explanatory induction is to find a hypothesis H such that $B \wedge H \models E$ and $B \wedge H$ is consistent [5]. By the principle of *inverse entailment* (IE) [9] this is logically equivalent to find a consistent hypothesis H such that $B \wedge \neg E \models \neg H$. This equivalence means that the inductive hypothesis H can be computed by deducing its negation $\neg H$ from B and $\neg E$. This allows highly developed *deductive* reasoning techniques to be used in the implementation of *inductive* learning systems.

IE is the basis of many successful Horn clause systems such as Progol [9], which have been used in real application domains. Recently, IE methods have been developed for full clausal theories to enable the solution of more complex problems in richer knowledge representation formalisms. One such method is CF-induction [5], which has two important benefits: unlike some related systems, such as FC-HAIL [15], CF-induction is complete for finding full clausal hypotheses; and unlike other related systems, such as the Residue Procedure

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 330–343, 2008.

[17], CF-induction can exploit language bias to focus the procedure on some relevant part of the search space specified by the user.

CF-induction computes a hypothesis in two steps: by first constructing an intermediate bridge formula, denoted CC, such that $B \wedge \neg E \models CC$ and then generalising the negation of this bridge formula to obtain a hypothesis H such that $\neg CC \mathrel{=\!\!\!|} H$, where $\mathrel{=\!\!\!|}$ denotes the inverse of the classical logical entailment relation $\models$. Hereafter, we call $\mathrel{=\!\!\!|}$ a *generalisation relation* and we refer to its associated operators as *generalisation operators* or simply *generalisers*.

The bridge formula CC is a finite set of instances of so-called *characteristic clauses*: which are the consequences of $B \wedge \neg E$ that satisfy a given language bias and are minimal with respect to subsumption. This bridge formula is generalised using a combination of several well-known generalisers, such as *least generalisation*, *inverse resolution* and *anti-subsumption*, to obtain H from $\neg CC$. In order to derive a particular hypothesis H it may be necessary to apply several different generalisers in some specific order, i.e.,

$$\neg CC \mathrel{=\!\!\!|} U_1 \mathrel{=\!\!\!|} U_2 \mathrel{=\!\!\!|} \cdots \mathrel{=\!\!\!|} U_{n-1} \mathrel{=\!\!\!|} U_n \mathrel{=\!\!\!|} H. \tag{1}$$

The fact that these generalisers can be sequenced in many different ways, results in a large search space. Our long-term research objective is to reduce the non-determinism of CF-induction in order to make the procedure easier to apply in real-world applications.

In this paper, we concentrate on simplifying the process of obtaining hypotheses from a given bridge formula. In particular, we propose an alternative approach for CF-induction based on the observation that the negated hypothesis $\neg H$ can be computed deductively from CC, i.e.,

$$CC \models V_1 \models V_2 \models \cdots \models V_{m-1} \models V_m \models \neg H. \tag{2}$$

In the interests of computational efficiency, it is convenient to work in clausal form logic. In this case, $\neg H$ will be represented by a clausal theory G in which any existentially quantified variables in $\neg H$ are substituted by ground (Skolem) terms. Therefore, our new approach works in three steps: first G is deduced from CC; then it is negated; and finally anti-instantiation is used (to replace any Skolem terms by variables) to result in the final hypothesis H.

To derive G from CC, we introduce a new deductive operator, which can be regarded as simplifying the existing operators of *subsumption*, *resolution* and *weakening*. This new operator, called the *γ-operator*, warrants the insertion of literals into the clauses of CC. We show that this single deductive operator (followed by negation and anti-instantiation) is sufficient to preserve the soundness and completeness of CF-induction. This new approach, called *CF-induction with γ-operator*, simplifies the original CF-induction in the sense that it only requires one deductive operator (γ-operator) and one generalisation operator (anti-instantiation) to be used in the construction of H from CC.

The rest of this paper is organized as follows. Section 2 introduces the theoretical background in this paper and reviews the previous procedure of CF-induction. Section 3 introduces three standard deductive operators of subsumption, resolution and weakening, and provides several results concerning the order of applying these operators. Section 4 describes the γ-operator and provides the basic idea for simplifying the CF-induction procedure. Section 5 discusses some related work, and Section 6 concludes.

2 Background

2.1 Notation and Terminology

Here, we review the notation and terminology used in the paper. A *clause* is a finite disjunction of literals which is often identified with the set of its literals. A clause $\{A_1, \ldots, A_n, \neg B_1, \ldots, \neg B_m\}$, where each A_i, B_j is an atom, is also written as $B_1 \wedge \cdots \wedge B_m \supset A_1 \vee \cdots \vee A_n$. A *definite clause* is a clause which contains only one positive literal. A *positive* (*negative*) *clause* is a clause whose disjuncts are all positive (negative) literals. A *Horn clause* is a definite clause or negative clause. A *unit clause* is a clause with exactly one literal. A *clausal theory* is a finite set of clauses. A clausal theory is *full* if it contains at least one non-Horn clause. A clausal theory Σ is often identified with the conjunction of its clauses and is said to be in *Conjunctive Normal Form* (CNF). The *complement* of Σ, denoted $\overline{\Sigma}$, is defined as a clausal theory obtained by translating $\neg\Sigma$ into CNF using a standard translation procedure [12]. (In brief, $\overline{\Sigma}$ is obtained by converting $\neg\Sigma$ into prenex conjunctive normal form with standard equivalence-preserving operations and skolemising it.) Two clausal theories Σ_1 and Σ_2 are said to be equivalent, denoted $\Sigma_1 \equiv \Sigma_2$ if $\Sigma_1 \models \Sigma_2$ and $\Sigma_2 \models \Sigma_1$. Let C and D be two clauses. C *subsumes* D, denoted $C \succeq D$, if there is a substitution θ such that $C\theta \subseteq D$. C *properly* subsumes D if $C \succeq D$ but $D \not\succeq C$. For a clausal theory Σ, $\mu\Sigma$ denotes the set of clauses in Σ not properly subsumed by any clause in Σ. Let Σ be a clausal theory and C be a clause. A *derivation* of C from Σ is a finite sequence of clauses $R_1, \ldots, R_k = C$ such that each R_i is either in Σ, or is a resolvent of two clauses in $\{R_1, \ldots, R_{i-1}\}$. We recall the following result, called the *Subsumption Theorem.*

Theorem 1. [7,12] *Let Σ be a clausal theory and C be a clause. Then $\Sigma \models C$ iff C is a tautology or there exists a derivation of a clause D that subsumes C.*

For a clausal theory Σ, a *consequence* of Σ is a clause entailed by Σ. We denote by $Th(\Sigma)$ the set of all consequences of Σ. The Subsumption Theorem states the completeness of the resolution principle for finding any non-tautological consequences in $\mu Th(\Sigma)$.

We give the definition of *hypothesis* H in the logical setting of learning from positive examples.

Definition 1 (Hypothesis). *Let B and E be clausal theories, representing a background theory and (positive) examples/observations, respectively. Then H is a hypothesis wrt B and E iff H is a clausal theory such that $B \wedge H \models E$ and $B \wedge H$ is consistent.*

If no confusion arises, then we refer simply to a "hypothesis" instead of a "hypothesis wrt B and E". Note that, for a treatment of negative examples in CF-induction, we refer the reader to [5].

2.2 CF-Induction

CF-induction is a sound and complete method for IE. It is based on the notion of *characteristic clauses* which represent "interesting" consequences of a given problem for users [4]. Each characteristic clause is constructed over a sub-vocabulary of the representation language called a *production field*. A production field $\mathcal{P}$ is defined as a pair, $\langle \mathbf{L}, \mathit{Cond} \rangle$, where $\mathbf{L}$ is a set of literals closed under instantiation, and *Cond* is a certain condition to be satisfied, e.g., the maximum length of clauses, the maximum depth of terms, etc. When *Cond* is not specified, $\mathcal{P}$ is simply denoted as $\langle \mathbf{L} \rangle$. A clause C *belongs* to $\mathcal{P} = \langle \mathbf{L}, \mathit{Cond} \rangle$ if every literal in C belongs to $\mathbf{L}$ and C satisfies *Cond*. For a set Σ of clauses, the set of consequences of Σ belonging to $\mathcal{P}$ is denoted $Th_{\mathcal{P}}(\Sigma)$. Then, the characteristic clauses of Σ wrt $\mathcal{P}$ are defined as:

$$Carc(\Sigma, \mathcal{P}) = \mu Th_{\mathcal{P}}(\Sigma)$$

Note that $Carc(\Sigma, \mathcal{P})$ can, in general, include tautological clauses [4].

When a new clause F is added to a clausal theory, some consequences are newly derived with this additional information. The set of such clauses that belong to the production field are called *new characteristic clauses*. Formally, the *new characteristic clauses* of F wrt Σ and $\mathcal{P}$ are defined as:

$$NewCarc(\Sigma, F, \mathcal{P}) = Carc(\Sigma \cup \{F\}, \mathcal{P}) - Carc(\Sigma, \mathcal{P})$$

In the following, we assume the production field $\mathcal{P} = \langle \mathbf{L} \rangle$ where $\mathbf{L}$ is a set of literals reflecting an *inductive bias* whose literals are the negations of those literals we wish to allow in hypothesis clauses. When no inductive bias is considered, $\mathcal{P}$ is just set to $\langle \mathcal{L} \rangle$, where $\mathcal{L}$ is the set of all literals in the first-order language. We say H is a *hypothesis wrt* B*,* E *and* $\mathcal{P}$ iff H is a hypothesis wrt B and E, and, for every literal L appearing in H, its complement $\overline{L}$ is in $\mathbf{L}$. Then it holds that for any hypothesis H wrt B, E and $\mathcal{P}$,

$$B \wedge \overline{E} \models Carc(B \wedge \overline{E}, \mathcal{P}) \models \neg H; \tag{3}$$

$$B \models Carc(B, \mathcal{P}) \not\models \neg H. \tag{4}$$

The two formulae above follow from the principle of IE and the definition of characteristic clauses. In particular, Formula (3) implies we can use characteristic clauses to construct intermediate bridge formulae for IE. Formula (4) ensures the consistency of the hypothesis and background theory. As explained in [5], this can always be ensured by including at least one clause from $NewCarc(B, \overline{E}, \mathcal{P})$ in an intermediate bridge formula. The definition of a bridge formula in CF-induction is as follows:

Definition 2 (Bridge formula [5]). *For given* B*,* E *and* $\mathcal{P}$*, a clausal theory* CC *is a bridge formula wrt* B*,* E *and* $\mathcal{P}$ *iff* CC *satisfies the following conditions:*

1. *Each clause* $C_i \in CC$ *is an instance of a clause in* $Carc(B \wedge \overline{E}, \mathcal{P})$*;*
2. *At least one* $C_i \in CC$ *is an instance of a clause from* $NewCarc(B, \overline{E}, \mathcal{P})$*.*

If no confusion arises, a "bridge formula wrt B, E and $\mathcal{P}$" will simply be called a "bridge formula".

Theorem 2. [5] *Let B, E be clausal theories and $\mathcal{P}$ be a production field. Then, for any hypothesis H wrt B, E and $\mathcal{P}$, there exists a bridge formula CC wrt B, E and $\mathcal{P}$ such that $H \models \neg CC$.*

Theorem 2 shows that any hypothesis can be computed by constructing and generalising the negation $\neg CC$ of a set of characteristic clauses CC. In the original CF-induction, a bridge formula CC is first selected. Then a clausal theory F is obtained by Skolemising $\neg CC$ and translating it to CNF. Finally, H is obtained by applying a series of generalisers to F under the constraint that $B \wedge H$ is consistent. Many such generalisers have been proposed, such as, *reverse skolemisation* [2] (converting Skolem constants/functions to existentially quantified variables), *anti-instantiation* (replacing ground subterms with variables), *anti-weakening* (adding some clauses), *anti-subsumption* (dropping some literals from a clause), *inverse resolution* [10] (applying the inverse of the resolution principle) and *Plotkin's least generalisation* [13].

Example 1. Define a background theory B, an example E and a production field $\mathcal{P}$ as follows:

$$B = natural(0) \vee even(0),$$
$$E = natural(s(0)),$$
$$\mathcal{P} = \langle \{natural(X), \neg natural(X), even(X), \neg even(X)\} \rangle.$$

Then, $NewCarc(B, \overline{E}, \mathcal{P})$ and $Carc(B \wedge \overline{E}, \mathcal{P})$ are as follows:

$$NewCarc(B, \overline{E}, \mathcal{P}) = \neg natural(s(0)),$$
$$Carc(B \wedge \overline{E}, \mathcal{P}) = (natural(0) \vee even(0)) \wedge \neg natural(s(0)) \wedge Taut.$$

where $Taut$ denotes the tautological clauses in $Carc(B \wedge \overline{E}, \mathcal{P})$.

Let CC be a clausal theory $(natural(0) \vee even(0)) \wedge \neg natural(s(0))$. Since each clause in CC is a clause in $Carc(B \wedge \overline{E}, \mathcal{P})$ and the unit clause $\neg natural(s(0))$ in CC is a clause in $NewCarc(B, \overline{E}, \mathcal{P})$, CC is a bridge formula wrt B, E and $\mathcal{P}$ by Definition 2. The clausal theory F is obtained by translating $\neg CC$ into CNF as follows:

$$F = (natural(0) \supset natural(s(0))) \wedge (even(0) \supset natural(s(0))).$$

Since $B \wedge F \models E$ and $B \wedge F$ is consistent, F is a hypothesis wrt B, E and $\mathcal{P}$.

Assume that an inverse resolution generaliser is applied to F in such a way that the clause $C_1 = natural(0) \supset natural(s(0))$ in F is replaced with the clause $D_1 = natural(0) \supset even(0)$, which is treated as a parent clause of C_1. This means C_1 is the resolvent of D_1 and $C_2 = even(0) \supset natural(s(0))$ in F. Then the following clausal theory F'_1 is constructed:

$$F'_1 = (natural(0) \supset even(0)) \wedge (even(0) \supset natural(s(0))).$$

Since $B \wedge F'_1 \models E$ and $B \wedge F'_1$ is consistent, F'_1 is a hypothesis wrt B, E and $\mathcal{P}$.

Next, assume that another inverse resolution generaliser is applied to F in such a way that the clause C_2 in F is replaced with the clause $D_2 = even(0) \supset natural(0)$. Then the following clausal theory F_2' is constructed.

$$F_2' = (natural(0) \supset natural(s(0))) \wedge (even(0) \supset natural(0))$$

Since $B \wedge F_2' \models E$ and $B \wedge F_2'$ is consistent, F_2' is also a hypothesis wrt B, E and $\mathcal{P}$. In addition to the above generaliser, if we apply an anti-instantiation generaliser to F_2' in such a way that the ground term 0 occurring in F_2' is replaced with the variable X, then the following theory is obtained:

$$F_3' = (natural(X) \supset natural(s(X))) \wedge (even(X) \supset natural(X)).$$

Since $B \wedge F_3' \models E$ and $B \wedge F_3'$ is consistent, F_3' is also a hypothesis wrt B, E and $\mathcal{P}$.

3 Ordering Deductive Operators

Our motivation in this section is to develop a simplified generalisation procedure for CF-induction that uses fewer operators while preserving its soundness and completeness for finding hypotheses. We present one way to simplify the generalisation process by computing generalisations deductively. Our approach is motivated by considering the following deductive operators, which are based on [17].

Definition 3 (Deductive operators). *Let S and T be clausal theories. Then T is directly-derivable from S iff T is obtained from S by one of the following three operators:*

1. *(resolution) $T = S \cup \{C\}$, where C is a resolvent of two clauses $D_1, D_2 \in S$.*
2. *(subsumption) $T = S \cup \{C\}$, where C is subsumed by some clause $D \in S$.*
3. *(weakening) $T = S - \{D\}$ for some clause $D \in S$.*

Two special cases of the subsumption operator can be further distinguished by the following two operators.

2a (instantiation) $T = S \cup \{D\sigma\}$ for some clause $D \in S$ and substitution σ.
2b (expansion) $T = S \cup \{C\}$, where C is a superset of some clause $D \in S$.

We write $S \vdash_r T$, $S \vdash_s T$, $S \vdash_w T$, $S \vdash_\alpha T$ and $S \vdash_\beta T$ to denote that T is directly derivable from S by resolution, subsumption, weakening, instantiation and expansion, respectively. $\vdash_X^*$ is the reflexive and transitive closure of $\vdash_X$, where X is one of the symbols r, s, w, α, β.

We now show that entailment between theories can be established by applying these operators in a particular order.

Theorem 3. *Let S be a clausal theory and T be a clausal theory without any tautological clauses. If $S \models T$, then there are two clausal theories U, V such that*

$$S \vdash_r^* U \vdash_s^* V \vdash_w^* T$$

Proof. Let $T = \{C_1, \ldots, C_n\}$. Then $S \models C_i$ for each clause C_i in T. By the Subsumption Theorem there is a derivation $R_1^i, \ldots, R_{m_i}^i$ from S of a clause $R_{m_i}^i$ that subsumes C_i. Hence, it is sufficient to let $U = S \cup \{R_j^i : 1 \leq i \leq n, 1 \leq j \leq m_i\}$ and $V = U \cup T$. □

We also mention the following two properties concerned with reordering of deductive operators.

Proposition 1. *Let U_1 and U_2 be clausal theories. If $U_1 \vdash_s^* U_2$, then there exist two clausal theories V_1 and V_2 such that $U_1 \vdash_\alpha^* V_1 \vdash_\beta^* V_2 \vdash_w^* U_2$.*

Proof. By Definition 3, there exist a substitution σ_i and a clause $D_i \in U_1$ such that $D_i\sigma_i \subseteq C_i$, for each clause $C_i \in U_2 - U_1$ $(1 \leq i \leq n)$. Let T be a finite set $\bigcup_{i=1}^n \{D_i\sigma_i\}$. Let V_1 be $U_1 \cup T$ and V_2 be $V_1 \cup \{C_1, \ldots, C_n\}$. Then it holds that $U_1 \vdash_\alpha^* V_1$ and $V_1 \vdash_\beta^* V_2$. Moreover, since it holds that $U_2 \subseteq V_2$, it holds that $V_2 \vdash_w^* U_2$. □

Proposition 2. *Let S, U and V be clausal theories. If $S \vdash_w^* U \vdash_s^* V$, then there exists a clausal theory U' such that $S \vdash_s^* U' \vdash_w^* V$.*

Proof. Let U' be the clausal theory $S \cup (V - U)$. Since $U \vdash_s^* V$, there exists a clause $D_i \in U$ such that $D_i \succeq C_i$ for each clause $C_i \in V - U$. Since $S \vdash_w^* U$, $U \subseteq S$ holds. Since $D_i \in U$, $D_i \in S$ holds. Therefore U' is obtained from S using the subsumption operator, that is, $S \vdash_s^* U'$. Next we show that $V \subseteq U'$ holds. Since $U' = S \cup (V - U)$, $V - U \subseteq U'$ holds. And also, $S \subseteq U'$ holds. Since $S \vdash_w^* U$, $U \subseteq S$ holds. Accordingly, $V \cap U \subseteq U'$ holds, since $V \cap U \subseteq U$. Since $V - U \subseteq U'$ and $V \cap U \subseteq U'$, $V \subseteq U'$ holds. Hence $U' \vdash_w^* V$ holds. □

4 Logical Reconstruction of CF-Induction

In this section, we use the above ordering results to show how the number of generalisation operators used in CF-induction can be reduced. Section 4.1 shows this result in the case of ground hypotheses and Section 4.2 shows it in the general case.

4.1 Logical Relation between Bridge Formulae and Ground Hypotheses

First, we show that resolution and instantiation can be incorporated into the selection of CC. We show this with the following two lemmas.

Lemma 1. *Let CC be a bridge formula wrt B, E and $\mathcal{P}$, and U be a clausal theory. If $CC \vdash_r^* U$, then there exist a clausal theory V and a bridge formula CC' wrt B, E and $\mathcal{P}$ such that $CC' \vdash_s^* V \vdash_w^* U$.*

Proof. The proof is in two parts.

(a) First we prove that, for each clause $C_i \in U$ $(1 \leq i \leq n)$, there exists a clause $D_i \in Carc(B \wedge \overline{E}, \mathcal{P})$ such that $D_i \succeq C_i$.

By the definition of characteristic clauses, for each clause $K \in Th_{\mathcal{P}}(B \wedge \overline{E})$ there exists a clause $M \in Carc(B \wedge \overline{E}, \mathcal{P})$ such that $M \succeq K$. Hence it is sufficient to show that every clause $C_i \in U$ is included in $Th_{\mathcal{P}}(B \wedge \overline{E})$. Since $CC \models U$ and $B \wedge \overline{E} \models CC$, it holds that every clause $C_i \in U$ is a consequence of $B \wedge \overline{E}$. Then it remains to show that every clause $C_i \in U$ belongs to $\mathcal{P}$, which is done by mathematical induction on the number n of the applications of $\vdash_r$ for deriving U from CC. In the following, we write $CC \vdash_r^n U$ to denote that U is derived from CC by n applications of $\vdash_r$.

Base step: If $n = 0$ then $U = CC$ and it trivially follows that every clause in U belongs to $\mathcal{P}$.

Induction step: If $n = k + 1$ for some $k \geq 0$, then it holds that $CC \vdash_r^k U' \vdash_r U$ where U' is a clausal theory. By the induction hypothesis, it holds that every clause in U' belongs to $\mathcal{P}$. Moreover, it follows that $U = U' \cup \{R\}$ for some resolvent R of two clauses in U'. Since every clause in U' belongs to $\mathcal{P}$ and $\mathcal{P}$ is closed under instantiation, the resolvent R also belongs to $\mathcal{P}$. Thus every clause in U belongs to $\mathcal{P}$ and so part (a) holds.

(b) Now we show how to construct the theories CC' and V. Start by defining the theory $T = \bigcup_{i=1}^{n} \{D_i\}$ using the clauses $D_i \in Carc(B \wedge \overline{E}, \mathcal{P})$ constructed above. Now define the bridge formula $CC' = CC \cup T$ and the theory $V = CC' \cup U$. Since for each clause $C_i \in U$ there exists $D_i \in T$ such that $D_i \succeq C_i$, $CC' \vdash_s^* V$ holds. Since $U \subseteq V$, $V \vdash_w^* U$ holds. Hence $CC' \vdash_s^* V \vdash_w^* U$ holds. Each clause in CC' is an instance of a clause in $Carc(B \wedge \overline{E}, \mathcal{P})$. Since $CC \subseteq CC'$ and CC is a bridge formula, there exists a clause in CC' which is an instance of a clause in $NewCarc(B, \overline{E}, \mathcal{P})$. Then CC' is a bridge formula. Therefore there exist a bridge formula CC' and a clausal theory V such that $CC' \vdash_s^* V \vdash_w^* U$. □

Lemma 2. *Let CC be a bridge formula wrt B, E and $\mathcal{P}$, and U be a clausal theory. If $CC \vdash_\alpha^* U$, then U is a bridge formula wrt B, E and $\mathcal{P}$.*

Proof. Every clause in U is an instance of a clause in $Carc(B \wedge \overline{E}, \mathcal{P})$. Since $U \supseteq CC$, there exists a clause $C_i \in U$ such that C_i is an instance of a clause from $NewCarc(B, \overline{E}, \mathcal{P})$. Therefore, U is a bridge formula. □

Then, using Lemmas 1 and 2, we can show the following theorem, which establishes the logical relation between bridge formulae and ground hypotheses.

Theorem 4. *Let H be a ground hypothesis wrt B, E and $\mathcal{P}$. Then there exist a bridge formula CC wrt B, E and $\mathcal{P}$ and a clausal theory V such that $CC \vdash_\beta^* V \vdash_w^* \overline{H}$.*

Proof. First, we consider the case that $\overline{H}$ has no tautological clauses. Then, by Theorem 2, there exists a bridge formula CC such that $CC \models \neg H$. Since H is ground, $\overline{H} \equiv \neg H$ holds. Thus $CC \models \overline{H}$ holds. By Theorem 3, there exist clausal theories V_1 and V_2 such that $CC \vdash_r^* V_1 \vdash_s^* V_2 \vdash_w^* \overline{H}$. By Lemma 1, there exist a clausal theory V_3 and a bridge formula CC'' such that $CC'' \vdash_s^* V_3 \vdash_w^* V_1 \vdash_s^* V_2 \vdash_w^* \overline{H}$. By Proposition 2, there exists a clausal theory V_4 such that

$CC'' \vdash_s^* V_3 \vdash_s^* V_4 \vdash_w^* V_2 \vdash_w^* \overline{H}$. Thus $CC'' \vdash_s^* V_4 \vdash_w^* \overline{H}$. By Proposition 1, there exist clausal theories V_5 and V_6 such that $CC'' \vdash_\alpha^* V_5 \vdash_\beta^* V_6 \vdash_w^* V_4 \vdash_w^* \overline{H}$. By Lemma 2, V_5 is a bridge formula, and by letting $CC' = V_5$, it holds that $CC' \vdash_\beta^* V_6 \vdash_w^* \overline{H}$.

Next, in the case that $\overline{H}$ contains tautological clauses, we let $T = \{D_1, \ldots, D_n\}$ denote the set of the tautological clauses in $\overline{H}$. Then, since every clause $D_i \in T$ $(1 \leq i \leq n)$ is a consequence of $B \wedge \overline{E}$ and D_i belongs to $\mathcal{P}$, for each $D_i \in T$ there exists a clause $C_i \in Carc(B \wedge \overline{E}, \mathcal{P})$ such that $C_i \succeq D_i$. Let S be the clausal theory $\bigcup_{i=1}^n \{C_i\}$. Then, it holds that $S \vdash_s^* S_1 \vdash_w^* T$ where $S_1 = S \cup T$. Now, since the clausal theory $\overline{H} - T$ has no tautological clauses, there exist a bridge formula CC and a clausal theory V_7 such that $CC \vdash_s^* V_7 \vdash_w^* \overline{H} - T$. Then, it holds that $CC \cup S \vdash_s^* V_7 \cup S_1 \vdash_w^* \overline{H}$. Since the clausal theory $CC \cup S$ satisfies Definition 2, the theorem holds. □

Next, we introduce a new operator, which can be regarded as concatenating weakening and expansion.

Definition 4 (γ-operator). *Let S and T be clausal theories. Then T is directly γ-derivable from S iff T is obtained from S under the following condition:*

$T = (S - \{D\}) \cup \{C_1, \ldots, C_n\}$ *for some* $n \geq 0$ *where* $C_i \supseteq D$ *for all* $1 \leq i \leq n$.

Analogously to Definition 3, we write $S \vdash_\gamma T$ iff T is directly γ-derivable from S and $\vdash_\gamma^*$ is a reflexive and transitive closure of $\vdash_\gamma$.

Theorem 5. *Let H be a ground hypothesis wrt B, E and $\mathcal{P}$. Then, there exists a bridge formula CC wrt B, E and $\mathcal{P}$ such that $CC \vdash_\gamma^* \overline{H}$.*

Proof. Since H is a ground hypothesis, by Theorem 4, there exist a bridge formula $CC = \{C_1, \ldots, C_n\}$ and a clausal theory U such that $CC \vdash_\beta^* U \vdash_w^* \overline{H}$. Let F_{C_i} be the clausal theory $\{C \mid C \in \overline{H} \text{ and } C_i \subseteq C\}$, for each clause $C_i \in CC$ $(1 \leq i \leq n)$. Then, by Definition 4, for each clause $C_i \in CC$, $\{C_i\} \vdash_\gamma^* F_{C_i}$ holds. Accordingly, $CC \vdash_\gamma^* \bigcup_{i=1}^n F_{C_i}$ holds.

Hence, it is sufficient to show that $\overline{H} = \bigcup_{i=1}^n F_{C_i}$. Since $F_{C_i} \subseteq \overline{H}$ for every clause $C_i \in CC$, it holds that $\bigcup_{i=1}^n F_{C_i} \subseteq \overline{H}$. Conversely, since $CC \vdash_\beta^* U$, for every clause $D \in U$, there exists a clause $C_i \in CC$ such that $C_i \subseteq D$. Also since $U \vdash_w^* \overline{H}$, it holds that $\overline{H} \subseteq U$. Then, for every clause $D \in \overline{H}$, there exists a clause $C_i \in CC$ such that $C_i \subseteq D$, that is, $D \in F_{C_i}$. This means that $\overline{H} \subseteq \bigcup_{i=1}^n F_{C_i}$. Hence, it holds that $\overline{H} = \bigcup_{i=1}^n F_{C_i}$. Therefore, it holds that $CC \vdash_\gamma^* \overline{H}$. □

4.2 Deriving Non-ground Hypotheses

We generalise the result of the previous section to non-ground hypotheses. We show that any hypothesis can be obtained from a bridge formula by applying the γ-operator followed by anti-instantiation.

Definition 5. *Let B and E be clausal theories and $\mathcal{P} = \langle \mathbf{L} \rangle$ be a production field. A clausal theory H is derived by CF-induction with γ-operator from B, E and $\mathcal{P}$ iff H is constructed as follows:*

Step 1. *Construct a bridge formula CC wrt B, E and $\mathcal{P}$.*
Step 2. *Construct a clausal theory G such that $CC \vdash_{\gamma}^{*} G$.*
Step 3. *Compute the complement $\overline{G}$ of G.*
Step 4. *H is obtained by applying anti-instantiation to $\overline{G}$, such that $B \wedge H$ is consistent, H contains no Skolem constants, and for every literal L in H, $\neg L$ belongs to $\mathcal{P}$.*

Several remarks are necessary for Definition 5.

1. Even if G satisfies $CC \vdash_{\gamma}^{*} G$ for some bridge formula CC at Step 2, any output H obtained from G cannot satisfy the conditions of Definition 1 unless $\overline{G} \wedge B$ is consistent and G belongs to $\mathcal{P}$. In this respect, the constraints of H at Step 4 are introduced to guarantee the soundness of H.
2. At Step 3, the complement of G can theoretically include redundant clauses such as tautological clauses and clauses properly subsumed by other clauses. Accordingly it might be necessary to remove such clauses from the complement of G computed in Step 3[1].
3. At Step 4, anti-instantiation allows us to replace subterms in $\overline{G}$ with variables. For example, for the clause $p(a) \vee q(a)$, it is possible to construct $p(X) \vee q(Y)$ obtained by replacing the constant a in $p(a)$ and $q(a)$ with two variables X and Y, respectively. In this way there are many possibilities to apply anti-instantiation for clauses.

We now give soundness and completeness results for CF-induction with γ-operator.

Theorem 6 (Soundness). *Let B, E and H be clausal theories, and $\mathcal{P}$ be a production field. If H is derived by CF-induction with γ-operator from B, E and $\mathcal{P}$, then H is a hypothesis wrt B, E and $\mathcal{P}$.*

Proof. Suppose a clausal theory H is derived by CF-induction with γ-operator from B, E and $\mathcal{P}$. Then, by Definition 5, there exist a bridge formula CC and a clausal theory G such that H is derived by applying anti-instantiation to $\overline{G}$ and $CC \vdash_{\gamma}^{*} G$. By Definition 2, it holds that $B \wedge \overline{E} \models CC$. By Definition 4, it holds that $CC \models G$. Accordingly, it holds that $B \wedge \overline{E} \models G$. Since H is derived by applying anti-instantiation to $\overline{G}$, $H \models \overline{G}$ holds. Since $\overline{G} \models \neg G$, $H \models \neg G$ follows. Equivalently, $G \models \neg H$. Therefore, it holds that $B \wedge \overline{E} \models \neg H$. Then, it holds that $B \wedge H \models \neg \overline{E}$. Since, from Step 4 of Definition 5, H contains no Skolem constants from $\overline{E}$, it holds that $B \wedge H \models E$. Hence, it holds that H is a hypothesis wrt B, E and $\mathcal{P}$, since, from Step 4 of Definition 5, $B \wedge H$ is consistent and for every literal L appearing in H, $\neg L \in \mathbf{L}$. □

[1] We refer the readers to [16] concerning an efficient algorithm for computing such clauses obtained by removing redundant clauses from the complement.

Theorem 7 (Completeness). *Let B, E and H be clausal theories, and $\mathcal{P}$ be a production field. If H is a hypothesis wrt B, E and $\mathcal{P}$, then there exists a theory $H^* \equiv H$ that is derived by CF-induction with γ-operator from B, E and $\mathcal{P}$.*

Proof. Suppose H is a hypothesis wrt B, E and $\mathcal{P}$. By Theorem 2, there is a bridge formula CC wrt B, E and $\mathcal{P}$ such that $CC \cup H$ is unsatisfiable. Using Herbrand's theorem[2] [12], there are two finite sets CC' and H' such that CC' (resp. H') is a finite set of ground instances of CC (resp. H) and $CC' \cup H'$ is unsatisfiable. In this case, H' can be chosen in such a way that for every clause C in H, there is an instance C' of C such that $C' \in H'$, and also, CC' can be chosen in such a way that CC' contains at least one instance of a clause in $NewCarc(B, \overline{E}, \mathcal{P})$. Then, H can be obtained by applying an anti-instantiation generaliser to H'. We prove that H' is a ground hypothesis wrt B, $\overline{\overline{E}}$, and $\mathcal{P}$. That is, we will show that (1) $B \wedge H' \models \overline{\overline{E}}$, (2) $B \wedge H' \not\models \Box$ and (3) $\neg L \in \mathbf{L}$ for every literal L appearing in H'.
Proof of (1): CC is a bridge formula wrt B, E and $\mathcal{P}$. Since every clause in CC' is an instance of a clause in CC, CC' satisfies the first condition of Definition 2. Also, CC' contains at least one clause C' such that C' is an instance of a clause from $NewCarc(B, \overline{E}, \mathcal{P})$. Then CC' satisfies the second condition of Definition 2. Hence, CC' is also bridge formula wrt B, E and $\mathcal{P}$. Thus $B \wedge \overline{E} \models CC'$ holds. By $CC' \models \neg H'$, $B \wedge \overline{E} \models \neg H'$ holds. Then $B \wedge H' \models \neg \overline{E}$ holds. By $\neg \overline{E} \equiv \overline{\overline{E}}$, $B \wedge H \models \overline{\overline{E}}$ holds.
Proof of (2): If it holds that $B \wedge H' \models \Box$, then it must hold that $B \wedge H \models \Box$, by $B \wedge H \models B \wedge H'$. It contradicts the fact H is a hypothesis.
Proof of (3): Since H is a hypothesis wrt B, E and $\mathcal{P}$, for every literal L appearing in H, $\neg L \in \mathbf{L}$ holds. Then it holds that $\neg L \in \mathbf{L}$ for every literal L appearing in H', since $\mathcal{P} = \langle \mathbf{L} \rangle$ is closed under instantiation.

Now, since H' is a ground hypothesis wrt B, $\overline{\overline{E}}$ and $\mathcal{P}$, there exists a bridge formula CC'' wrt B, $\overline{\overline{E}}$ and $\mathcal{P}$ such that $CC'' \vdash_{\gamma}^{*} \overline{H'}$ by Theorem 5. Since $\overline{\overline{\overline{E}}} \equiv \overline{E}$, it holds that for every clause C in $Carc(B \wedge \overline{\overline{\overline{E}}}, \mathcal{P})$, C is contained in $Carc(B \wedge \overline{E}, \mathcal{P})$. Then, it also holds that for every clause C in CC'', C is contained in $Carc(B \wedge \overline{E}, \mathcal{P})$. Hence, CC'' is also bridge formula wrt B, E and $\mathcal{P}$. Then, from Step 1 of Definition 5, CC'' can be constructed in a CF-induction with γ-operator from B, E and $\mathcal{P}$. Moreover, $\overline{H'}$ can be also constructed from CC'' at Step 2. Since H' is ground, it holds that H' is logically equivalent to $\overline{\overline{H'}}$ computed from $\overline{H'}$ at Step 3. Recall that H is obtained by applying anti-instantiation to H'. Therefore a formula H^* is obtained at Step 4 with the application of anti-instantiation to $\overline{\overline{H'}}$ such that $H^* \equiv H$. $\Box$

Example 2. Recall Example 1. Let CC be the following bridge formula, which appears in Example 1.

$$CC = (natural(0) \vee even(0)) \wedge \neg natural(s(0))$$

[2] A set of clauses Σ is unsatisfiable iff a finite set of ground instances of clauses of Σ is unsatisfiable.

Assume that a γ-operator is applied to CC so that $\neg natural(s(0))$ is replaced with the two clauses $\neg natural(s(0)) \vee \neg even(0)$ and $\neg natural(s(0)) \vee natural(0)$, then the following clausal theory G_1 is constructed:

$$\begin{aligned} G_1 = &(natural(0) \vee even(0)) \\ &\wedge (\neg natural(s(0)) \vee \neg even(0)) \\ &\wedge (\neg natural(s(0)) \vee natural(0)). \end{aligned}$$

Then, we can obtain the complement $\overline{G_1}$ of G_1, which is logically equivalent to F'_1 in Example 1. Next assume that another γ-operator is applied to CC so that $\neg natural(s(0))$ is replaced with the two clauses $\neg natural(s(0)) \vee even(0)$ and $\neg natural(s(0)) \vee \neg natural(0)$, then the following clausal theory G_2 is constructed:

$$\begin{aligned} G_2 = &(natural(0) \vee even(0)) \\ &\wedge (\neg natural(s(0)) \vee even(0)) \\ &\wedge (\neg natural(s(0)) \vee \neg natural(0)) \end{aligned}$$

Then, the complement $\overline{G_2}$ of G_2 is logically equivalent to F'_2 in Example 1. Accordingly, we can obtain a clausal theory, which is logically equivalent to F'_3 in Example 1 by applying an anti-instantiation generaliser to $\overline{G_2}$. In this way, the inverse resolution generaliser can be realised with applications of the γ-operator.

5 Related Work

The γ-operator can be regarded as a particular *downward refinement operator* [1,6,8,12] for the $\vdash_{\gamma}^{*}$ order, which is closely related to the subsumption order. Let S and T be clausal theories such that $S \vdash_{\gamma}^{*} T$. Then $S \succeq T$ holds. Compared with the subsumption order, one important feature of γ-operator lies in restraint of the operation of instantiation, which leads to a large number of choice points. There are certain desirable properties that a “good” downward refinement operator should satisfy and we intend to study which of these the γ-operator satisfies.

We can reduce generalisation under the entailment relation in the previous version to generalisation under the γ-operator. It is based on the notion that any series of processes of inductive operations on the inverse relation of entailment between the negation of a bridge formula and a hypothesis connects a certain series of processes of deductive operations on entailment between a bridge formula and the negation of hypothesis. Accordingly, there are two sides where we can grasp generalisation processes. Yamamoto and Fronhöefer [18] and Yamamoto [17] first have studied the connection between two clausal theories related by entailment and negation. It will be interesting to consider about the relation between two clausal theories ordered by the γ-operator instead of entailment and these negations.

CF-induction is related to several methods [9,14,15,18] developed in the field of Inductive Logic Programming (ILP) [11], and the relationship between CF-induction and such methods as Progol [9] and the residue procedure [18] has been studied in [5]. We briefly discuss the relationship between CF-induction and the more recent HAIL/FC-HAIL [14,15] approaches.

HAIL [14] uses a combination of abduction and induction to overcome some of the limitations of Progol. In particular, HAIL learns more than one clause in response to a single example using a multi-clause generalisation of the *Bottom Clause* called a *Kernel Set*, which can be regarded as a bridge formula. The HAIL generaliser is anti-subsumption applied to this set of clauses. Recently, a full clausal generalisation of HAIL, called FC-HAIL, has been proposed in [15] that extends the mode-directed algorithm of HAIL to full clausal theories. Although FC-HAIL is not complete for hypothesis finding in full clausal logic unlike CF-induction, it can use a form of language bias called mode-declarations to restrict the hypothesis space like Progol. It will be fruitful to consider about incorporating the notion of mode-declarations into CF-induction for finding efficient hypotheses.

6 Conclusion and Future Work

We have studied how the generalisation procedure of CF-induction can be simplified while preserving its soundness and completeness. In Section 4.1, we introduced the γ-operator whose task is removing some clause D in an input clausal theory and adding a set of clauses $C_1, \ldots, C_n$ for some $0 \leq n$ where each clause C_i is a super set of D. In Section 4.2, we showed that the γ-operator and ant-instantiation are sufficient to ensure the completeness of CF-induction.

The results shown in this paper will contribute toward promoting practical real-world applications of CF-induction. CF-induction is now expected to be applied in real domain such as Systems Biology [3], and we are trying to apply CF-induction with the γ-operator to examples shown in [3]. Some initial experiments have confirmed that these examples can be more easily solved using CF-induction with γ-operator, as compared with the previous version.

Future work will investigate ways of automatically finding which literals must be added to selected clauses by the γ-operator. In addition, we believe that studying various restrictions of the γ-operator may allow us to systematically compare the generalisation power of previously proposed operators. Other important future work is the selection of clauses in CC, which currently requires assistance from the user. We intend to address this issue by ranking characteristic clauses in such a way that the user is allowed to more easily select appropriate clauses for CC.

Acknowledgements. The authors would like to thank the anonymous reviewers for giving us useful and constructive comments.

References

1. Badea, L.: A Refinement Operator for Theories. In: Rouveirol, C., Sebag, M. (eds.) ILP 2001. LNCS (LNAI), vol. 2157, pp. 1–14. Springer, Heidelberg (2001)
2. Cox, P.T., Pietrzykowski, T.: A Complete, Nonredundant Algorithm for Reversed Skolemization. Theoretical Computer Science 28(3), 239–261 (1984)

3. Doncescu, A., Inoue, K., Yamamoto, Y.: Knowledge Based Discovery in Systems Biology Using CF-Induction. In: Okuno, H.G., Ali, M. (eds.) IEA/AIE 2007. LNCS (LNAI), vol. 4570, pp. 395–404. Springer, Heidelberg (2007)
4. Inoue, K.: Linear Resolution for Consequence Finding. Artificial Intelligence 56(2-3), 301–353 (1992)
5. Inoue, K.: Induction as Consequence Finding. Machine Learning 55(2), 109–135 (2004)
6. Laird, P.D.: Learning from Good and Bad Data. Kluwer Academic Publishers, Dordrecht (1988)
7. Lee, C.T.: A Completeness Theorem and Computer Program for Finding Theorems Derivable from Given Axioms. PhD thesis, Dept. of Electronic Eng. and Computer Sci., Univ. of California, Berkeley, CA (1967)
8. Midelfart, H.: A Bounded Search Space of Clausal Theories. In: Džeroski, S., Flach, P.A. (eds.) ILP 1999. LNCS (LNAI), vol. 1634, pp. 210–221. Springer, Heidelberg (1999)
9. Muggleton, S.H.: Inverse Entailment and Progol. New Generation Computing 13(3-4), 245–286 (1995)
10. Muggleton, S.H., Buntine, W.L.: Machine Invention of First Order Predicates by Inverting Resolution. In: Proc. of the 5th Int. Conf. on Machine Learning, pp. 339–352 (1988)
11. Muggleton, S.H., De Raedt, L.: Inductive Logic Programming: Theory and Methods. Logic Programming 19(20), 629–679 (1994)
12. Nienhuys-Cheng, S., de Wolf, R.: Foundations of Inductive Logic Programming. LNCS, vol. 1228. Springer, Heidelberg (1997)
13. Plotkin, G.D.: A Further Note on Inductive Generalization. In: Machine Intelligence, vol. 6, pp. 101–124. Edinburgh University Press (1971)
14. Ray, O., Broda, K., Russo, A.M.: Hybrid Abductive Inductive Learning: a Generalisation of Progol. In: Horváth, T., Yamamoto, A. (eds.) ILP 2003. LNCS (LNAI), vol. 2835, pp. 311–328. Springer, Heidelberg (2003)
15. Ray, O., Inoue, K.: Mode Directed Inverse Entailment for Full Clausal Theories. In: Proc. of the 17th Int. Conf. on Inductive Logic Programming ILP 2007. LNCS, vol. 4894. Springer, Heidelberg (to appear 2008)
16. Satoh, K., Uno, T.: Enumerating Maximal Frequent Sets Using Irredundant Dualization. In: Grieser, G., Tanaka, Y., Yamamoto, A. (eds.) DS 2003. LNCS (LNAI), vol. 2843, pp. 256–268. Springer, Heidelberg (2003)
17. Yamamoto, A.: Hypothesis Finding Based on Upward Refinement of Residue Hypotheses. Theoretical Computer Science 298, 5–19 (2003)
18. Yamamoto, A., Fronhöfer, B.: Hypotheses Finding via Residue Hypotheses with the Resolution Principle. In: Arimura, H., Sharma, A.K., Jain, S. (eds.) ALT 2000. LNCS (LNAI), vol. 1968, pp. 156–165. Springer, Heidelberg (2000)

Part V

Juris-Informatics

First International Workshop on Juris-Informatics

Katsumi Nitta, Ken Satoh, and Satoshi Tojo

The First International Workshop on Juris-Informatics (ie: JURISIN workshop) was held on June 19, 2007 at World Convention Center Summit in Miyazaki, Japan, as a part of the Twenty First Annual Conference of Japanese Society for Artificial Intelligence (JSAI-2007). This workshop was organized to study legal issues from the perspective of informatics. Law is one of the oldest practical applications of computer science. Though lots of legal reasoning systems have been developed thus far, they were not supported by the lawyers, or they didn't have a positive impact on juris-prudence. One of the reasons is that legal reasoning mechanisms currently implemented are too simple from the lawyer's viewpoint. Another reason is that legal reasoning has been studied mainly from the viewpoint of logical aspects, but it has not been studied so much from the view point of natural language processing. If we can bring lawyers and informatics people and natural language processing people together, we can expect great advances in both informatics and jurisprudence by implementing legal reasoning systems clear to what lawyers expect.

The main purpose of the JURISIN workshop is to discuss both the fundamental and practical issues in juris-informatics among people from diverse backgrounds such as law, social science, information and intelligent technology, logic and philosophy, including the conventional "AI and law" area. The program committee (PC) was organized with the help of leading researchers in AI and Law as follows; Kevin Ashley (Univ. Pittsburgh, USA), Trevor Bench-Capon (Univ. Liverpool, UK), Aditya Ghose (Univ. Wollongong, Australia), Guido Governatori (Univ. Queenland, Australia), Tokuyasu Kakuta (Univ. Nagoya, Japan), Ronald Loui (Univ. Washington, USA), Henry Prakken (Univ. Utrecht & Griningen, The Netherland), Seiichiro Sakurai (Meijigakuin Univ., Japan) Giovanni Sartor (European Univ. Institute, Italy), Hajime Sawamura (Univ. Niigata, Japan), Akira Shimazu (JAIST, Japan), Katsuhiko Toyama (Univ. Nagoya, Japan), Takahira Yamaguchi (Keio Univ., Japan), and John Zeleznikow (Victoria Univ., Australia).

Though the announcement period was short, eleven papers were submitted. Each paper was reviewed by three PC members, and eight papers were accepted in total. The accepted eight papers covers various topics such as legal reasoning, argumentation theory, legal ontology, computer-aided law education, use of informatics and AI in law, and so on. The workshop became a provoking and stimulating opportunity for opening up new research areas.

After the workshop, five papers were submitted for the post proceedings. They were reviewed by PC members again and four papers were selected. Followings are their synopses.

Nakamura et al. investigate linguistic characteristics of legal texts, and propose a framework for translating legal sentences into logical forms. They examine the translating system with actual data of legal documents and show this system provides high accuracy in terms of predicates corresponding to words and their semantic relations.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 347–348, 2007.

Ogawa et al. focus on the fact that some statutes are not easy to manage because they are often amended and their every version must be restored in the legal database. They show that amendment clauses can be formalized in terms of sixteen kinds of regular expressions, and propose an automatic consolidation system for Japanese statutes based in the formalization and experts' knowledge about consolidation.

Tanaka et al. show an online mediation support system and an argument agent which participates in moot mediation as a disputant for self-training purposes. An agent's text responses are generated by retrieving from a case base. The characteristic of an argument agent is controlled by three parameters such as single-mindedness, selfishness and argumentativeness.

Toni provides a mapping of defeasible reasoning into assumption-based argumentation, and shows that the framework obtained has properties of closedness and consistency which are important for defeasible reasoning in the presence of strict rules. Though other argumentation approaches have been proven closed and consistent under some semantics, assumption-based argumentation is closed and consistent under all argumentation semantics.

Finally, we wish to express our gratitude to all those who submitted papers, PC members, discussant and attentive audience.

October 2007

Katsumi Nitta(Tokyo Tech)
Ken Satoh(NII)
Satoshi Tojo(JAIST)

Towards Translation of Legal Sentences into Logical Forms

Makoto Nakamura, Shunsuke Nobuoka, and Akira Shimazu

School of Information Science
Japan Advanced Institute of Science and Technology
1-1, Asahidai, Nomi, Ishikawa, 923-1292, Japan
{mnakamur,shimazu}@jaist.ac.jp

Abstract. This paper proposes a framework for translating legal sentences into logical forms in which we can check for inconsistency, and describes the implementation and experiment of the first experimental system. Our logical formalization conforms to Davidsonian Style, which is suitable for languages allowing expressions with zero-pronouns such as Japanese. We examine our system with actual data of legal documents. As a result, the system was 78% of accurate in terms of deriving predicates with bound variables. We discuss our plan for further development of the system from the viewpoint of the following two aspects: (1) improvement of accuracy (2) formalization of output necessary for logical processing.

1 Introduction

In recent years, a new research field called *Legal Engineering* was proposed in order to achieve a trustworthy electronic society [1,2]. Legal Engineering serves to examine and verify whether the following issues are satisfied:

- A law is established appropriately according to its purpose,
- There are no logical contradictions or no problems as a document *per se*,
- The law is consistent with related laws, and
- It is modified, added, and deleted consistently for its revision.

Legal Engineering also serves to design an information system which works based on laws. Towards the achievement of the goal of Legal Engineering, we need to develop a system with advanced technology which deals with electronically processed legal texts in computers. The core of the system roughly consists of two procedures:

1. Translating legal texts into logical forms
2. Proving consistency in terms of the given logical forms [3]

This paper reports our ongoing research effort to develop a system for the translation into forms. Hence, our purpose in this paper is to develop a system which translates legal texts into logical forms in which we can check for inconsistency.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 349–362, 2008.

Acquisition of knowledge bases by automatically reading natural language texts has widely been studied, and is one of the main themes of the field of natural language processing [4]. Because the definition of semantic representation differs depending on what the language processing systems deal with, a few systems try to generate logical formulae based on first order predicate logic [5]. A study of knowledge acquisition by Mulkar et al. [6,7] is one of those systems. They extracted well-defined logical formulae from textbooks of biology and chemistry. Although the final stage of their work was to apply high school AP exam questions to the system in order to measure its ability relative to high school students, it is not yet as robust as our target which checks for inconsistency of a set of logical formulae.

Let us change the viewpoint to AI in law. Logical processing in the legal domain has widely been studied by AI researchers for a long time [8,9]. They have, however, aimed at finding what kind of law can apply to a particular incident, not at proving law *per se*. Furthermore, most of the systems require manual transcription of legal documents and authoritative examples into logical formulae. Therefore, our system would help them as a pre-processor, which automatically processes law texts.

Because law sentences[1] have to describe the details of their intentions precisely, they are usually very long and thus complicated. These long and complicated sentences potentially have ambiguities in syntax, although ambiguous description must never be permitted. This is the primary reason that reading legal texts is more difficult for people than reading other daily-use documents. However, we consider that it is easier to process such characteristic expressions in legal texts with an appropriate method than that of daily-use documents. In this paper, we pay attention to linguistic characteristics in the surface form of legal texts, such as:

If-then Statement. A law sentence can roughly be separated into two parts. The former part is called the law requisite part, and the latter is the law effectuation part. Thus, a sentence is basically described as an *'if-then'* statement [10].

Coordinate Rule. There are some kinds of expressions for conjunction and disjunction, which are used in a different layer.

Conventional Expression. Making a specific case frame dictionary is effective.

Our logical formalization conforms to Davidsonian Style [11,12], in which a relation between events can be represented by a predicate which has more than one event variables. Otherwise, it is necessary to define a higher order predicate form in order to express predicates referring to predicates. This was pointed out in the previous researches. For example, Yoshino [13] proposed his own representation. On this aspect, we use the well-known representation formalism. We expect output of this style to be easily converted into other first order predicate logic forms [3].

[1] Throughout this paper, the word 'sentence' never indicates judicial decision nor logical formula, but linguistic meaning.

In this research, we linguistically investigated legal documents such as sentences of the Income Tax Law (100 articles, 255 clauses, 247 items, among 244 articles), sentences of the National Pension Law (100 articles among 148 articles) in addition to sentences in municipal laws of Toyama Pref. and Chiyoda Ward, Tokyo, Japan (38 articles, 90 clauses). Taking into account the investigation, we propose a framework for generating a logical formula corresponding to an input legal sentence. Based on the framework and investigation, we implement the first version which realize the framework.

In Section 2 we first describe how to deal with law documents with the methodology of natural language processing, based on linguistic analyses. We propose a framework of our system in Section 3. We show the implementation of the first version and its experimental results in Section 4. Finally, we conclude and describe our future work in Section 5.

2 Linguistic Analysis of Legal Documents

In this section, we aim at finding an appropriate method of replacing a law sentence to a logical formula, investigating some linguistic characteristics of legal texts. In Section 2.1, we explain our logical formalization, which is suitable for Japanese legal texts, and then consider some grammatical constraints in the following subsections.

2.1 Davidsonian Formalization

Our logical formalization conforms to Davidsonian Style [11,12], which is a logical formalism focusing on verb meanings, which are interpreted as properties of events or relations between individuals and events. In this style, an event in a proposition is expressed by predicates with an event variable, some of which are of thematic roles of the event. A relation between events can be represented by a predicate which has more than one event variables without defining a higher order predicate form in order to express predicates referring to predicates. Because words or phrases modifying a verb in a target sentence can directly be assigned to a predicate, this style is suitable for languages allowing expressions with zero-pronouns, as seen in Japanese.

Davidson's original motivation for his proposal comes from the phenomenon of adverbial modification [12]. The Davidsonian approach has been extended and modified by Parsons [11]. According to them, events should not be taken to be structured entities at all, but should be taken instead to be primitive entities. This approach includes an assumption that events are not a subclass of propositions. For example, in Parsons's theory, both action predicates like *run* and predicates that do not describe actions, like *is drunk*, express properties of eventualities: the difference between them is that, while the former expresses a property of events, the latter expresses a property of states. Sentences in (1) and (2) are translated respectively as the following formulae (Tense is ignored here):

$$\text{Jones ran:} \quad \exists e(\text{running}(e) \wedge agent(e, j)) \tag{1}$$

$$\text{Jones was drunk:} \quad \exists e(\text{being-drunk}(e) \wedge patient(e, j)) \tag{2}$$

2.2 Structure of Law Sentences

In most cases, a law sentence consists of a law requisite part and a law effectuation part, which designate its legal logical structure [10,14]. Structure of a sentence in terms of these parts is shown in Fig. 1. The law requisite part is further divided into a subject part and a condition part, and the law effectuation part is into an object, detail, and provision part.

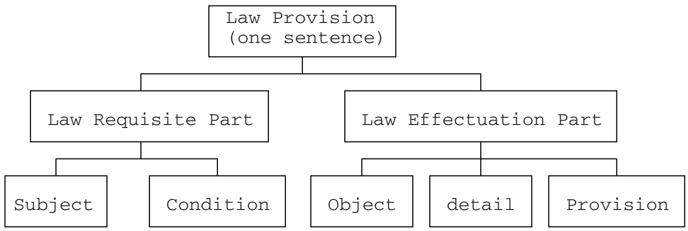


Fig. 1. Structure of requisition and effectuation [10]

Dividing a sentence into these two parts in the pre-processing stage makes the main procedure more efficient and accurate. Nagai et al. [14] proposed an acquisition model of this structure from Japanese law sentences. Dealing with strict linguistic constraints of law sentences, their model succeeded to acquire the structures at fairly high accuracy for a simple method, which specifies surface forms of law sentences. Our approach is different from theirs in that we consider some semantic analyses in order to represent logical formula.

We analyzed plain texts consisting of a total of 501 sentences from 138 provisions of the National Pension Law, and the municipal laws of Toyama Pref. and Chiyoda Ward in Tokyo Pref., and found 84 patterns of cue phrases which represent a combination of the subject part and the condition part as shown in Fig. 1. Example patterns are shown in Table 1, in which the subject and object parts[2] terminate with particular particles, and the condition part with phrases corresponding to 'if' or 'when.' If a sentence matches one of the patterns, each clause in the sentence can be assigned to the subject part or the condition part in the law requisite part, and the rest to the law effectuation part. The object part includes phrases with a dative case particle in addition to ones for the subject part. For the law effectuation part, the provision part denotes the predicate in the main clause of a sentence, which is the last phrases including the main verb in the sentence, and the remaining phrase(s) between the object part and the provision part is assigned to the detail part.

Let us consider how to associate the structure with the logical framework. In general, the law requisite part and the law effectuation part are related to

[2] For the object part, we analyzed 38 provisions of the municipal laws of Toyama Pref. and Chiyoda Ward in Tokyo Pref.

Table 1. Patterns for Subject, Object and Condition Parts

Subject/Object	Condition
'*...wa,*' (〜は、)[Theme]	'*...suru-toki-wa,*' (〜するときは、)
'*...ga,*' (〜が、)[Nominative Case]	'*...ni-tsuite-wa,*' (〜については、)
'*...mo,*' (〜も、)[too/also]	'*...ni-oite-wa,*' (〜においては、)
Object	'*...suru-baai-niwa,*' (〜する場合には、)
'*...ni*' (〜に、) [Dative Case]	'*...niyori,*' (〜により、)

the logical implication (→) or the logical equivalence (↔)[3]. In order to classify phrases of a sentence into the detailed five parts in the Fig. 1, we separate a sentence into clauses based on the cue phrases. After that, we assign the sentence to a logical frame according to its structure, shown in the following items:

- The law requisite part consists of a subject part and a condition part:

$$(\text{Subject}) \wedge (\text{Condition}) \rightarrow (\text{Provision})$$

- The law effectuation part consists of an object part and a provision part:

$$(\text{Condition}) \rightarrow (\text{Object}) \wedge (\text{Provision})$$

- There is neither a subject part or an object part:

$$(\text{Condition}) \rightarrow (\text{Provision})$$

The condition part may appear twice, or may not, depending on the clauses.

According to our investigation about the first 60 sentences described in the National Pension Law, we found that 32 of the sentences included the implication and 20 sentences for the equivalence. Therefore, most of the sentences have the structure consisting of both the law requisite and effectuation parts. We found cue phrases from 30 out of 52 sentences including the implication or equivalence.

Modal Operator. In Japanese law sentences, there exist many kinds of expressions, where the difference of the meaning among them is sometimes delicate. Particularly, such delicate expressions appear at the end of the sentence, which corresponds to auxiliary verbs in English. Therefore, we consider to assign a modal operator to the law effectuation part. The modal operators specify the meaning of the sentence such as obligation, permission, possibility, and inhibition. We generally classify modal linguistic expressions and assign each expression to its corresponding modal operator. Strictly speaking, there are expressions which does not correspond to such a modal operator in some cases. For example, "*monotosuru*" generally corresponds to obligation but does not represent obligation in some cases. In order to have accurate logical representation, we must

[3] We are still examining the cue phrases for distinguishing the logical implication (→) and the logical equivalence (↔), which tends to be used for the definitions of words. In this paper, we focus on the logical implication.

analyze such expressions more and clarify how many kinds of such operators we need to represent Japanese legal sentences and what kind of information we can use to transform expressions into such modal operators.

2.3 Analysis of Noun Phrases – Coordinate Structure

There are strict constraints in terms of coordinate structures which appear very frequently in law sentences [14]. For example, '*matawa*' (又は) and '*moshikuwa*' (若しくは), both of which are equivalent to an English word 'or,' have different precedence in embedding order. '*moshikuwa*' (若しくは) is used in deeper embedding level than '*matawa*' (又は). Therefore, a phrase 'A *moshikuwa* B *matawa* C' should be interpreted as:

$$((A \;\; moshikuwa \;\; B) \;\; matawa \;\; C) \tag{3}$$

We replaced the disjunction markers into logical connectives, that is, '∨,' regardless of the embedding order. Hence, the above expression is translated into the following logical formula:

$$((A \;\vee\; B) \;\vee\; C) \tag{4}$$

There are similar constraints for the other coordinate structure markers, too. The conjunctions, '*oyobi*' (及び) and '*narabini*' (並びに), correspond to an English word 'and,' which is replaced by '∧.'

For a parallel phrase consisting of three or more coordinate noun phrases, the last one that follows a conjunction or disjunction, *e.g.* '*sonota-no*' (その他の) corresponding to "or other," tends to be a hypernym of the precedent noun phrases. An example is shown in the following expression:

機関に係る	申請、	届出	その他の	手続き等
(*kikan-ni-kakaru*	*shinsei,*	*todoke-de*	*sonota-no*	*tetsuzuki-tou*)
concerned with the organization	applications,	notifications	or other	procedures

The precedent words 'applications' and 'notifications' imply the last phrase 'procedures.' The first phrase 'concerned with the organization' should be considered to modify each of the following phrases. We examined the number of distinct expressions of conjunctive phrases from 38 provisions. As a result, we found 5 kinds of conjunctions or disjunctions used in parallel phrases.

Taking the characteristics of expressions into account, we cope with the problem of complexity in the hierarchical coordinate structure.

2.4 Analysis of Noun Phrases – Adnominal Particles 'no'

Japanese has many noun phrase patterns of the type '*A no B*' consisting of two nouns *A* and *B* with an adnominal particle '*no*,' which is interpreted as some relation between *A* and *B*. This type of noun phrase has been widely studied by many researchers. Shimazu et al. [15] classified it into many semantic relations, according to the properties or functions of *A* and *B*. For example, if the noun *B* expresses an action or an event, A is its case element such as agent, object, and so on. In this case, B is typically a *sahen*-noun, which can become a verb with the suffix *-suru*. For example, '*teishutsu-suru*' (submit) is a verb while '*teishutsu*' (submission) is a noun. In the accordance with the previous works, we manually made transformation rules from 430 sentences in 110 provisions. As a result, we classified 'A *no* B' expressions into 10 patterns. We show three examples in the next paragraph.

From the viewpoint of representing the semantics of 'A no B' in logical forms, most of the expressions of 'A no B' consist of predicates corresponding to the words A and B, and to a relation between them as follows:

1. A logical form of typical expressions consists of predicates corresponding to A and B, and to a relation between them. A and/or B is a *sahen*-noun.

   shinseisho$_{(A)}$ *no* *teishutsu*$_{(B)}$ (申請書の提出)
   "submission$_{(B)}$ of an application form$_{(A)}$"

   申請書$(x) \wedge$ 提出$(e) \wedge \mathrm{obj}(e, x)$
   $\mathrm{application\_form}(x) \wedge \mathrm{submit}(e) \wedge \mathrm{obj}(e, x)$

2. In a case as B is an attribute of A, a logical form consists of two predicates corresponding to A and B.

   hi-hokensha$_{(A)}$ *no* *shimei*$_{(B)}$ (被保険者の氏名)
   "the name$_{(B)}$ of the person insured$_{(A)}$"

   被保険者$(x) \wedge = ($氏名$(x), n)$
   $\mathrm{person\_insured}(x) \wedge = (\mathrm{name}(x), n)$

3. In a case that A or B is a compound noun, a logical form of A or B generally consists of more than one predicate.

   hatachi-miman$_{(A)}$ *no* *mono*$_{(B)}$ (二十歳未満の者)
   "a person$_{(B)}$ below the age of twenty$_{(A)}$"

   者$(x) \wedge = ($年齢$(x), t) \wedge < (t, 20)$
   $\mathrm{person}(x) \wedge = (\mathrm{age}(x), t) \wedge < (t, 20)$

3 Proposal of Framework

Here, we explain an outline of our framework, which reflects our linguistic investigation mentioned in Section 2. The following list is the procedure for one sentence. We repeat it during processing a set of sentences.

1. Analyzing morphology and parsing a target sentence.
2. Splitting the sentence based on the characteristic structure of a law sentence.
3. Assignment of modal operators with the cue of auxiliary verbs.
4. Making a paraphrase of some similar expressions to a unified expression.
5. Analyzing clauses and noun phrases using a case frame dictionary.
6. Assigning variables and predicates. We assign verb phrases and *sahen*-nouns to a predicate and an event variable, e_i, and other content words to x_j.

7. Building a logical formula from the fragments of logical connectives, modal operators, and predicates.

The procedure is roughly divided into two parts. One is to make the outside frame of the logical form (Step 1 to 3 and 7), which corresponds to the legal logical structure shown in Fig. 1. The other (Step 4 to 6) is for the inside frame. We assign noun phrases to bound variables and predicates using a case frame dictionary.

4 Implementation and Experiments

4.1 Implementation

In the current system, we use JUMAN [16] and KNP [17], which are a Japanese morphological analyzer and a Japanese dependency analyzer, respectively. Both are representative tools for language processing of Japanese. A Japanese thesaurus [18] is also used for the calculation of similarity among words.

Each sentence is divided depending on the structure consisting of a law requisite part and a law effectuation part. In the current system, there are three types of modal operators, O, M, and P, which correspond to Obligation, Possibility (may), and Permission, respectively. Especially, a sentence expressing a ban on something is represented by the use of the modal operator for Permission with negation ($\neg P$).

In order to assimilate a variety of similar representations into a unified logical form, we make paraphrases of particular expressions. We consider that legal texts are easier to analyze than daily-use documents because of unfamiliar but typical expressions, nevertheless this process is still necessary for stable output.

We describe details about case frame and noun phrase analyses in the next subsections.

4.2 Case Frame Analysis

Using a case frame dictionary, we can search for semantic relations between a verb and modifier nouns in a sentence. We assign semantic relations to predicates which connect to other predicates sharing a common event variable.

We built a case frame dictionary by extracting relations between verbs and their modifier nouns from 818 sentences of 366 provisions in 13 prefectures. As a result, a total of 517 verbs were registered into the dictionary. In the dictionary, each verb is an index and has a number of case frames, each of which stores semantic relations between nouns and the index verb. A case frame consists of a number of case slots, each of which is composed of a deep case, a case particle, a semantic category of nouns, and a set of example nouns and their frequency in use, as:

$$\begin{aligned} V &: \{CF_1, CF_2, \ldots, CF_n\} \\ CF_i &: \{CS_1, CS_2, \ldots, CS_m\} \\ CS_j &: (\text{[deep case]}, \text{[case particle]}, \text{[semantic category of nouns]}, \\ &\quad \{(noun_1, freq_1), (noun_2, freq_2), \ldots\}) \quad , \end{aligned} \tag{5}$$

verb idx : *kakeru* [掛ける]
CF_1 [*meaning "hanging"*]:
((**DC:** agent **CP:** '-*ga* [が]' **SEM_CAT:** 'person [人]'
examples: (('he [彼]' 6)))
(**DC:** object **CP:** '-*wo* [を]' **SEM_CAT:** 'clothing [衣類]'
examples: (('jacket [上着]' 2)('coat [コート]' 4))))
CF_2 [*meaning "calling"*]:
((**DC:** agent **CP:** '-*ga* [が]' **SEM_CAT:** 'person [人]'
examples: (('father [父]' 2)))
(**DC:** destination **CP:** '-*ni* [に]' **SEM_CAT:** 'person [人]'
examples: (('friend [友達]' 1) ('mother [母]' 1)))
(**DC:** object **CP:** '-*wo* [を]' **SEM_CAT:** 'things [物品]'
examples: (('telephone [電話]' 2))))
CF_3 [*meaning "sitting"*]:
((**DC:** agent **CP:** '-*ga* [が]' **SEM_CAT:** 'person [人]'
examples: (('Bob' 1)('mother [母]' 1)))
(**DC:** instrument **CP:** '-ni [に]' **SEM_CAT:** 'things [物品]'
examples: (('chair [椅子]' 2)))
(**DC:** object **CP:** '-wo [を]' **SEM_CAT:** 'body [身体]'
examples: (('waist [腰]' 2))))

DC, **CP**, and **SEM_CAT** denote deep case, case particle, and semantic category, respectively.

Fig. 2. Case Frame and Case Slot

where V, CF_i, and CS_j denote an index of verb phrases, a case frame, and a case slot, respectively. We manually annotated one deep case for each slot. We show in Fig. 2 an example of the dictionary, in which a verb '*kakeru*' has three case frames.

We search for a case frame candidate corresponding to a verb modified by a number of sets of a noun and a case particle. Let us consider which case frame in Fig. 2 the following word '*kakeru*' belongs to: "*Kare-ga watashi-ni denwa-wo kakeru* (彼が私に電話を掛ける)," meaning "He calls me." First of all, we screen out the case frames which do not match the number of case frames. Therefore, CF_1 leaves out of the selection, and CF_2 and CF_3 become the candidates because of the same case particles.

When there are multiple candidates, we choose the one with the highest score by the following calculation method.

1. If a case particle of a noun matches one of the case slots, the current frame scores 5 points, otherwise, 2.5 points either for a sub particle, or for a hidden case used in '*A no B*' and a relative clause.
2. If one of the target nouns matches one of the examples in a case slot, the current frame obtains 5 more points.

3. Using a thesaurus of Japanese [18], we calculate a value of similarity between the head noun of the case element and an example noun stored in the case frame dictionary, and add the value to the score.

Here, we explain the calculation method of 3. Let w_1 be the head noun of the case element, and w_2 be an example noun stored in the case slot. The similarity between the two nouns is calculated as:

$$Sim_w(w_1, w_2) = \frac{2L}{l_1 + l_2} \quad , \tag{6}$$

where l_1 denotes the depth of w_1 from the root node in the thesaurus, l_2 is for w_2, and L denotes the depth of the least upper bound of the category node between w_1 and w_2. If we assume that the case frame dictionary holds n words for a particular case belonging to a predicate, then the similarity between a word, w_1, and the set of n examples, $w_{2,1}, \ldots w_{2,n} \in S$, is calculated in the following equation:

$$Sim_c(w_1, S) = \frac{\sum_{i=1}^{n} Sim_w(w_1, w_{2,i}) \times c_i}{\sum_{i=1}^{n} c_i} \quad , \tag{7}$$

where c_i denotes appearance frequency of $w_{2,i}$.

For the example sentence, the case frame CF_2 gets 45.5 points, while CF_3 gets 28.4. Therefore, our system consequently chooses CF_2 as the appropriate case frame of the sentence. Because the score greatly depends on the example words of the case frame dictionary, it is important to extract the case frames from the large number of actual law texts.

4.3 Noun Phrase Analysis

We especially put our efforts into analysis of noun phrases concerned with relative clauses and '*A no B*' relations. For a relative clause, because a predicate variable of the modificand noun phrase is shared by the events both of the relative clause and of the main verb, we took care in assignment of predicate variables. For example, for a sentence of "He hit the man who sold the book," the man is the agent of an event 'sell the book' as well as the object of the other event 'hit the man.'

For an '*A no B*' relation, because we regard a noun phrase with a *sahen*-noun as a verb, it is transformed to an event. Even though there is only one semantic relation to the event, it is easy to transform the noun phrase into a logical form with an event variable. If we introduced anaphora analysis for the event, we could generate better output, adding predicates in terms of obligatory cases. However, we can make a temporary result without obligatory cases in the current system. This is the reason that Davidsonian formalism is suitable for some languages allowing zero-pronouns.

4.4 Experiments and Results

We examined our system with actual data, which is an ordinance of Hiroshima city on a ban on dumping cans and cigarette butts, which consists of 61 predicate verbs in 20 provisions. Because of the difficulty of testing correctness of logical connectives or the logical formula itself, we focus on testing only whether the system correctly derives predicates, which correspond to words and semantic relations between nouns and verbs. Because we could not find any other models in terms of translating Japanese law sentences into logical forms, we do not compare experimental results with others.

We assume a baseline model which derives predicates of a semantic relation chosen by the surface form of a case particle instead of by the case frame dictionary. For example, a noun phrase with a case particle of '-*ga*' (が) is likely to become an agent. With some verbs, however, this particle has a different meaning, and the case frame dictionary may refuse to assign it to agent. As a result, our system realized 78.6% accuracy, while the baseline model was 61.2%.

Here, we show an example of results, as follows. The following text is a provision of Hiroshima city, concerned with obligations of the mayor.

Hiroshima city provision 13-2. When the mayor designates a district for promoting beautification, s/he must in advance listen to opinions from the organizations and the administrative agencies which are recognized to be concerned with the district[4].

Our system worked out the following logical formula, in which the implication (→) forms the boundary between the law requisite part and the law effectuation part, and this formula includes a modal operator for obligation in the law effectuation part:

$$\begin{aligned}
&\text{designate}(e_2) \wedge agt(e_2, x_0) \wedge \text{mayor}(x_0) \wedge obj(e_2, e_1)\\
&\quad \wedge \text{district_for_promoting_beautification}(e_1)\\
&\rightarrow \text{O}(\text{listen}(e_{12}) \wedge agt(e_{12}, x_0) \wedge obj(e_{12}, e_{11})\\
&\quad \wedge \text{opinion}(e_{11}) \wedge obj(e_{11}, x_{10}) \wedge \text{administrative_agency}(x_{10})\\
&\quad \wedge \text{organization}(x_9)\\
&\quad \wedge \text{recognize}(e_8) \wedge obj(e_8, x_{10}) \wedge obj(e_8, e_7)\\
&\quad \wedge \text{concern}(e_7) \wedge res(e_7, x_6) \wedge \text{district}(x_6)) \quad ,
\end{aligned} \tag{8}$$

where *agt*, *obj*, and *res* denote thematic roles of verbs in terms of agent, object and result, respectively. Because the noun phrase 'a district for promoting beautification' is a compound noun in Japanese, it is represented as one predicate. We point out incorrect parts as follows:

1. The variable attached to 'district_for_promoting_beautification' should not be an event variable but an object as x_1[5]. Hence, '$obj(e_2, e_1)$' should be '$obj(e_2, x_1)$.'

[4] The original sentence is as follows: (広島市条例第 13 条 2 項) 市長は、美化推進区域を指定しようとするときは、あらかじめ、当該区域に関係すると認められる団体等および行政機関の意見を聴くものとする。

[5] This problem has already been solved.

2. The predicate of '$obj(e_{11}, x_{10})$' should be '$agt(e_{11}, x_{10})$.'
3. The predicate of 'organization(x_9)' is neglected. It should be '(administrative_agency(x_{10}) ∨ organization(x_{10})).'
4. There are two object terms modifying the predicate 'recognize(e_8).' One of the predicates '$obj(e_8, e_7)$' should be '$res(e_8, e_7)$.'
5. The variable of the predicate 'district(x_6)' should be unified with 'x_1.'

The problems of 2 and 4 come from the lack of examples in the case frame dictionary. The problems of 3 and 5 took place, because we have not yet dealt with an appropriate semantic analysis to unify these variables.

Owing to Davidsonian-style, some predicates can refer to events corresponding to event variables. An example is shown in a part of the output; 'listen(e_{12}) ∧ $obj(e_{12}, e_{11})$ ∧ opinion(e_{11}),' in which the object of the event 'listen' is the event 'opinion,' which is recognized as an event due to a *sahen*-noun in Japanese. Using this style, we can represent such a simple formula. Otherwise, we would have to define a higher order predicate logic form in order to express predicates referring to predicates.

As long as we do not realize anaphora analysis, the system cannot find out from the sentence the agent of the event 'recognize,' which is an obligatory case of the verb. Davidsonian style allows us to make a logical formula for the verb, regardless of the number of arities for which the predicate is necessary.

5 Concluding Remarks

Our research purpose is to develop a system which translates legal texts into logical forms. We took into account linguistic characteristics of legal texts, regarding them as suitable for language processing, proposed a processing framework and showed the implementation of the present version.

Our present system provided high accuracy in terms of predicates corresponding to words and their semantic relations. Because the accuracy of semantic relations is mostly affected by that of the case frame dictionary, some errors in the example shown in Section 4 come from the lack of examples of the dictionary.

We reported our ongoing study in this paper. The rest of this section is spent on our plan for further development of the system. The remaining problems are concerned with (1) improvement of accuracy, and (2) formalization of output necessary for logical processing. Firstly, we expect to improve the analytical accuracy of our system by attaching syntactic rules to the parser. Particularly, complex sentences with paragraphs would be analyzed by structural analysis of law article sentences, and the hierarchical rules for conjunction and disjunction would be effective for legal texts, as was mentioned in Section 2.3. In addition, making a case frame dictionary adequate for stochastic processing is important. The more texts provided to the language analyzer, the greater the size of case frame dictionary, eliminating the problem of data sparseness. In fact, our current database lacks word use frequency. Although we manually annotated each item of the database for deep case makers, in the next version we aim for the realization of automatic annotation.

Secondly, forming logical representation is very important for the next step in logical processing. Assignment of quantifiers to logical forms is necessary for logical processing. Because the Japanese language tends not to describe quantifiers explicitly, it is difficult to do this. We expect to solve the problem also by linguistic characteristics of legal texts. For example, we can interpret a quantifier of '$\forall x$ Citizen(x)' from the sentence "All the citizens have the right," while Japanese tends to lack the expression 'all.' We can, however, recognize that subjective nouns denoting a person or an organization in a law sentence tend to be applied to a universal quantifier, while objective nouns are applied to an existential quantifier. Therefore, we could make a logical form by attaching quantifiers from the sentence "The mayor can dismiss a deputy mayor." to "$\forall x \exists y \exists e P(\text{mayor}(x) \wedge \text{dismiss}(e) \wedge \text{deputy\_mayor}(y) \wedge \text{agent}(e, x) \wedge \text{object}(e, y))$."

We have two strategies about modal operators. On the one hand, we consider adding other kinds of modal operators in order to allow flexible expressions. Tense operator is one of the major candidates. On the other hand, we consider removing modal operators from the formalization of logic representation, in order to realize smooth logical processing. This problem is a trade-off between language processing and logical processing.

Acknowledgment. This research was partly supported by the 21st Century COE Program 'Verifiable and Evolvable e-Society' and Grant-in-Aid for Scientific Research. We would like to give special thanks to Akio Kitada and Akira Ejiri, who developed a pilot version of the system as members of our laboratory, and to Yoko Shibata, Kaori Kohno and Yoshiko Oyama, who analyzed law sentences and their logical representations.

References

1. Katayama, T.: The current status of the art of the 21st COE programs in the information sciences field (2): Verifiable and evolvable e-society - realization of trustworthy e-society by computer science - (in Japanese). IPS Japan (Information Processing Society of Japan) Journal 46(5), 515–521 (2005)
2. Katayama, T.: Legal engineering – an engineering approach to laws in e-society age –. In: Proc. of the 1st Intl. Workshop on JURISIN (2007)
3. Hagiwara, S., Tojo, S.: Stable legal knowledge with regard to contradictory arguments. In: AIA 2006. Proceedings of the 24th IASTED international conference on Artificial intelligence and applications, Anaheim, CA, USA, pp. 323–328. ACTA Press (2006)
4. Jurafsky, D., Martin, J.H.: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice Hall PTR, Upper Saddle River, NJ, USA (2000)
5. Hobbs, J.R., et al.: Interpretation as abduction. In: Proceedings of the 26th annual meeting of the Association for Computational Linguistics, Morristown, NJ, USA. Association for Computational Linguistics, pp. 95–103 (1988)
6. Mulkar, R., Hobbs, J.R., Hovy, E.: Learning from reading syntactically complex biology texts. In: Proceedings of the 8th International Symposium on Logical Formalizations of Commonsense Reasoning. part of the AAAI Spring Symposium Series (2007)

7. Mulkar, R., et al.: Learning by reading: Two experiments. In: Proceedings of IJCAI 2007 workshop on Knowledge and Reasoning for Answering Questions (2007)
8. Nitta, K., et al.: New HELIC-II: A software tool for legal reasoning. In: ICAIL 1995. Proceedings of the 5th international conference on Artificial intelligence and law, pp. 287–296. ACM Press, New York, NY, USA (1995)
9. Ashley, K.D.: Modeling Legal Argument: Reasoning with Cases and Hypotheticals. In: Artificial Intelligence and Legal Reasoning Series, MIT Press (Bradford), Cambridge, Massachusetts (1990)
10. Tanaka, K., Kawazoe, I., Narita, H.: Standard structure of legal provisions - for the legal knowledge processing by natural language - (in Japanese). In: IPSJ Research Report on Natural Language Processing, pp. 79–86 (1993)
11. Parsons, T.: Events in the Semantics of English: A Study in Subatomic Semantics. MIT Press, Cambridge, MA (1990)
12. Zucchi, A.: The Language of Propositions and Events. In: Studies in Linguistics and Philosophy, Kluwer, Dordrecht (1993)
13. Yoshino, H.: Representation of legal knowledge by compound predicate formula. In: ICLP 1994. Proc. of Workshop on Legal Application of Logic Programming, pp. 128–137 (1994)
14. Nagai, H., Nakamura, T., Nomura, H.: Skeleton structure acquisition of Japanese law sentences based on linguistic characteristics. In: Proc. of NLPRS 1995, vol. 1, pp. 143–148 (1995)
15. Shimazu, A., Naito, S., Nomura, H.: Semantic structure analysis of Japanese noun phrases with adnominal particles. In: Proceedings of the 25th annual meeting on Association for Computational Linguistics, Morristown, NJ, USA. Association for Computational Linguistics, pp. 123–130 (1987)
16. Kurohashi, S., et al.: Improvements of Japanese morphological analyzer JUMAN. In: Proceedings of the Workshop on Sharable Natural Language Resources, pp. 22–28 (1994)
17. Kurohashi, S., Nagao, M.: KN parser: Japanese dependency/case structure analyzer. In: Proceedings of the Workshop on Sharable Natural Language Resources, pp. 48–55 (1994)
18. Ikehara, S., et al.: Nihongo Goi Taikei – A Japanese Lexicon (In Japanese). In: Iwanami Shoten, Tokyo (1997)

Automatic Consolidation of Japanese Statutes Based on Formalization of Amendment Sentences

Yasuhiro Ogawa, Shintaro Inagaki, 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. To realize computer-supported works in the areas of legislation and the practical use of laws, a statute database is indispensable so that users can easily retrieve desired versions of desired statutes. However, almost all former versions of statutes need to be restored since they cannot be easily obtained or digitalized. This task can be achieved by repeatedly consolidating amendment statutes with each version of the statute from the first one. In this paper, for Japanese statutes, we show that amendment clauses, which are parts of amendment sentences, can be formalized in terms of sixteen regular expressions. We also propose an automatic consolidation system for Japanese statutes based on the formalization and experts' knowledge about consolidation. System evaluation is shown through a consolidation experiment.

Keywords: automatic consolidation, statute document, amendment sentence.

1 Introduction

Recently, information technology has been introduced to legislation and the practical use of laws. To realize computer-supported works in such areas, an electronic statute database is indispensable. Since a non-retroactive principle for applying statutes and transitional measures for enforcing, amending, and repealing statutes exist, former versions and repealed statutes are also necessary for lawyers so that desired versions of desired statutes can be easily retrieved from the database. Thus, not only the current versions of every statute but also all former versions including repealed statutes must be stored in the database.

However, the texts of the former versions cannot be easily obtained. Law books in printed publications usually only include the currently effective statutes. Although some law books that include former statutes are published, they are not digitalized in general. On the other hand, recently some electronic databases of statutes (e.g., [1, 2]) have been established; however, they still usually include only the current versions at the time when retrieved, when the database was established, or when the statutes in it were enforced. Some databases include

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 363–376, 2008.

former versions; however, they are often only those since the databases were established. Thus, to realize the database that consists of every version of statutes, most must be restored in a digitalized form.

In principle, by repeating amendment again from the first version to the current one, every version can be acquired. However, this is not an easy task since the number of statutes is large and increasing. Even if we limit the statutes to acts enacted by the National Diet in Japan, about 1,800 are currently effective. In addition, more than 12,000 acts have been enacted during this 120 years, each of which generates a version of at least one act, and about 200 new acts, which include about 150 amendment acts, are enacted every year. In addition to the acts, about 5,500 orders and regulations enacted by the cabinet and ministries are currently in effect, and each of about 1,800 local governments has a large number of their own local ordinances.

Another problem exists. To amend statutes, there are two ways in general: enlargement and consolidation. The former is just only to add new provisions to the tail of a former version. The latter is to revise provisions in a former version word by word according to other provisions enacted as an amendment statute, where the details of the revision are clearly specified in terms of amendment sentences. In Japan, as in many countries, the latter is adopted.

For example of consolidation, consider the following article included in the current version of Wholesale Market Act (Act No.35, 1971) [1] before 1999.

第七十九条 次の各号のいずれかに該当する者は、五万円以下の罰金に処する。
/ **Article 79** Any person who falls under any of the following items shall be punished by a fine of not more than fifty thousand yen [2].

After that, it was decided to raise the price of the fine provided in this article from fifty thousand yen to five hundred thousand yen. To realize this amendment, Partial Amendment Act for Wholesale Market Act (Act No.109, 1999) was enacted that included the following amendment sentence:

Ex. 1 第七十九条中「五万円」を「五十万円」に改める。
/ In Article 79, 'fifty thousand yen' shall be replaced with 'five hundred thousand yen.' (Act No.109, 1999)

When this amendment act was enforced, this sentence was consolidated with the current version so that the article was revised as follows and the new version of Wholesale Market Act was generated:

第七十九条 次の各号のいずれかに該当する者は、五十万円以下の罰金に処する。
/ **Article 79** Any person who falls under any of the following items shall be punished by a fine of not more than five hundred thousand yen.

[1] "Act No.35, 1971" means that this was promulgated as 35th act in 1971.

[2] Hereafter, an English sentence following the slash is a translation of the Japanese one before it.

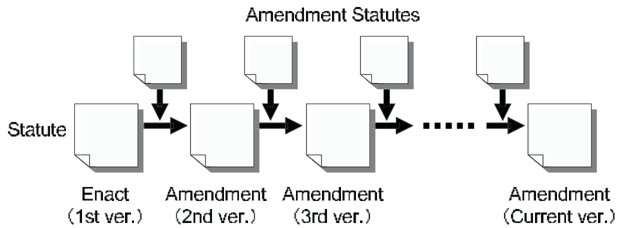


Fig. 1. Repetition of consolidation

Although every version can be acquired by repeating consolidation (Fig. 1), it obviously requires more works than enlargement. In addition, a new version sometimes cannot be easily and correctly obtained since consolidation needs knowledge. In fact, consolidation has so far been achieved manually by experts on legislation as paper-based work, and the knowledge must usually be acquired from technical guidebooks on legislation (e.g., [3, 4]) or from other experts.

These problems suggest the necessity of an automatic consolidation system for statutes, whose design is the purpose of this paper. To realize the system, the meaning of amendment sentences must be clarified. Furthermore, if knowledge about consolidation is used to clarify the meaning, it must be explicit and be implemented in the system. Fortunately, if we limit the problem to Japanese legislation, the amendment sentences seem to be restricted to a subset of Japanese sentences, i.e., their syntax obey a certain grammar, and their meaning is also considered unambiguous at least by experts. Actually, the syntax and meaning of amendment sentences have been already introduced by the guidebooks on legislation. However, since amendment sentences are usually explained by examples and might not be exhaustively mentioned, they should be investigated in detail.

In fact, such an automatic consolidation system has already been proposed [5]. This system, however, was designed for English statutes, where the whole of amendment sentences which can be processed is not explicitly shown and it seems unclear whether knowledge about consolidation is necessary in the system implementation.

In this paper, from a viewpoint of Japanese language processing, we show that, for Japanese statutes, amendment clauses, which are parts of amendment sentences, can be formalized in terms of sixteen regular expressions. Furthermore, we propose an automatic consolidation system based on the formalization, which is specific to Japanese statutes. Knowledge about consolidation in Japanese legislation is also implemented as programs. The system utilizes XML (eXtensible Markup Language) techniques for implementation. Supported by the system, every version of statutes can be easily restored.

This paper is organized as follows: in Section 2, the formalization of amendment clauses is shown by concentrating on the verbs in them. Section 3 shows the design and implementation of an automatic consolidation system, where the intended meaning of amendment clauses and some knowledge about consolidation are represented as programs. In Section 4, system evaluation is shown through a consolidation experiment, and Section 5 summarizes this paper.

2 Formalization of Amendment Sentences

Amendment sentences are fundamental for amending statutes since consolidation is executed step by step for a sequence of the sentences based on their intended meaning. That is, amendment sentences can be regarded as instructions for consolidation. The following are other examples of amendment sentences:

Ex. 2 第十三条第二項に次のただし書を加える。
/ The following proviso shall be added to Paragraph 2 of Article 13. (Act No.152, 2002)

Ex. 3 第四十一条第二項第四号を削る。
/ Item 4 of Paragraph 2 of Article 41 shall be deleted. (Act No.102, 1999)

Ex. 4 第十六条の次に次の章名を付する。
/ The following chapter title shall be attached to the next to Article 16. (Act No.138, 2002)

Ex. 5 第二十八条を第三十条とし、第二十二条から第二十七条までを二条ずつ繰り下げる。
/ Article 28 shall be set as Article 30, and Articles from 22 to 27 shall be shifted down two articles at a time. (Act No.102, 1999)

To realize an automatic consolidation system, the intended meaning of amendment sentences must be identified. Thus, we compiled a corpus of the sentences to analyze their meaning. In this section, we describe the details.

2.1 Compilation of an Amendment Sentence Corpus

First, we manually extracted and checked 8,114 amendment sentences from 192 acts [1] enacted in 2002. Then we found that the main predicates in the sentences, which decide the amendment content, include only six verbs: (1) "改める / replace," (2) "加える / add," (3) "削る / delete," (4) "とする / set," (5) "付する / attach," and (6) "繰り上げる; 繰り下げる / shift up; shift down." [3] Hereafter, we call them *amendment verbs*. This suggests that the meaning of amendment sentences can be roughly identified by the amendment verbs in them.

Next, we developed an automatic extraction tool of amendment sentences from statute texts that utilizes both the occurrences of amendment verbs and conventional formatting rules peculiar to amendment sentences in Japanese statute documents [3, 4]. For example, the rules include that there are two full-size spaces as indentation in the head of each amendment sentence in the main provision when amendment acts are article-based and that there are one full-size space when item-based, etc. By using this tool, we were able to extract

[3] The last kind of verbs includes two. The only difference between their meanings, however, is the shift direction so that we consider them the same kind.

8,004 amendment sentences from the acts enacted in 2002 so that the extraction recall and precision were 98.6 and 99.0%, respectively, when the above 8,114 sentences are set as correct answers. The reasons that valid amendment sentences were missing and invalid ones were extracted are typos, especially in verbs, and wrong indentations in digitalized statutes.

Next, from 2,200 acts [1] enacted from 1989 to 2004, we extracted amendment sentences by using this tool and manually excluded invalid ones from them so that we succeeded in compiling a corpus of 76,156 amendment sentences.

Here, note that an amendment sentence may consist of several clauses, which we call *amendment clauses*. For example, the sentence in Ex. 5 consists of two amendment clauses. Since each amendment clause specifies a primitive action of an amendment, we divided the extracted amendment sentences into clauses at the points just after a Japanese punctuation symbol "、" that follows amendment verbs. So the sentence in Ex. 5 is divided into

"第二十八条を第三十条とし、
/ Article 28 shall be set as Article 30"

and

"第二十二条から第二十七条までを二条ずつ繰り下げる。
/ Articles from 22 to 27 shall be shifted down two articles at a time."

Thus, we acquired 157,878 amendment clauses from the corpus.

2.2 Analysis of Amendment Clauses

By carefully observing the extracted amendment clauses, we found that intended actions in them can be classified into the following ten kinds:

1. Actions on strings in a statute text: (a) Replacement, (b) Addition, and (c) Deletion.
2. Actions on structure elements of a statute such as sections, articles, items, etc.: (a) Replacement, (b) Addition, and (c) Deletion.
3. Actions on numbers of structure elements: (a) Renumber, (b) Attachment, and (c) Shift.
4. Combined action of renumbering of a structure element and replacement of its title strings [4].

Table 1 shows the distribution of these actions by amendment verbs, where we can observe that almost half of the actions are string replacements by using amendment verb "改める / replace." Furthermore, paying attention to twelve non-zero elements in Table 1, we can recognize that the usage of amendment verbs is restricted to specifying each action of amendments so that the actions can be classified into twelve classes by their meaning and amendment verbs.

[4] For example,
"「第五章　第二種指定製品」を「第七章　指定表示製品」に改める。
/ 'Chapter 5 Second Class Designated Products' shall be replaced with 'Chapter 7 Designated Products with Marks'." (Act No.113, 2000).

Table 1. Distribution of amendment actions by amendment verbs

Action		Amendment verbs						Total	Ratio (%)
		改める replace	加える add	削る delete	とする set	付する attach	繰り上げる; 繰り下げる shift up; shift down		
String	Repl.	72,757	0	0	0	0	0	72,757	46.1
	Add.	0	17,500	0	0	0	0	17,500	11.1
	Del.	0	0	9,762	0	0	0	9,762	6.2
Structure	Repl.	9,292	0	0	0	0	0	9,292	5.9
	Add.	0	17,656	0	0	1,068	0	18,724	11.9
	Del.	0	0	7,302	0	0	0	7,302	4.6
Number	Renum.	0	0	0	21,527	0	0	21,527	13.6
	Attach	0	0	0	0	197	0	197	0.1
	Shift	0	0	0	5	0	774	779	0.5
Combined		38	0	0	0	0	0	38	0.0
Total		82,087	35,156	17,064	21,532	1,265	774	157,878	100.0
Ratio (%)		52.0	22.3	10.8	13.6	0.8	0.5	100.0	—

2.3 Regular Expressions for Amendment Clauses

Although amendment verbs are important clues to identify the meaning of amendment clauses, except "繰り上げる; 繰り下げる / shift up; shift down", each of the verbs is used to describe two or three actions. For example, as shown in Table 1, "加える / add" describes both the addition of a string in a statute text and the addition of a structure element to a statute. Furthermore, two kinds of amendment verbs are used to add structure elements and to shift their numbers.

To distinguish them, more information has to be acquired from amendment clauses. Therefore, we concentrate on words that grammatically depend on the verb, such as objectives and datives. For example, in the following amendment clause, the objective of verb "加える / add" is "「等」 / 'etc.'," so we recognize that this action adds a string to another string in the text:

Ex. 6 第一条中「融資」の下に「等」を加える。

/ In Article 1, 'etc.' shall be added after 'loan.' (Act No.1, 2002)

On the other hand, although Ex. 2 has the same verb "加える / add " as Ex. 6, its objective is "ただし書 / a proviso," which does not mean the addition of a string but the addition of a structure element called a proviso.

It is also necessary to acquire other parameterized information for each action: the position to be amended in a statute such as "第一条中 / in Article 1" in Ex. 6 or "第十六条の次 / the next to Article 16" in Ex. 4, the kinds of structure elements to be added, the shift steps of attached numbers, etc. To acquire such detailed information and precisely identify the meaning of amendment clauses,

Table 2. Regular expressions for Japanese amendment clauses

<table>
<tr><th>No.</th><th>Regular expression</th></tr>
<tr><td>1</td><td>/ (.*?) 中?(「.*」) を、?(「.*」) に改め (、|る。) /</td></tr>
<tr><td>2</td><td>/ (.*) の (見出し|. 名) を「(.*)」に改め (、|る。) /</td></tr>
<tr><td>3</td><td>/ (.*) を次のように改め (、|る。) /</td></tr>
<tr><td>4</td><td>/ ((.*?中)?「(.*)」の下に、?「(.*)」を加え (、|る。)) /</td></tr>
<tr><td>5</td><td>/ ((.*) に)?(.+?) として、?次の ((.+) を|ように) 加え (、|る。) /</td></tr>
<tr><td>6</td><td>/ ((.*) に)?次の (後段|ただし書) を加え (、|る。) /</td></tr>
<tr><td>7</td><td>/ ((.*) の (前|次)) に次の ((.+?) を|ように) 加え (、|る。) /</td></tr>
<tr><td>8</td><td>/ (.*?) に次の ((.+?) を|ように) 加え (、|る。) /</td></tr>
<tr><td>9</td><td>/ ((.*?) 中?「(.*)」を削 (り、|る。)) /</td></tr>
<tr><td>10</td><td>/ (.*) を削 (り、|る。) /</td></tr>
<tr><td>11</td><td>/ (.*?)?([ア-ン]) から ([ア-ン] まで) を (.*?)?([ア-ン]) から [ア-ン] までと (し、|する。) /</td></tr>
<tr><td>12</td><td>/ (.*) を (.*) と (し、|する。) /</td></tr>
<tr><td>13</td><td>/ (.*) に見出しとして「(.*)」を付 (し、|する。) /</td></tr>
<tr><td>14</td><td>/ ((.*) に)?項番号を付 (し、|する。) /</td></tr>
<tr><td>15</td><td>/ ((.*) に)?(次の)?(.*) を付 (し、|する。) /</td></tr>
<tr><td>16</td><td>/ (.*?)(から|及び)(.*?)(まで)?を (.*) ずつ繰り (上げ|下げ)(、|る。) /</td></tr>
</table>

we formalized their syntactical patterns as sixteen extended regular expressions used in programming language Ruby; Table 2 shows all of them. For example, amendment sentences Exs. 1 and 2 match expressions 1 and 6, respectively. In fact, we confirmed that every amendment clause in the corpus is matched with one of these regular expressions when verifying them in order from the top one in Table 2.

Here, except expression 11, all the syntactical patterns that match these expressions can be found in some of many guidebooks on legislation (e.g., [3, 4]). This fact reflects the suggestion from the guidebooks that the syntax of amendment sentences are restricted. On the other hand, the amendment clauses that match expression 11, which denote the shift of some subitems, are not mentioned as really used examples in any guidebooks at all. In fact, only five clauses that match this expression can be found in the corpus and all of them are enacted after 2003. Thus, this pattern can be considered as a new one which began to be used recently.

Note that there are some different regular expressions that describe identical kinds of actions and use the same verb. For example, the following clause, which matches expression 5, describes the same kind action as Ex. 2:

Ex.7 第九条にただし書として次のように加える。
/ The following shall be added to Article 9 as a proviso. (Act No. 65, 2002)

Thus, we acquired sixteen expressions although the actions are classified into twelve.

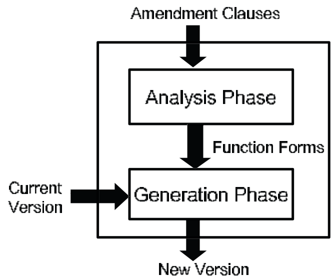


Fig. 2. Design of the automatic consolidation system

3 The Automatic Consolidation System

In this section, we describe the design and implementation of an automatic consolidation system of Japanese statutes. The system has two phases: the analysis phase of amendment clauses and the generation phase of a new version (Fig. 2).

3.1 Analysis of Amendment Clauses

In the first phase, the system retrieves a regular expression that matches a given amendment clause by verifying the expressions in the order shown in Table 2. At this time, the system extracts the necessary information to execute consolidation in the next phase from the expression and formalize it in a function form, which is one of eleven kinds as shown in Table 3. The last row of the table shows the numbers of corresponding regular expressions in Table 2 for each action class. For example, Exs. 1 and 2 in the previous section, which correspond to regular expressions 1 and 6, respectively, are transformed into the following function forms, respectively:

```
substitute_string ("第七十九条", "五万円", "五十万円")
/ substitute_string ("Article 79", "fifty thousand yen",
                     "five hundred thousand yen")
add_structure ("第十三条第二項", doc, "ただし書")
/ add_structure ("Paragraph 2 of Article 13", doc, "a proviso")
```

Except the one denoted as `doc`, each argument in function forms corresponds to the necessary information for consolidation that explicitly appears in an amendment sentence, while `doc` corresponds to the part that follows the sentence, to which the term "次の / the following" in the sentence refers.

Furthermore, in principle, each action class in Section 2.2 corresponds to each function form. However, since the case when the shift of the attached numbers of structure elements is described by using amendment verb "とする / set" is quite rare (Table 1), the function form for this class is unified with the one for the same kind of action by amendment verb "繰り上げる; 繰り下げる/ shift up; shift down" merely for convenience of system implementation.

Table 3. Function forms for amendment clauses

Action		Amendment verb		Function form	Regular exp.
String	Repl.	改める	replace	`substitute_string (pos, str1, str2)`	1
	Add	加える	add	`add_string(pos, str)`	4
	Del.	削る	delete	`delete_string(pos, str)`	9
Structure	Repl.	改める	replace	`substitute_structure (pos, doc)`	2, 3
	Add	加える	add	`add_structure (pos, doc, elm)`	5, 6, 7, 8
		付する	attach	`attach_structure (pos, doc, elm)`	13, 15
	Del.	削る	delete	`delete_structure(pos)`	10
Number	Renum.	とする	set	`renumber(str1, str2)`	12
	Attach	付する	attach	`attach_number(pos)`	14
	Shift	とする	set	`shift(pos, stp)`	11
		繰り上げる; 繰り下げる	shift up; shift down		16
Combined		改める	replace	`rename(str1, str2)`	1

3.2 Additional Processing for Generating Function Forms

Some cases need additional processing to generate function forms. We illustrate four of them below.

First, as shown in Table 3, there are two action classes that match regular expression 1 in Table 2; replacement of strings, and combined action of renumbering of a structure element and replacement of its title string. In fact, as illustrated in Ex. 1 for the former and in footnote 4 for the latter, amendment clauses for describing both actions are similar on the surface. Thus, to distinguish them we use the following properties peculiar to amendment clauses for the latter: (1) A string which denotes a position to be amended in a statute such as "第七十九条中 / in Article 79" in Ex. 1 does not appear, and (2) Strings before and after the replacement match the regular expression

```
/ (第 [一二三四五六七八九十百千]+([編章節款目])([のノ][一二三四五六七
  八九十百]+)*) /
```

which describes the patterns of title strings to which this action can apply.

Next, if an amendment clause includes an expression for several structure elements to be amended at a time, e.g., "第一条、第三条及び第四条中 / In Articles 1, 3, and 4," the clause is extended and transformed into several function forms. This is realized by finding a punctuation symbol "、" and a conjunction such as "及び / and" in the part of the clause that specifies the position to be amended. The part is divided so that each generated function form include only one structure element to be amended.

Thirdly, repetition of the same kind of action such as replacement, addition, and deletion of strings in statute texts is sometimes described in a single amendment clause where an amendment verb occurs only once as follows:

Ex. 8 第八条中「十二人」を「七人」に、「三人」を「二人」に改める。
/ In Article 8, 'twelve persons' shall be replaced with 'seven persons,' and 'three persons' with 'two persons.' (Act No.16, 2000)

Since this clause essentially denotes two executions of string replacement, we extend and transform it into two function forms by repeating string matching with another kind of regular expression, e.g.,

```
/ (([^「.+」]+?) 中)?「(.*?)」を、?「(.*?)」に、? /
```

for the replacement, so that we can deal with any times of the repetition and extract necessary information for each function form.

Finally, consider the following amendment sentence:

Ex. 9 第二十五条中「第四条第一項」を「この法律」に改め、「又は都道府県知事」を削り、同条の次に次の一条を加える。
/ In Article 25, 'Paragraph 1 of Article 4' shall be replaced with 'this act,' 'or a prefectural governor' shall be deleted, and the following one article shall be added to the article. (Act No.87, 1999)

This sentence is divided into the three clauses. However, from the second one,

"「又は都道府県知事」を削り、
/ 'or a prefectural governor' shall be deleted,"

we can not extract the structure element that includes the string to be deleted. On the other hand, in the third one,

"同条の次に次の一条を加える。
/ the following one article shall be added to the article,"

we can not identify which article is exactly denoted as "同条 / the article." To solve these problems of anaphora resolution so that all arguments in function forms are completed, we employed a heuristic rule: *the structure element that appears immediately before in a sequence of function forms is the one desired.* Although anaphora resolution is problematic in general, this simple rule works very well in the processing of amendment sentences. In fact, for example, for the latter case in the consolidation experiment shown in Section 4, there were 378 structure elements to be identified, denoted as "同条 / the article," "同項 / the paragraph," and "同号 / the item," all of which were completely resolved.

3.3 Generation of New Versions of Statutes

The second phase of the system is the generation of a new version of a statute, whose inputs are both the function form transformed from a given amendment clause and the current version of a statute to be amended.

In this system, the statute documents are supposed to be structured in terms of XML. Generally, a statute has logical structure elements such as a title, contents, the main provision, supplementary provisions, etc. The main provision has also a hierarchical structure that consists of parts, chapters, sections, and subsections, etc. In addition, the elements below chapters are also hierarchical, whose substructure includes articles, paragraphs, items, subitems, etc. We designed DTD (Document Type Definition) to specify these structure elements of Japanese statutes and marked-up the statute documents based on this DTD.

On the other hand, the function forms given to this phase are regarded as functions of programming language Ruby and directly executed. Each function is implemented as an operation to strings in the texts or to trees that denote the logical structures of statutes, where the intended meaning of each corresponding action and the knowledge about consolidation acquired from guidebooks and experts are reflected. For example, as shown in Ex. 8, in the case of amendment clauses that describe the repetition of string replacements in a statute text, all the replacements have to be executed in parallel [4]. Without this knowledge, consolidation based on the following type of amendment clause

"第一条中「A」を「B」に、「B」を「C」に改める。
/ In Article 1, 'A' shall be replaced with 'B', and 'B' with 'C'. "

will result in a new version with no occurrences of 'B' in Article 1 since we sequentially execute string replacement twice, which is incorrect. To implement this knowledge, we blocked the occurrences of 'B' that were replaced from 'A' so that they can be prevented from the replacements with 'C.'

4 Consolidation Experiments

To evaluate the behavior of the developed system, we carried out an experiment on a series of consolidations. The test set of statutes to be amended consists of the first versions of seventeen acts [1] enacted after 1947. With each of them, at least one and at most sixteen amendment acts [1] need to be consolidated to obtain the current versions. From these amendment acts, 965 amendment clauses were extracted, 187 of them were not included in the corpus compiled in Section 2.1. The distribution of the kinds of actions intended by these clauses is shown in Table 4, which nearly equals to the one in the corpus (Table 1).

For each act in the test set, related amendment clauses are sequentially transformed into function forms based on the order of the enforcement dates of the amendment acts and the order of their appearance in each amendment act. As a result, the system generated 1,164 function forms for all clauses.

Then we compared the final version of each act generated by our system with the current version obtained from the existing database [2]. The texts to be compared are leafs when the structures of the acts are regarded as trees. There were 4,355 such texts, and 4,332 in them were identical, which means that consolidation succeeded. The causes of the 23 errors are classified in Table 5, where all except two cases are in the test set. Thus, the system's behavior is almost confirmed.

Table 4. Distribution of amendment actions in the test set

Action		Number of clauses	Ratio (%)
String	Repl.	419	43.4
	Add	88	9.1
	Del.	54	5.6
Structure	Repl.	60	6.2
	Add	114	11.8
	Del.	41	4.2
Number	Renum.	183	19.0
	Attach	0	0.0
	Shift	5	0.5
Combined		1	0.1
Total		965	100.0

Table 5. Causes of errors in the consolidation experiment

Place	Kinds	Number
Source texts (1st ver.)	Typo	17
Source texts (current ver.)	Typo	2
	Incorrect consolidation	2
System	Incorrect generation of function forms	1
	Incorrect consolidation	1
Total		23

4.1 Incorrect Generation of Function Forms

One of the two errors is incorrect generation of function forms which is caused by replacement of a proviso. The former version of a proviso that were included in Article 5 of Quarantine Act (Act No.201, 1951) was a single sentence as follows:

但し、検疫所長の許可を受けた場合は、この限りでない。
/ provided, however, this shall not apply to the case when permission of a quarantine station's head is obtained.

The following amendment sentence intends to replace this proviso with the one that includes several items [5]:

Ex. 10 第五条ただし書を次のように改める。
ただし、次の各号のいずれかに該当するときは、この限りでない。
一　…を運び出すとき。
二　…を運び出すとき。
/ The proviso of Article 5 shall be replaced with the following:
provided, however, this shall not apply to the case which falls under any of the following items:

[5] Sentences in Exs. 10 and 11 are shortened for convenience.

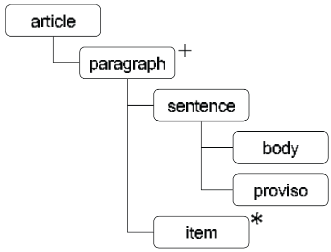


Fig. 3. DTD for Japanese statues (part)

(1) The case when ... are carried out.
(2) The case when ... are carried out. (Act No.45, 1970)

In our DTD, however, items are not formalized as subtrees of a proviso but as the ones of a paragraph (Fig. 3). Thus, in this case, the system should generate two function forms: for the replacement of a body sentence in the proviso and for the attachment of items to the paragraph that is a parent of the proviso. According to the DTD, this amendment clause essentially requires grafts of several trees at a time, where each parent of them is on different levels. Since the system was designed to process only a single tree, it could not process this situation. In fact, such a structure was overlooked when we designed the DTD since its occurrence is quite rare. Thus, to solve this problem, revising the DTD could be considered.

4.2 Incorrect Consolidation

The other error is caused by the following amendment sentence:

Ex. 11 第二十六条第一項中「次項」を「第四項」に改め、…、第二十六条第一項第四号中「(… 次項第四号において同じ。)」を削り、
/ In Paragraph 1 of Article 26, 'the next paragraph' shall be replaced with 'Paragraph 4,' ..., and in Item 4 of Paragraph 1 of Article 26 '(... the same shall apply in the next paragraph)' shall be deleted. (Act No.30, 1995)

The following function forms were generated from two clauses in this sentence:

f_1: `substitute_string ("第二十六条第一項","次項", "第四項")`
`/ substitute_string ("Paragraph 1 of Article 26", "the next paragraph", "Paragraph 4")`
f_2: `delete_string ("第二十六条第一項第四号中", "（… 次項第四号において同じ。）")`
`/ delete_string ("Item 4 of Paragraph 1 of Article 26", "(... the same shall apply in the next paragraph)")`

The system executes f_1 and f_2 in this order, that is, it executes the string replacement first and then the deletion of another string. However, the execution of f_2 had no effects at all in Item 4 since the string in it, which is to be deleted by f_2, has been already replaced with another string "（··· 第四項第四号において同じ。） / (... the same shall apply in Paragraph 4)" by f_1.

As is the case with parallel executions of string replacement in Section 3.3, this case requires another kind of knowledge about consolidation which suggests that there is a case where replacement and deletion of strings have to be executed in parallel. To clarify it, we will investigate more examples as a future problem.

5 Conclusion

In this paper, to compile a database of all versions of Japanese statutes, we proposed an automatic consolidation system based on the fact that amendment clauses for Japanese statutes can be formalized in terms of sixteen regular expressions, where some knowledge about consolidation is implemented as programs. Through a consolidation experiment, we confirmed the behavior of the system.

However, as pointed out in Section 4.2, when several string operations are included in a single amendment sentence, it is a problem to decide whether they are to be executed in parallel or in sequential.

Furthermore, other two cases of amendments still remain to be considered in this system, which were excluded in the experiment in the previous section: amendments of tables in statutes and amendments of amendment sentences. The ratios of amendment sentences for these cases in the corpus are about 7.0 and 0.6%, respectively. In the former, formalization of the structure and expressions of amendment sentences are complicated since the parts of tables including ruled lines can be amended. The latter includes a problem to identify amendment sentences to be amended since they are referred to by their contents.

However, except these problems that remain as future ones, since most cases are typical, appear frequently, and can be processed, the system can support the restoration of former versions of Japanese statutes.

Acknowledgments. The authors would like to thank Prof. Yoshiharu MATSUURA, Assoc. Prof. Tokuyasu KAKUTA, and Mr. Tomoya SANO, Graduate School of Law, Nagoya University, for their support. This research project was partly supported by Grant-in-Aid for Scientific Research (B) of JSPS and Grant-in-Aid for Exploratory Research of MEXT.

References

1. The House of Representatives: Seitei Horitsu, `http://www.shugiin.go.jp/index.nsf/html/index_housei.htm`
2. Ministry of Internal Affairs and Communications: Horei Data Teikyo System, `http://law.e-gov.go.jp/cgi-bin/idxsearch.cgi`
3. Maeda, M.: Workbook Hosei Shitsumu (revised ed.), Gyosei, Tokyo (2003)
4. Hasegawa, S.: Horei Kaishaku no Kiso, Gyosei, Tokyo (2000)
5. Arnold-Moore, T.: Automatically Processing Amendments to Legislation. In: Proc. Int. Conf. Artificial Intelligence and Law, pp. 297–306 (1995)

Characterized Argument Agent for Training Partner

Takahiro Tanaka, Norio Maeda, Daisuke Katagami, and Katsumi Nitta

Interdisciplinary Graduate School of Science and Engineering
Tokyo Institute of Technology,
4259 Nagatsuta, Midori-ku, Yokohama, Japan
{takat,maeda,katagami,nitta}@ntt.dis.titech.ac.jp

Abstract. For the resolution of disputes, Alternative Dispute Resolution (ADR) has become a popular replacement for trials. However, for mediation to work, the mediator must undergo extensive training. To help with mediator training, we have developed an online mediation support system. In this paper, we present an overview of the system and an argument agent. The argument agent participates in moot mediation as a disputant for self-training purposes. We also explain how an agent's text response can be generated by retrieving from a case base situations which are likely to be met with in dispute resolution.

1 Introduction

In recent years, the number of disputes arising from online shopping and online auctions has been steadily increasing. Instead of settlement by trial, prompt and low-cost mediation and arbitration called Alternative Dispute Resolution (ADR) is now gaining popularity.

Training by moot mediation is an effective way of teaching mediation skills. However, when students participate in a moot mediation, they always need to be accompanied and guided by an instructor and, ideally, they need to be at the same physical location, which presents a potential problem. In an attempt to overcome these problems, we have developed an online mediation support system [2]. Using this system, students can participate in a moot mediation from any network-enabled device, regardless of physical location. This system collects logs of past mediations, and during mediation training, this system shows students past discussions that followed a similar pattern to the scenario being dealt with. This information was found to be an effective substitute for an actual instructor's advice [8][11]. However, the current system still has a problem in that at least three students – a mediator and two disputants - are necessary to initiate a moot mediation. To solve this problem, we have developed an argument agent which participates in the moot mediation as a disputant for self-training.

In this paper, we explain a way of generating the argument agent's text response using similar situations from a case base. In Section 2, we introduce the online mediation support system and in Section 3, we present an overview of the argument agent and the case-based text response generation for the agent. In Section 4, we show the evaluation of the argument agent, and in Section 5, we summarize this paper.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 377–389, 2008.

2 Online Mediation Support System

2.1 Overview of the System

Figure 1 shows an overview of the *Online Mediation Support System*. Online mediations are carried out by students connecting to the server through the Internet and using the Argument Interface. A trainee mediator and the two parties' agents participate. The *case ontology* includes background knowledge about the case to be mediated. The *mediation model* includes the basic rules of mediation. Each text response is indexed by referring to the case ontology, and is stored in the case base. During the mediation, a mediator can retrieve similar past cases and find suggestions to help in selection of his/her next response.

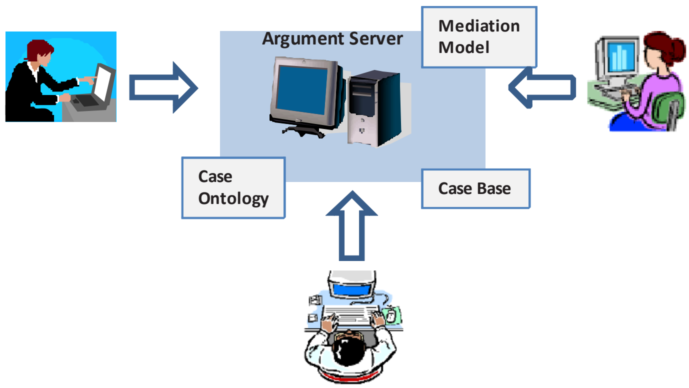


Fig. 1. Overview of the Online Mediation System

2.2 Case Ontology

The case data used in this research were natural-language texts representing past mediation dialogues. To be able to analyze them statistically and to search for similar situations from the case base, we indexed the remarks referring to the case ontology. The ontology consists of three types of information: a factor list, relationship among factors, and keywords which characterize the factors [2].

Factors are legal concepts which appear in the conversation during a mediation, and are used to summarize the conversation by observing only the relevant key concepts. Factors are almost the same concepts of abstract factors and leaf factors introduced by Aleven [4]. The following are examples of factors which appeared in a case that dealt with an auction.

c1: <u>The product auctioned has a defect.</u> [B, Share]

f11: It has a scratch. [B, Share]

f12: It is broken. [B, Share]

c2: The product description in the auction site was not complete. [B, Share]

f21: The product picture showed the defective part. [S, Share]

f22: The description didn't identify the defect. [B, Share]

c3: A complaint was justified. [B, Share]

f31: The complaint was made in proper time. [B, Share]

f32: It was a serious enough defect to cancel the contract. [B, Secret(S)]

c4: The seller had announced "NO CLAIM NO RETURN" on the auction site. [S, Share]

These factors are not independent. There is a relationship among them, which we call, *attack* and *support*. For example, f11 is a more concrete issue than c1. Factor c3 and c4 contradict each other because c3 is advantageous for the buyer (B) and c4 for the seller (S). In this example, we say "f11 supports c1" and "c4 attacks c3." These relationships are represented in the diagram (Figure 2). In this Figure, '+' and '-' represent 'support' and 'attack', respectively.

Recognition of factors from natural language sentences is generally a hard problem because it requires the ability to parse high-level language. However, in our case, the task is simpler because all dialogue is limited by mediation subject. Therefore, instructors can anticipate and prepare keywords for each factor before the moot mediation starts. If keywords related to a factor appear in a text response, our system would determine that the response is related to that factor.

Moreover, by referring to the relationship among factors, our system estimates the argument formula for a particular text response. For example, if a text response includes both *c3* and *f31*, then by referring to Figure 2, our system estimates that the text response has the argumentation formula "*c3 <- f31"* (*f31* implies *c3* holds). Furthermore, if the next text response by the opponent includes *c4*, we estimate that it has the formula "^*c3 <- c4*" (*c3* doesn't hold because *c4* holds). Here, '<-' and '^' represent 'implication' and 'negation', respectively. As it is often hard to decide on a

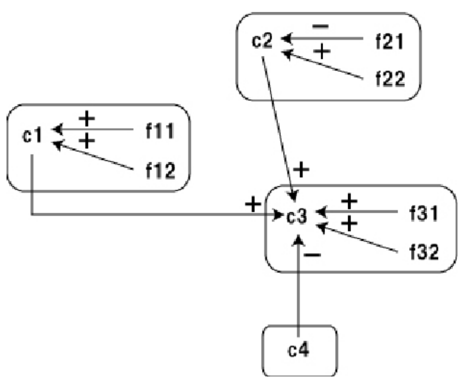


Fig. 2. Relationship among factors

proper set of keywords beforehand, we decided on one by analyzing sample mediation logs. However, the keyword set could still be incomplete.

Our system may also extract more than one factor set from the same text response. Some vagueness is resolved by discarding factor sets which don't fit the relationship among factors.

2.3 Mediation Model

The mediation model defines the rule of mediation that consists of the requirement and the effects of mediation moves. We discriminate among the following mediation moves: *claim*, *concede*, *deny*, *explain*, *withdraw*, *ask*, *answer, call-on, require-explanation* and other. These moves are related to each other. For example, when a party X claims (or proposes) P, the mediator will usually "call on" party Y. Then, if a mediator doesn't agree to P, he may deny it and require an explanation of P to party X. When party Y is called on after party X proposes P, and if Y disagrees to P, then he may deny P. Then party X has to explain why P holds. Such relations among mediation moves are defined in the mediation model (Table 1). Here,

'explain(P, P<-A)'

shows that P holds because A holds.

Table 1. Mediation Model

Action of Party X	Action of Mediator	Action of Party Y	
		Disagree	Agree
claim(P)	call-on(Y)	deny(P) require-explanation(P)	concede(P)
claim(P)	deny(P) require-explanation(P)		
deny (P)	call-on(Y)	explain(P, P<-A)	withdraw(P)
explain(P, P<-A)	call-on(Y)	deny(P<-A) explain(^P, ^P<-B)	concede(P<-A)
require- explanation(P)	call-on(Y)		explain(P, P<-Q)
deny(P<-A)	call-on(Y)	explain(P, P<-B)	withdraw(P<-A)
concede(P)	call-on(Y)		
concede(P<-A)	call-on(Y)		
withdraw(P)	call-on(Y)		

2.4 Argument Interface

The trainee mediator conducts the online mediation with the Argument Interface as shown in Figure 3. This interface is implemented in Adobe Flash which is software for running moving images on web browsers. The mediator inputs "remark texts (text response)", "link data (mediation move)" and "avatar expressions". "Link data" includes data that represents how the present remark relates to which past remarks. "Avatar expressions" are facial expressions and gestures of animated agents: *Cool*, *Happy*, *Angry*, *Sad* and *Surprised*. The animated agents are able to change facial expressions and read out the received text. Facial expressions are essential in the mediation process, since they convey important nonverbal information [6].

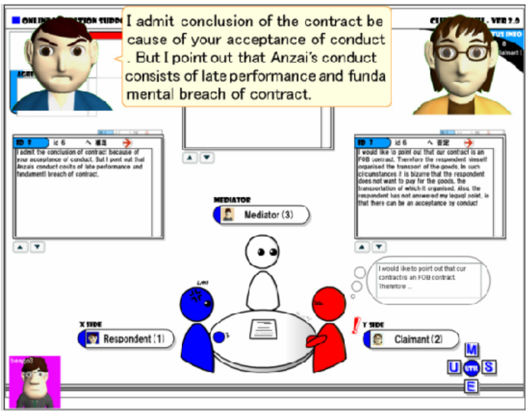


Fig. 3. Argument Interface

2.5 Construction of a Case Base

Each participant's typed text is provided as a part of natural language text and there is a *link data specifier*. A link data specifier is the relationship between the current and the previous one. Our system extracts the point at issue (factor) and the argumentation formula in the text response. Consequently, one text response has three pieces of indexed data; (1) the point at issue (factor) in the text response, (2) the argumentation formula in the text response, and (3) a link data specifier. After the end of mediation, the system automatically annotates the following indices: speaker name, text, facial expression, issue point (factor), argument formula and a link data specifier. Finally, the system stores the mediation record in the case base.

A case base is used in two ways. One way is for the mediation logs in a case base to be compared to each other and analyzed statistically. Statistical analysis helps to evaluate the mediator's mediation skills and the disputant's character. The other way is for similar scenario case to be retrieved during mediation. Similar scenarios are useful when a participant becomes stuck, doesn't know what the next text response should be, and needs words to bring about a satisfactory mediation.

(1) Analysis of mediation logs in a case base

Here, we show a sample analysis of 11 mediation logs which share a common theme of a dispute between a seller and a buyer. The story is as follows.

A secondhand automobile muffler was entered in the online auction by Mr. X. Though mufflers are usually made of steel, this muffler is made of poor quality material, and the fact that the material of the muffler was not described in the auction site because Mr. X didn't know its material. The muffler had been made by Z-corporation. Z-corporation had made mufflers of poor quality material in the past. These mufflers were listed in the catalogue and Web pages of Z-corporation three years ago. However, currently, Z-corporation doesn't make the mufflers of poor quality. Mr. Y bought this muffler, and he left it untouched in his room for a while. Two month later, he used the muffler first time, and found it wasn't made of steel. He wanted to cancel the contract and get the money back. However, Mr. X denies to return the money because Mr. X had declared as "No claim, No return" in the auction site.

Each mediation was performed by three students – one mediator and two disputants. By analyzing mediation logs, we identified 3 characteristics of disputants – single-mindedness, selfishness and argumentativeness.

Table 2 shows the number of text responses which changed the issue points (factors). In case 6, the disputants changed the issue points only once and the mediator changed it 4 times. In case 10, the disputants changed it 12 times and the mediator changed it only 2 times.

This means that the disputant of case 6 tried to concentrate on a few issue points while the mediator acted as an *evaluative* one. The disputant of case 10 discussed as many issue points as possible while the mediator acted as a *facilitative* one.

Table 2. Character of Disputants (Single-mindedness)

case ID	1	2	3	4	5	6	7	8	9	10	11
number of text response of mediator which changed the issue points	5	4	8	5	6	4	11	4	5	2	6
number of text response of disputants which changed the issue points	6	4	6	3	6	1	8	5	7	12	9

Table 3 shows the second characteristic of disputants. The text responses of the disputant of case 6 had 7 factors all of which were favorable to him. The disputant of case 8 referred to 4 favorable factors and 5 unfavorable ones. Usually, in mediation, friendly disputants will talk about both favorable and unfavorable factors to reach agreement. Therefore, in the case of case 6, we estimate that the disputant was more selfish than the disputant of case 8 and not so friendly.

Table 3. Character of Disputants (Selfishness)

case ID	1	2	3	4	5	6	7	8	9	10	11
number of favorable factors	7	5	8	4	7	7	12	4	10	11	6
number of unfavorable factors	7	1	1	0	2	0	8	5	1	1	6

Table 4 shows the third characteristic of disputants. The text responses of the disputant of case 2 had 3 claims and 4 arguments. The disputant of case 10 referred 13 claims and 6 arguments. We estimate that the disputant of case 2 is more argumentative than the disputant of case 10 because the ratio of arguments of case 2 is higher.

Table 4. Character of Disputants (Argumentativeness)

case ID	1	2	3	4	5	6	7	8	9	10	11
number of claims	9	3	6	7	9	4	10	7	8	13	10
number of arguments	4	4	5	3	6	3	5	4	3	6	9

(2) *Similar scenario search*

Searching similar scenes is usually not easy because similar text responses can always be slightly different one from one another, as the expressions are different, and what we understand from them differs depending on previous text responses. Therefore, the traditional vector-space model which evaluates similarity between documents by counting common terms appearing in them is not appropriate here.

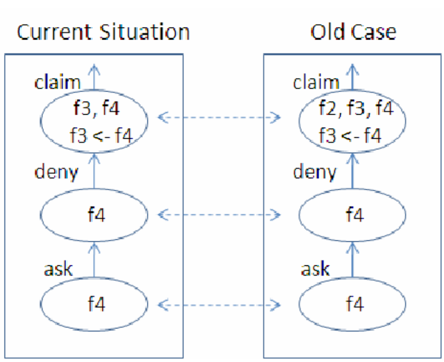


Fig. 4. Similar Scene Search

To make searching for similar scenarios easier, we propose a search method based on text response indexing [2]. This method analyses both the current text response, as well as the previous two text responses. By observing a sequence of three text responses, we can grasp the flow of issue points, which improves the accuracy of assessing the similarity of two scenes. The factors, argument formulas and the relation (link data specifiers) are compared within these three text responses and similar scenes are assessed by counting the number of common factors, common formulas and common link specifiers. In the case of Figure 4, each node denotes a text response and its contents as factors and formulas and each link denotes a link data specifier. The upper nodes in this figure share two factors, *f3* and *f4* and a formula "*f3 <- f4*." The middle nodes share a factor, *f4*. And the lower nodes share factor, *f4*. Furthermore, the first links share a link specifier, *claim*, and the second links share a link specifier, *deny*, and the third link share a link specifier, *ask*. Consequently, in this figure, two scenes share 8 common components.

3 Argument Agent

In the place of a human, the argument agent argues in the online mediation as a disputant on either party or both parties. When a disputant or the mediator inputs some text response to the online mediation system, this system extracts factors, argument formulas and a link data specifier as explained in Section 2.2. The argument agent receives this data and recognizes what issue points (factors) had been agreed upon and what issue points are still being debated. Then, it generates the text response as follows:

Firstly, it generates all possible reply candidates by referring to the case ontology. Then it selects the one to respond to by evaluating them based on a selection strategy. Finally, by referring to the case base, it selects the proper text response which corresponds to the selected reply.

3.1 Reply Candidate Generation Stage

Firstly, an argument agent lists possible mediation moves as the agent's reply from the opponent's link data by referring to the Mediation Model (Table 1). Then, the argument agent lists possible argument formula by referring to the case ontology (Figure 2). Finally, the agent generates suitable combinations of moves and formula.

For example, in Figure 2, if party X claimed *c3* (*claim(c3)*) and showed the reason "*c3 holds because c2 holds*" (e*xplain(c3, c3 <- c2)*), the agent will list the following possible reply candidates: *concede(c2)*, *deny(c2)*, *explain(^c2, ^c2 <- f21),* and *explain(^c3, ^c3 <- c4).*

3.2 Selection Stage

After generating reply candidates, the argument agent evaluates them and selects the best one.

To select the most suitable reply candidate, we propose a 3-step evaluation process. Firstly, the argument agent evaluates contents based on the issue point relationship in the case ontology and current situation. Then, the agent revises the evaluation by referring to the agent characters. Finally, the agent selects the highest scoring evaluated candidate as the content of the agent text response.

(1) Evaluation based on the case ontology

The argument agent evaluates both the argument formulas and the mediation move.

The argument formulas are evaluated by counting the number of advantageous issue points and the number of formulas which support or attack the issue point. For example, the formula "*c3 <- c2*" includes two advantageous points, c2 and c3, and there is a formula which attacks c2 ("*^c2 <- f21*") and there is a formula which supports c2 ("*c2 <- f22*"). The score of "*c3 <- c2*" is calculated as follows.

$$\mathrm{Score(c3{<}{-}c2) = Score(c3) + Score(c2) + Score(\hat{\ }c2{<}{-}f21)+Score(c2{<}f22)}$$

$$= 2+2+(-1)+1 = 4$$

Here, Score(c3) and Score(c2) are 2 because c3 and c2 are advantageous for the agent. If they are disadvantageous, their scores become -2. Score(^c2<-f21) is -1 because it attacks c2. Score(c2<-f22) is 1 because it supports c2.

The mediation move is evaluated as follows. The agent counts advantageous issue points based on the case ontology; shared and secret issue points. Also, the agent counts disadvantageous issues in the issue point list generated by the Mediation Monitor. Attempting to maintain a balance of advantageous and disadvantageous issue points, the agent tends to select the *Deny* move if the agent is in "overbalanced", and tends to select the *Concede* move when it is in "under-balanced." This means that the mediator should elicit as many shared and secret issue points as possible from the agent for a fair settlement.

(2) Revision of evaluation score by referring to personality

In order to generate variations in text response, we introduce agents which have different personalities. To revise the evaluation referring to the personality of an argument agent, the agent can generate different argument processes even if they are based on the same mediation theme.

There has been similar research in the analysis of human personalities. For example, The Big Five model of human traits describes human personality as organized along five dimensions: as extraversion/introversion, friendliness/hostility, conscientiousness/impulsiveness, emotional stability/neuroticism, and openness to experience [10]. However, results of these analyses are not directly applicable to our agent, because the relationship between these dimensions and their text response is not clear. Therefore, as we explained in Section 2.5, we analyzed the case base and devised our own attributes (dimensions). These characteristics are easier to devise from our set of mediation logs (Table 2, Table 3 and Table 4). An agent's personality type is defined on a scale of 1 to 10. The value is calculated based on the degree of each of the following:

Single-mindedness: Tendency to focus on a single topic.

Revise from the difference of issue point category between current and previous response texts. For example, in Figure 2, if the agent is single-minded, and if the opponent showed the reason "*c3 holds because c2 holds*" (*explain(c3, c3 <- c2)*), then the score value 3 is added to the score of reply candidate "*c2 doesn't hold because f21 holds*" (*explain(^c2, ^c2 <- f21)*) because *f21* belongs to the category *c2*. However, the score value isn't added to the score of reply candidate "*c3 doesn't hold because c4 holds*" (*explain(^c3, ^c3 <- c4)*) because *c3* and *c4* are different categories.

Selfishness: Tendency to take a position which is advantageous to the agent.

Revise from the advantage of issue points to the agent or not. For example, in Figure 2, if the agent is selfish, and if the opponent showed the reason "*c3 holds because c2 holds*" (*explain(c3, c3 <- c2)*), then the score value 3 is added to the score of reply candidate "*c2 doesn't hold because f21 holds*" (*explain(^c2, ^c2 <-f21)*) because it includes only advantageous issue points. On the other hand, the score value is not added to the score of reply candidate "*I will concede the reason you mentioned*" (*concede(c3 <- c2)*) because it is not advantageous to the agent.

Argumentativeness: Tendency to assert all valid arguments available.

Revise from the type of argument formula; "*P <- Q*" or "*P*." For example, in Figure 2, if the agent is argumentative, then the score value 3 is added to the score of reply candidate "*c2 doesn't hold because f21 holds*" (*explain(^c2, ^c2 <- f21)*) because it is the argument. However, the score value is not added to the score of reply candidate "*I will concede the reason you mentioned*" (*concede(c3 <- c2)*) because it is not an argument.

The following examples are selected from mediation logs which are described in Section 2.5. They show how the personality of an agent affects the reply. Here are three replies which are generated by three agents with different personalities. Mr. Y1 is very single-minded (M-agent), Mr.Y2 is very selfish (S-agent) and Mr.Y3 is very argumentative (A-agent). To observe the difference of replies, comparing whole reply texts seems to be appropriate. However, because space is limited, I cannot show a long sequence of reply texts.

Example1:

Mr. X> I will not accept the cancellation of the contract because you made a mistake over the material of the muffler.

Mediator> (To Buyer) Did you know the muffler was not made of steel?

Mr. Y1(M-agent)> No, I didn't. Whether I knew it or not is not important. Steel is standard material of the muffler.

Mr. Y2(S-agent)> Steel is the most popular material of the muffler. In the catalogue of Z company of last year, only mufflers made of steel are listed.

Mr. Y3(A-agent)> I believed the muffler was made of steel because the material was not described in the auction site.

Example2:

Mr. X> I will not cancel the contract because I had announced as "No claim, no return" in the auction site.

Mediator> (To Buyer) Do you still want to cancel the contract?

Mr. Y1(M-agent)> I want to cancel the contract and return the muffler to him.

Mr. Y2(S-agent)> "No claim, no return" declaration are legally ineffective.

Mr. Y3(A-agent)> Yes. He can't deny the cancellation because unexpected situation occurred.

Example3:

Mr. X> I cannot return the money, but I may exchange your muffler for another one which is also secondhand but which is made of steel.

Mediator> (To Buyer) How about his proposal? Do you accept it?

Mr. Y1(M-agent)> I want to cancel the contract and return the muffler to him.

Mr. Y2(S-agent)> He should return the money and make an apology.

Mr. Y3(A-agent)> No, I don't accept the proposal because he is insincere and I am afraid the quality of the muffler which he proposed is not good.

(3) Selection of the best reply

The agent evaluates all reply candidates based on the score of each reply candidate and selects one which has highest score.

3.3 Text Response Selection Stage

When the best candidate is selected, the argument agent selects a response text using the similar scene search explained in Section 2.5. Even if the content is the same, various expressions are included in a case base. Therefore, searching returns a list of reply candidates. Among them, the agent selects the best response text and uses it as the reply of the agent. For example, if the selected reply candidate is "*explain(c3, c3<- f31)*" and previous two text responses are "*claim(c3)*" and "*require-explanation(c3)*", then the argument agent searches similar scene using "*claim(c3)*", "*require-explanation(c3)*" and "*explain(c3, c3<- f31)*" , and selects the text response in the searched scene.

If the agent fails to find any similar scene, it refers to the template text which has been prepared in advance for each reply content.

4 Evaluation of Argument Agent

In the previous section, we introduced three personality attributes based on observations made from the mediation logs. The name of each personality attribute is tentative because we have not examined how mediation students feel about each characteristic. To confirm that the argument agent shows his or her intended personality in the mediation training, we substituted the buyer (Mr. Y) by the argument agent and asked the following questions to the seller (Mr. X) and the mediator after the moot mediation. In this experiment, we prepared four argument agents, M-agent, S-agent, A-agent and N-agent. We have already defined M-agent, S-agent and A-agent in Section 3.2. N-agent is a neutral agent which doesn't show a specific personality.

10 graduate school students whose major is Computer Science joined this experiment. They are separated into 5 couples, and each couple joined the moot mediation with a different type of argument agent.

Table 5 shows the average scores of these questions. The score of Q2 of M-agent and the score of Q3 of S-agent are very low. They match the personalities of M-agent and S-agent very well. According to comments by students, text responses of S-agent were appropriate and natural in most cases.

Question (0 (I don't think so at all) – 10 (I think so very much))
Q1) Are the buyer's text responses consistent?
Q2) Did the buyer change the issue point frequently?
Q3) Was the buyer cooperative to reach the settlement.
Q4) Was the buyer argumentative?
Q5) Was the buyer persuasive?

However, the scores of Q4 and Q5 show that the personality of A-agent is not recognized. To examine the reason, we counted the number of formulas given by the argument agent, and found that the number of formulas spoken by A-agent was the maximum. Nevertheless, the impression of argumentativeness of A-agent is weak. According to comments by students, we found that the main reason is that A-agent often refers the inappropriate arguments, which were ignored by the students.

The quality of response by the argument agent is affected by the quality and quantity of old cases in a case base. We used 11 cases in this experiment and they are represented using 21 factors. These numbers are not sufficient, and sometimes the argument agent fails to generate sufficient reply candidates and fails to find a proper old text response.

Table 5. Results of Questionare

	M-agent	S-agent	A-agent	N-agent
Q1) Consistency	4.9	6.1	5.2	5.7
Q2) change issues frequently	<u>3.4</u>	5.7	5.4	5.6
Q3) Cooperative	4.3	<u>3.3</u>	3.9	4.1
Q4) Argumentative	4.5	6.9	<u>3.3</u>	4.3
Q5) Persuasive	4.2	7.1	<u>5.1</u>	4.3

5 Conclusion

We presented an overview of the online mediation support system and introduced the argument agent which interacts with the system. The agent recognizes text response from the issue point, link data specifiers and argument formula, and then it makes a proper reply by comparing possible reply candidates according to the personality of the agent. By introducing our own three dimensions of personality, we showed different types of text response are made according to the personality of the argument agent.

The experiment results show that we could control the personality of an argument agent to some extent. However, the control mechanism is still naive. For example, the argument strategy is very simple because it doesn't consider the opponent's future reply nor compromising procedure. We have to improve the argument strategy and the personality control mechanism.

We should also evaluate differences in the argument process and mediation training between each agent personality type.

References

[1] Barkai, J.: Teaching Negotiation and ADR: The Savvy Samurai Meets the Devil. 75 Nebraska Law Review 704 (1996)

[2] Tanaka, T., et al.: Online Support System for Mediator Education. In: Proceedings of IJCAI 2005, Poster Session, pp. 1702–1703 (2005)

[3] Tanaka, T., Katagami, D., Nitta, K.: Advice Agent for Online Mediator Education. In: International Workshop of Agent-Based Systems for Human Learning, AAMAS 2006, pp. 43–48 (2006)

[4] Aleven, V.: Teaching Case-Based Argumentation through a Model and Examples, ph D. Thesis, University of Pittsburgh (1997)

[5] Yuasa, M., Yasumura, Y., Nitta, K.: Negotiation Support Tool Using Emotional Factors. In: Proceeding of IFSA-NAFIP 2001, pp. 2906–2911 (2001)

[6] Gordon, T.: The Pleadings Game. Kluwer Academic Publishers, Dordrecht (1995)

[7] Tanaka, T., et al.: Case Based Online Training Support System for ADR Mediator. In: International Workshop on Artificial Intelligence and Legal Education, ICAIL 2005, pp. 22–27 (2005)

[8] Tanaka, T., et al.: Case Based Utterance Generating for An Argument Agent. In: 5th Workshop of Knowledge and Reasoning in Practical Dialogue Systems, IJCAI 2007, pp. 10–14 (2007)

[9] Goldberg, L.R.: The structure of phenotypic personality traits. American Psychologist 48, 26–34 (1993)

[10] Prakken, H.: Modeling Reasoning about Evidence in Legal Procedure. In: Proc. ICAIL 2001, pp. 119–128 (2001)

[11] Tanaka, T.: Case Based Online Mediation Support System (in Japanese), ph D. thesis, Tokyo Institute of Technology (2006)

Assumption-Based Argumentation for Closed and Consistent Defeasible Reasoning

Francesca Toni

Imperial College London, Department of Computing

Abstract. Assumption-based argumentation is a concrete but general-purpose argumentation framework that has been shown, in particular, to generalise several existing mechanisms for non-monotonic reasoning, and is equipped with a computational counterpart and an implemented system. It can thus serve as a computational tool for argumentation-based reasoning, and for automatising the process of finding solutions to problems that can be understood in assumption-based argumentation terms. In this paper we consider the problem of reasoning with defeasible and strict rules, for example as required in a legal setting. We provide a mapping of defeasible reasoning into assumption-based argumentation, and show that the framework obtained has properties of closedness and consistency, that have been advocated elsewhere as important for defeasible reasoning in the presence of strict rules. Whereas other argumentation approaches have been proven closed and consistent under *some* specific semantics, we prove that assumption-based argumentation is closed and consistent under *all* argumentation semantics.

1 Introduction

Argumentation has proven to be a useful abstraction mechanism for understanding several AI problems. In particular, several mechanisms for defeasible, non-monotonic reasoning have been proven to be instances of argumentation frameworks [5,3], and defeasible logic can be understood in argumentation terms [11]. Moreover, argumentation has been extensively applied in a legal setting (see [18,14,19]), that require defeasible reasoning. In this paper we consider a simple form of defeasible reasoning, whereby defeasible and strict rules may be combined to support conclusions. Informally, the strict rules are conflict-free and unquestionable, and should always be taken into account. Instead, the defeasible rules (together with the strict rules) may give rise to conflicts, and should be picked and chosen when trying to support conclusions so as to avoid conflicts. The emergence of conflicts provides a "defeat" against choosing certain defeasible rules. This form of defeasible reasoning is a simplified version of the one given by Nute's defeasible logic [15] and most later frameworks for defeasible reasoning which also include defeaters and preferences over rules.

K. Satoh et al. (Eds.): JSAI 2007, LNAI 4914, pp. 390–402, 2008.

Argumentation lends itself well to modelling defeasible reasoning, and indeed many approaches to defeasible reasoning are equipped with an argumentation-based semantics (e.g. see [18,11,9,13]). [4] have proposed two principles to which any argumentative formulation of defeasible reasoning should obey. The first principle states that the set of arguments and justified conclusions sanctioned as acceptable by any given argumentative formalisation of defeasible reasoning should be "closed", namely take all strict rules into account, by making sure that their conclusions are justified if their premises are. The second principle states that acceptable arguments and conclusions should be "consistent", namely they should not give rise to any conflicts. Caminada and Amgoud [4] demonstrate that many of the existing approaches to defeasible reasoning via argumentation do not obey these principles [1].

In this paper we provide an argumentative formulation for the problem of reasoning with defeasible and strict rules and show that it obeys the principles of closedness and consistency proposed in [4]. The proposed formulation makes use of an existing general-purpose framework for argumentation, referred to as assumption-based argumentation [3,14], that has been shown, in particular, to generalise several existing mechanisms for non-monotonic reasoning, by admitting them as concrete instances. Assumption-based argumentation can be seen as an instance of abstract argumentation [5] providing the "building blocks" for arguments and the attack relationship, that are seen as primitive notions in abstract argumentation but are derived from the notions of assumption, backward deduction and contrary of assumptions in assumption-based argumentation. Assumption-based argumentation is also equipped with a computational counterpart [6,7]. It can thus serve as an effective computational tool for argumentation-based reasoning, and for automatising the process of finding solutions to AI problems that can be understood in assumption-based argumentation terms. In assumption-based argumentation, a set of assumptions stands for a set of arguments and all the conclusions they justify. Some of the computational advantages of reasoning with assumptions rather than full arguments are outlined in [6]. In particular, recomputation of overlappings amongst arguments is not required in assumption-based argumentation. The results in this paper identify some further advantages in terms of formal properties. Note that the formulation of defeasible reasoning within assumption-based argumentation requires a generalisation of conventional assumption-based argumentation. We foresee that this generalisation will widen the applicability of the framework.

The paper is organised as follows. First we give some background on abstract argumentation and assumption-based argumentation. Then, we define the framework for defeasible reasoning we consider in this paper and the general principles adapted from [4] and we give the required generalisation of assumption-based argumentation. Further, we provide a mapping from defeasible reasoning onto generalised assumption-based argumentation and prove our formal results. Finally, we conclude, identifying in particular directions for future work.

[1] However, note that these principles are fulfilled by conventional approaches to defeasible reasoning, such as [16].

2 Abstract and Assumption-Based Argumentation

An *abstract argumentation framework* is a pair $(Arg, attacks)$ where Arg is a finite set, whose elements are referred to as *arguments*, and $attacks \subseteq Arg \times Arg$ is a binary relation over Arg. Given sets $X, Y \subseteq Arg$ of arguments, X *attacks* Y iff there exists $x \in X$ and $y \in Y$ such that $(x, y) \in attack$.

Given an abstract argumentation framework, several notions of *acceptable sets of arguments* can be defined [5]. A set X of arguments is

- *conflict-free* iff it does not attack itself
- *admissible* iff X is conflict-free and X attacks every argument Y such that Y attacks X;
- *preferred* iff X is maximally (wrt set inclusion) admissible;
- *complete* iff X is admissible and X contains all arguments x such that X attacks all attacks against x;
- *grounded* iff X is minimally (wrt set inclusion) complete;
- *ideal* iff X is admissible and it is contained in every preferred set of arguments.

The last notion was not in the original [5], but has been proposed recently [7] as an alternative, less sceptical semantics than the grounded semantics.

Most instances of abstract argumentation are equipped with a notion of *sets of justified conclusions*, namely those supported by the chosen acceptable sets of arguments. Each such set provides a possible *output*. In the case of a sceptical semantics (such as that given by grounded and ideal sets of arguments), only one acceptable set of arguments and corresponding output exist. Instead, for credulous semantics (such as that given by admissible sets of arguments) multiple outputs exist. These outputs will typically be in conflict (except when they are not maximal, as in the case of admissible sets).

The abstract view of argumentation does not deal with the problem of actually finding arguments and attacks amongst them. Typically, arguments are built by joining rules in the belief set of the proponent of arguments, and attacks arise from conflicts amongst such arguments. In assumption-based argumentation, arguments are obtained by reasoning backwards with a given set of inference rules (the "beliefs") from conclusions to assumptions, and attacks are defined in terms of a notion of "contrary" of assumptions. Concretely, assumption-based argumentation frameworks [3,6,7] are instances of abstract argumentation frameworks where arguments in Arg are defined as *backward deductions* from *assumptions* in an underlying logic, viewed as a *deductive system*, and where *attack* is defined in terms of a notion of *contrary*.

Definition 1. *A* deductive system *is a pair* $(\mathcal{L}, \mathcal{R})$ *where*

- $\mathcal{L}$ *is a formal language consisting of countably many sentences, and*
- $\mathcal{R}$ *is a countable set of inference rules of the form* $\alpha \leftarrow \alpha_1, \ldots, \alpha_n$
  $\alpha \in \mathcal{L}$ *is called the* conclusion *of the inference rule,* $\alpha_1, \ldots, \alpha_n \in \mathcal{L}$ *are called the* premises *of the inference rule and* $n \geq 0$.

If $n = 0$, then the inference rule represents an axiom (written simply as α). A deductive system does not distinguish between domain-independent axioms/rules, which belong to the specification of the logic, and domain-dependent axioms/rules, which represent a background theory.

Definition 2. *Given a deductive system* $(\mathcal{L}, \mathcal{R})$ *and a selection function* f*, a* (backward) deduction *of a conclusion* α *based on (or supported by) a set of premises* P *is a sequence of sets* $S_1, \ldots, S_m$*, where* $S_1 = \{\alpha\}$*,* $S_m = P$*, and for every* $1 \leq i < m$*, where* σ *is the sentence occurrence in* S_i *selected by* f*:*

1. *If* σ *is not in* P *then* $S_{i+1} = S_i - \{\sigma\} \cup S$ *for some inference rule of the form* $\sigma \leftarrow S$ *in the set of inference rules* $\mathcal{R}$*;*
2. *If* σ *is in* P *then* $S_{i+1} = S_i$*.*

As an example, consider $\mathcal{R} = \{a \leftarrow b, c; a \leftarrow b, e; c \leftarrow f; b \leftarrow g\}$. Then, the following is a (backward) deduction of a supported by premises $\{f, g\}$, with left-most selection function: $\{a\}$, $\{b, c\}$, $\{g, c\}$, $\{g, f\}$.

Deductions are the basis for the construction of arguments in assumption-based argumentation, but to obtain an argument from a backward deduction we restrict the premises to ones that are *assumptions*. Moreover, to specify when one argument attacks another, we need to specify *contraries* of assumptions.

Definition 3. *An* assumption-based argumentation framework *is a tuple*

$$\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \overline{} \rangle$$

where

- $(\mathcal{L}, \mathcal{R})$ *is a deductive system;*
- $\mathcal{A} \subseteq \mathcal{L}$*,* $\mathcal{A} \neq \{\}$ *(*$\mathcal{A}$ *is the set of* assumptions*);*
- *if* $\alpha \in \mathcal{A}$*, there is no inference rule of the form* $\alpha \leftarrow \alpha_1, \ldots, \alpha_n \in \mathcal{R}$*;*
- $\overline{}$ *is a (total) mapping from* $\mathcal{A}$ *into* $\mathcal{L}$*.* $\overline{\alpha}$ *is the* contrary *of* α*.*

Note that, by the third bullet, following [6] we restrict ourselves to *flat* frameworks [3], whose assumptions do not occur as conclusions of inference rules. Flat frameworks are restricted but still interesting and general, as, for example, they admit default logic and logic programming as concrete instances (see [3]). Concretely, in the instance of assumption-based frameworks for logic programming, the inference rules are the clauses in the logic program, the assumptions are all the negation-as-failure literals, and the contrary of any assumption of the form *not p* is p [2]: since no negation-as-failure literal may occur in the conclusion of a clause, the resulting assumption-based framework is flat.

In the assumption-based approach to argumentation, arguments are deductions to conclusions, based upon assumptions, and the attack relationship between arguments depends solely on sets of assumptions and the notion of contrary: an argument attacks another if the first argument supports a deduction for the contrary of an assumption in the support of the second argument. Formally:

[2] Please note that the notion of contrary needs to be given solely for assumptions.

Definition 4. *An* argument $A \vdash \alpha$ *is a deduction of* $\alpha \in \mathcal{L}$ *whose premises* A *are all assumptions (in* $\mathcal{A}$*). Below, we will use* $A \vdash \alpha$ *also as a shorthand for "there exists a deduction for* α *supported by* A*".*

Definition 5.
- *An* argument $A \vdash \alpha$ attacks an argument $B \vdash \beta$ *iff* $A \vdash \alpha$ *attacks an assumption in* B*.* [3]
- *An* argument $A \vdash \alpha$ attacks an assumption β *iff* α *is the contrary* $\overline{\beta}$ *of* β*.*
- *A* set of assumptions A attacks a set of assumptions B *iff there exists an assumption* $\alpha \in B$ *and an argument* $A' \vdash \overline{\alpha}$ *such that* $A' \subseteq A$*.*

Note that in this approach to argumentation, the attack relationship between arguments depends solely on sets of assumptions. In some other approaches, however, such as those of Pollock [17] and Prakken and Sartor [19], an argument can attack another argument by contradicting its conclusion. As discussed and illustrated in [14,6], such "rebuttal" attacks can be reduced to "undermining" attacks against supporting assumptions, by adding explicit assumptions to the premises of inference rules so that undermining the additional assumption is equivalent to rebutting the conclusion of the inference rule.

A set of arguments univocally identifies a set of assumptions (the union of all supports of the arguments). Conversely, a set of assumptions stands for the set of all arguments whose premises are contained in the given set of assumptions. Thus, the computation of acceptable sets of arguments and the corresponding outputs amounts to computing acceptable sets of assumptions:

Definition 6. *A set of assumptions* A *is*

- conflict-free *iff* A *does not attack itself*
- admissible *iff* A *conflict-free and* A *attacks every set of assumptions* B *that attacks* A
- preferred *iff it is maximally admissible*
- complete *iff it is admissible and contains all assumptions* x *such that* A *attacks all attacks against* $\{x\}$
- grounded *iff it is minimally complete*
- ideal *iff* A *is admissible and it is contained in every preferred set of assumptions.*

Definition 7. *Given one of the notions in definition 6,*

- *An* acceptable set of arguments *is a set of arguments such that the union of all premises of these arguments is sanctioned by the chosen notion.*
- *Given an acceptable set of arguments Acc, let* X *be the union of all sets of assumptions serving as premises of arguments in Acc. Then, the* output *corresponding to Acc is the set of all sentences in* $\mathcal{L}$ *for which there is an argument with premises a subset of* X*.*

[3] As discussed in [6], here the attack relationship between arguments depends solely on sets of assumptions. In some other approaches, however, an argument can attack another argument by contradicting its conclusion. Such "rebuttal" attacks can be reduced to "undermining" attacks, as described in [14].

3 Reasoning with Defeasible and Strict Rules

Defeasible reasoning may be performed within a framework consisting of defeasible and strict rules and facts [15]. Defeasible rules represent subjective or uncertain information, whereas strict rules represent objective or certain information. Defeasible facts may represent potential assumptions, whereas strict facts may represent observations. Before defining our adopted frameworks for defeasible reasoning, we give some preliminary notions.

Definition 8. *A* language $\mathcal{L}_d$ *is a set of ground literals, which can be atoms* A *or negations of atoms* $\neg A$*.*

In the remainder of this paper, given a literal L, with an abuse of notation, $\neg L$ will stand for the complement of L, namely $\neg L$ if L is an atom, and A if L is $\neg A$ (with A an atom).

Definition 9. *Given a language* $\mathcal{L}_d$*, a* rule *is of the form* $B_1, \ldots, B_n \rightarrow B_0$ *where* $B_0, \ldots, B_n$ *are literals in* $\mathcal{L}_d$ *and* $n \geq 0$*.* B_0 *is referred to as the* conclusion *and* $B_1, \ldots, B_n$ *as the* premises *of the rule. When* $n = 0$ *the rule is also referred to as a* fact*.*

Definition 10. *Let* $\mathcal{L}_d$ *be a language. A* defeasible framework *is a pair* $\langle D, S \rangle$*, with* D *and* S *sets of rules wrt* $\mathcal{L}_d$ *such that* $D \neq \{\}$*. We refer to* D *as the* defeasible *and* S *as the* strict *components, and to rules in* D *(*S*) as defeasible (strict, respectively).*

Note that we see facts as a special kind of rules. Also, other approaches in the literature, starting from [15], equivalently distinguish defeasible and strict rules by using different implications symbols. Moreover, some approaches allow a third kind or rules, called defeaters [15,11]. Further, some existing approaches, e.g. [18,9], allow negation as failure in both strict and defeasible rules. Finally, many approaches also allow for preferences amongst defeasible rules to be expressed explicitly (e.g. [18]) or taken into account (e.g. via specificity, [15,9]) within a defeasible framework. Until the conclusions we will ignore these extended features of defeasible reasoning.

Intuitively, defeasible rules may or may not be applied by a rational reasoner, whereas strict rules need to be always applied. A rational reasoner needs to avoid conflicts in its chosen lines of reasoning. Conflicts in defeasible frameworks arise from "deriving" complementary conclusions from sets of chosen strict and defeasible rules, of the form L and $\neg L$. As strict rules cannot be disregarded ever, it is reasonable to assume that conflicts cannot arise by applying strict rules only. Conflicts will however typically arise from defeasible rules. The semantics of defeasible frameworks needs to resolve these conflicts and identify suitable "acceptable subsets of the set of defeasible rules" and corresponding sets of "supported conclusions". These notions can be defined by means of argumentation (several examples exist, e.g. [18,9,4]), in terms of acceptable sets of arguments. [4] suggests that, when argumentation is used for defeasible reasoning, the properties of "closedness" and "consistency" should hold for the notion of acceptable set of arguments and the corresponding output. These properties can be defined as follows:

Definition 11. *Let* $(Arg, attacks)$ *be an argumentation framework for a defeasible framework* $\langle D, S\rangle$ *wrt* $\mathcal{L}_d$*, Acc an acceptable set of arguments wrt the framrwork* $(Arg, attacks)$*, Conc the set of all conclusions of all arguments in Acc and Out the output corresponding to Acc:*

- *Acc is* closed *iff for each strict rule* $X \rightarrow Y \in S$*, if* $X \subseteq Conc$ *then* $Y \in Conc$ [4]*;*
- *Out is* closed *iff for each strict rule* $X \rightarrow Y \in S$*, if* $X \subseteq Out$ *then* $Y \in Out$*;*
- *Acc is* consistent *iff for no* $L \in \mathcal{L}_d$*, both* $L \in Conc$ *and* $\neg L \in Conc$*;*
- *Out is* consistent *iff for no* $L \in \mathcal{L}_d$*, both* $L \in Out$ *and* $\neg L \in Out$*.*

[4] points out that most existing argumentation-based approaches to defeasible reasoning (e.g. [18,9]) are not guaranteed to exhibit the features of closedness and consistency, either at the level of acceptable sets of arguments or output. We provide an argumentation framework for defeasible reasoning that is closed and consistent at all levels. This framework is an instance of a generalised form of assumption-based argumentation that we define next.

4 Generalised Assumption-Based Argumentation

Definition 12. *A* generalised assumption-based framework *is a tuple*

$$\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C}\rangle$$

with $\mathcal{L}$*,* $\mathcal{R}$ *and* $\mathcal{A}$ *defined as for conventional assumption-based frameworks in definition 3, and* $\mathcal{C}$ *is a non-empty set of pairs of sets of sentences in* $\mathcal{L}$ *such that, for each pair* $(X, Y) \in \mathcal{C}$*:*

- X *and* Y *are both non-empty,*
- X *contains at least one assumption in* $\mathcal{A}$ *(*X *is referred to as* retractible*).*

Basically, each element (X, Y) of $\mathcal{C}$ represents a combination $X \cup Y$ of sentences that cannot hold together, with the retractible set X the designated "culprit" to be "withdrawn" should such a combination arise. Conventional assumption-based frameworks are generalised assumption-based frameworks where the set $\mathcal{C}$ is $\{(\{\alpha\},\{\overline{\alpha}\}) \mid \alpha \in \mathcal{A}\}$. Generalised frameworks are a variant of the frameworks adopted in [20].

Definition 13. *Given a generalised framework, a* set of assumptions A attacks a set of assumptions B *iff there exists* $\alpha \in X$ *for some* $(X, Y) \in \mathcal{C}$ *such that*

- $\alpha \in B$*, and*
- *if* $X - \{\alpha\} \neq \{\}$*, there exists an argument* $B' \vdash x$ *with* $B' \subseteq B$ *for each* $x \in X - \{\alpha\}$*, and*
- *there exists* $A' \vdash y$ *with* $A' \subseteq A$ *for every* $y \in Y$*.*

[4] With an abuse of notation, the premises of rules are seen here as sets.

Intuitively, A and B together derive all elements of X and Y (and thus conflict), and B is to blaim, by being responsible for deriving the elements of the "culprits" in X. Note that if X and Y are singleton sets then this notion of attack amounts to the notion of attack for conventional assumption-based frameworks. Note also that, if $X \subseteq \mathcal{A}$ in all pairs in $\mathcal{C}$, then A attacks B iff there exists $(X, Y) \in \mathcal{C}$ such that $X \subseteq B$ and there exists $A' \vdash y$ with $A' \subseteq A$ for every $y \in Y$. The generalised framework for defeasible reasoning that we will provide next is indeed such that $X \subseteq \mathcal{A}$ in all pairs in $\mathcal{C}$.

Generalised assumption-based argumentation frameworks maintain the same features of conventional assumption-based argumentation frameworks in that (1) they still rely upon assumptions and backward deductions as the building blocks for arguments, (2) the notion of attack can be computed backwards, supported by assumptions (precisely any of the assumptions in the retractible set in any pair in $\mathcal{C}$), and (3) the notions of acceptable set of arguments and corresponding outputs are borrowed from abstract argumentation.

5 Generalised Assumption-Based Argumentation for Defeasible Reasoning

In order to define the specific instance of generalised assumption-based frameworks for capturing defeasible reasoning, we will make use of the following notion.

Definition 14. *Given a sentence s, a deductive system $(\mathcal{L}, \mathcal{R})$ and a set of sentences $\mathcal{A}$, $reduct(s, \mathcal{R}, \mathcal{A}) = \{\Delta \subseteq \mathcal{A} | \Delta \vdash s\}$, with $\Delta \vdash s$ standing for: there exists a backward deduction of s whose premises are Δ, given $(\mathcal{L}, \mathcal{R})$.*

Example 1. Let $\mathcal{R} = \{p \leftarrow q, r; q \leftarrow a, b; r \leftarrow c; r \leftarrow d\}$ and $\mathcal{A} = \{a, b, c, d\}$. Then, $reduct(p, \mathcal{R}, \mathcal{A}) = \{\{a, b, c\}, \{a, b, d\}\}$.

Definition 15. *The assumption-based framework corresponding to a defeasible framework $\langle D, S\rangle$ wrt $\mathcal{L}_d$ is $\delta = \langle \mathcal{L}_\delta, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta\rangle$ whereby*

- *$\mathcal{A}_\delta$ is a set of literals not already in $\mathcal{L}_d$ such that there exists a bijective mapping asm from rules in D into $\mathcal{A}_\delta$;*
- *$\mathcal{L}_\delta = \mathcal{L}_d \cup \mathcal{A}_\delta$;*
- *$\mathcal{R}_\delta = \{X \leftarrow Y, asm(Y \rightarrow X) | Y \rightarrow X \in D\} \cup S$*
- *$\mathcal{C}_\delta = \{(K, \{\neg L\}) | K \in reduct(L, \mathcal{R}_\delta, \mathcal{A}_\delta), L \in \mathcal{L}_d$ and both L and $\neg L$ occur as conclusions of some rule in $S \cup D\}$,*

Basically, we associate an assumption to each defeasible rule in D, exactly as done in Poole's Theorist and other abductive approaches. Moreover, we reduce the conflict amongst complementary literals to the assumptions corresponding to the defeasible rules that may occur in any "relevant" reasoning lines giving rise to the conflict.

Note that the construction of *reduct* will terminate only if there are no "loops" in the inference rules. For example, given the set of inference rules $\mathcal{R} = \{p \leftarrow q, a; q \leftarrow q\}$, $reduct(p, \mathcal{R}, \{a\})$ cannot be computed finitely. We will assume below that there are no "loops" in the inference rules. Moeover, we will assume that $\langle D, S\rangle$ has at least two conflicting rules and that conflicts always involve defeasible rules on "both sides" (namely for reasoning to both L and $\neg L$ in the conflict), as otherwise, if for some L, L can be derived by strict rules only, it would be natural to adopt L instead of a defeasibly derived $\neg L$. By making these assumptions we can prove the following result:

Theorem 1. *The assumption-based framework* $\delta{=}\langle \mathcal{L}_\delta, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta\rangle$ *corresponding to a defeasible framework* $\langle D, S\rangle$ *wrt* $\mathcal{L}_d$ *is well-defined.*

Proof. Assume that $\langle \mathcal{L}_\delta, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta\rangle$ is not well-defined. This can only mean that $\mathcal{C}_\delta$ is empty or one of $(\{\}, Y)$ or $(X, \{\})$ belongs to $\mathcal{C}_\delta$. However, note that $\mathcal{C}_\delta$ may be empty only if no rules with conflicting conclusions exist in $D \cup S$. Moreover, $(\{\}, Y)$ or $(X, \{\})$ may belong to $\mathcal{C}_\delta$ only if there exist some literal $L \in \mathcal{L}_d$ such that there is a reasoning line for L supported solely by strict rules. But we have assumed that this cannot be the case. Moreover, by construction, each retractible set in $\mathcal{C}_\delta$ consists solely of assumptions. Thus, $\langle \mathcal{L}_\delta, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta\rangle$ is well-defined.

The following example illustrates the mapping from $\langle D, S\rangle$ to $\langle \mathcal{L}_\delta, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta\rangle$.

Example 2. Given $\langle D, S\rangle$ with $S = \{\}$ and $D = \{q; q \rightarrow p; r; r \rightarrow \neg p\}$, $\mathcal{R}_\delta$ may be $\{q \leftarrow w; p \leftarrow q, x; r \leftarrow y; \neg p \leftarrow r, z\}$, $\mathcal{A}_\delta = \{w, x, y, z\}$, and $\mathcal{C}_\delta = \{(\{w, x\}, \{\neg p\}), (\{y, z\}, \{p\})\}$.

Given $\langle D, S\rangle$ with $S = \{a; b; c; d\}$ and $D = \{q; q, r \rightarrow p; a \rightarrow r; b \rightarrow r; c \rightarrow \neg p; d \rightarrow \neg p\}$, $\mathcal{R}_\delta$ may be $\{q \leftarrow v; p \leftarrow q, r, u; r \leftarrow a, w; r \leftarrow b, x; \neg p \leftarrow c, y; \neg p \leftarrow d, z\}$, $\mathcal{A}_\delta = \{u, v, w, x, y, z\}$, and $\mathcal{C}_\delta = \{\ (\{u, v\}, \{\neg p\}), (\{u, x\}, \{\neg p\}), (\{y\}, \{p\}), (\{z\}, \{p\})\ \}$.

Below, we will assume a defeasible framework $\langle D, S\rangle$ wrt $\mathcal{L}_d$ and the corresponding assumption-based argumentation framework $\langle \mathcal{L}_\delta, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta\rangle$.

The mapping into assumption-based argumentation automatically provides notions of acceptable sets of arguments and corresponding output, by instantiating the notions in definition 7.

Theorem 2. *Let Acc be any acceptable set of arguments wrt* $\langle \mathcal{L}_\delta, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta\rangle$ *and Out the corresponding output. Acc and Out are closed.*[5]

Then, let us observe that no "inconsistency" may be obtained from conflict-free (see definition 6) sets of assumptions in $\langle \mathcal{L}_\delta, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta\rangle$:

Lemma 1. *Let* $\Delta \subseteq \mathcal{A}_\delta$ *be conflict-free. Then, there exists no* $L \in \mathcal{L}_\delta$ *and no* $\Delta', \Delta'' \subseteq \Delta$ *such that* $\Delta' \vdash_\delta L$ *and* $\Delta'' \vdash_\delta \neg L$.

[5] All proofs are omitted for lack of space, but can be found in an accompanying forthcoming technical report.

Then, the following result holds:

Theorem 3. *Let Acc be any conflict-free set of arguments wrt* $\langle \mathcal{L}_\delta, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta \rangle$ *and Out the corresponding output. Both Acc and Out are consistent.*

Since every set of assumptions acceptable according to any notion in definition 6 is conflict-free, the following is a direct corollary of theorems 2 and 3.

Corollary 1. *Let Acc be any acceptable set of arguments wrt* $\langle \mathcal{L}_\delta, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta \rangle$ *according to any notion in definition 6 and Out the corresponding output. Acc and Out are closed and consistent.*

Below we illustrate our approach and results using (variations of) some examples from [4].

Example 3 (example 2 in [4]). Let $\langle D, S \rangle$ with $S = \{b \rightarrow \neg w; m \rightarrow w; r; o\}$ and $D = \{r \rightarrow m; o \rightarrow b\}$. The corresponding $\langle \mathcal{L}_\delta, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta \rangle$ has

$\mathcal{R}_\delta = \{\neg w \leftarrow b; w \leftarrow m; r; o; m \leftarrow r, \alpha; b \leftarrow o, \beta\}$,
$\mathcal{A}_\delta = \{\alpha, \beta\}$, and
$\mathcal{C}_\delta = \{(\{\alpha\}, \{\neg w\}), (\{\beta\}, \{w\})\}$,

obtained as follows: the only conflicting literals occurring as conclusions of rules are w and $\neg w$, and

- $reduct(w, \mathcal{R}_\delta, \{\alpha, \beta\}) = \{\{\alpha\}\}$
- $reduct(\neg w, \mathcal{R}_\delta, \{\alpha, \beta\}) = \{\{\beta\}\}$.

There are two maximally admissible sets of assumptions, $\{\alpha\}$ and $\{\beta\}$, with corresponding outputs $\{r, o, m, w\}$ and $\{r, o, b, \neg w\}$ (respectively), and one grounded set of assumptions, $\{\}$, with output $\{r, o\}$: all are consistent (and closed).

Example 4 (adaptation of example 3 in [4]). Let $\langle D, S \rangle$ with $S = \{b, c, e, f \rightarrow \neg g; a; d\}$ and $D = \{a \rightarrow b; b \rightarrow c; d \rightarrow e; e \rightarrow f; g\}$. [6] The corresponding assumption-based framework $\langle \mathcal{L}_\delta, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta \rangle$ has

$\mathcal{R}_\delta = \{b \leftarrow a, \alpha; c \leftarrow b, \beta; e \leftarrow d, \gamma; f \leftarrow e, \delta; g \leftarrow \epsilon; \neg g \leftarrow b, c, e, f; a; d\}$,
$\mathcal{A}_\delta = \{\alpha, \beta, \gamma, \delta, \epsilon\}$, and
$\mathcal{C}_\delta = \{(\{\alpha, \beta, \gamma, \delta\}, \{g\}), (\{\epsilon\}, \{\neg g\})\}$,

obtained as follows: the only conflicting literals occurring as conclusions of rules are g and $\neg g$, and

- $reduct(g, \mathcal{R}_\delta, \mathcal{A}_\delta) = \{\{\alpha, \beta, \gamma, \delta\}\}$
- $reduct(\neg g, \mathcal{R}_\delta, \mathcal{A}_\delta) = \{\{\epsilon\}\}$.

[6] Note that in [4] g (and a and d) was assumed to be in a component $\mathcal{K}$ of defeasible theories, holding facts. In our defeasible frameworks, facts can be either defeasible or strict (in D or S, respectively). In our formalisation of this example, we treat g as defeasible.

Then, the $\{\epsilon\}$ attacks $\{\alpha, \beta, \gamma, \delta\}$ which attacks it back. There are four preferred (and maximally conflict-free) sets of assumptions: $\{\alpha, \beta, \gamma, \epsilon\}$, $\{\alpha, \gamma, \delta, \epsilon\}$, $\{\beta, \gamma, \delta, \epsilon\}$, $\{\alpha, \beta, \gamma, \delta\}$, with outputs $\{b, c, e, g\}$, $\{b, e, f, g\}$, $\{c, e, f, g\}$ and $\{\neg g\}$ (respectively), all closed. There is one grounded set of assumptions, $\{\}$, with output $\{a, d\}$, also closed.

In our approach to defeasible reasoning, it is also easy to see that every argument in an acceptable set is necessarily consistent, in the following sense.

Theorem 4. *Let Acc be any acceptable set of arguments wrt* $\langle \mathcal{L}_\delta, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta \rangle$*. For no argument* $\Delta \vdash_\delta x$ *in Acc there exist* $L, \neg L \in \mathcal{L}_d$ *and sets* $\Delta', \Delta'' \subseteq \Delta$ *such that* $\Delta' \vdash_\delta L$ *and* $\Delta'' \vdash_\delta \neg L$*.*

As a consequence, in assumption-based argumentation we do not force/need to check that arguments are consistent, as some approaches do (e.g. [9]), since arguments in acceptable sets in assumption-based argumentation are guaranteed to be consistent, as a side effect. Other approaches instead need to check that all arguments are consistent when they compute acceptable sets, at an extra cost.

6 Conclusions

We have proposed a formulation of reasoning with defeasible rules and facts, in the presence of conflicts, within the framework of assumption-based argumentation. The formulation has required a generalisation of the conventional framework for assumption-based argumentation whereby contraries of individual assumptions are replaced by conflicting sets, with assumptions specified as retractible in case a conflict arises. We have proven that the given formulation obeys principles of closedness and consistency under *all* argumentation semantics, identified in [4] as important for argumentative defeasible reasoning, and proven to hold there only for *some* specific semantics.

For simplicity, in this paper we have focused our attention to defeasible reasoning with strict and defeasible rules and facts only. In particular, we have ignored defeaters and preferences over rules and/or arguments, and other more sophisticated forms of defeasible reasoning [12,10,16,2]. As a consequence of this representation choice, the notion of attack obtained in the framework we have proposed is symmetric. Some existing argumentative approaches adopt preference mechanisms to break this symmetry in attacks and to fulfil the demands of many applications, e.g. in the legal setting. For example, [9] uses a mechanism of specificity to prefer some arguments over some others, and [18,13] use user-defined preferences over rules. In the future, we plan to consider more-sophisticated forms of defeasible reasoning and, in particular, consider adding preferences over rules to our framework and propose a formulation in generalised assumption-based argumentation, e.g. following the one suggested in [14], by reducing preference-defining rules to defeaters against the less preferred rules. This approach is also followed by [1].

We have focused on a specific form of conflict in this paper, namely that arising amongst complementary literals (of the form L and $\neg L$). However, our approach applies also with more general forms of conflicts, for example, given rules:

$\{gunshot;\ gunshot \rightarrow crime-of-manslaughter;$
$child;\ child \rightarrow not-prosecuted\}$

the conflict may be amongst $crime-of-manslaughter$ and $not-prosecuted.$

Some approaches (e.g. [9]) also consider negation as failure literals in the premises of rules, as in extended logic programming. Negation as failure can be dealt with within assumption-based argumentation [3], by appropriate assumptions and contraries. Thus, the extension of our approach to include negation as failure should be possible and straightforward.

Finally, the approach we presented could benefit from the computational mechanisms that assumption-based argumentation is equipped with [6,7,8], if these mechanisms were extended to deal with the generalisation of the framework we have proposed in this paper. This is left for future work.

Acknowledgements

The author has been supported by a UK Royal Academy of Engineering/ Leverhulme Trust Senior Fellowship and by the Sixth Framework IST programme of the EC, under the 035200 ARGUGRID project. The author wishes to thanks Carlos Chesñevar for useful discussions leading to this paper, and the anonymous referees for helpful suggestions.

References

1. Antoniou, G., et al.: Representation results for defeasible logic. ACM Transactions on Computational Logic 2(2), 255–287 (2001)
2. Billington, D., Rock, A.: Propositional plausible logic: Introduction and implementation. Studia Logica 67(2), 243–269 (2001)
3. Bondarenko, A., et al.: An abstract, argumentation-theoretic framework for default reasoning. Artificial Intelligence 93(1-2), 63–101 (1997)
4. Caminada, M., Amgoud, L.: An axiomatic account of formal argumentation. In: Proc. AAAI (2005)
5. Dung, P.M.: The acceptability of arguments and its fundamental role in nonmonotonic reasoning and logic programming and n-person game. Artificial Intelligence 77, 321–357 (1995)
6. Dung, P., Kowalski, R., Toni, F.: Dialectic proof procedures for assumption-based, admissible argumentation. Artificial Intelligence 170, 114–159 (2006)
7. Dung, P., Mancarella, P., Toni, F.: A dialectic procedure for sceptical, assumption-based argumentation. In: 1st International Conference on Computational Models of Argument (COMMA 2006) (September 2006)
8. Dung, P., Mancarella, P., Toni, F.: Computing ideal sceptical argumentation. Artificial Intelligence, Special Issue on Argumentation in Artificial Intelligence (to appear, 2007)
9. Garcia, A., Simari, G.: Defeasible logic programming: An argumentative approach. Journal of Theory and Practice of Logic Prog. 4(1-2), 95–138 (2004)

10. Governatori, G., et al.: A formal approach to protocols and strategies for (legal) negotiation. In: Procedings of the 8th International Conference on Artificial Intelligence and Law, pp. 168–177 (2001)
11. Governatori, G., et al.: Argumentation semantics for defeasible logics. Journal of Logic and Computation 14(5), 675–702 (2004)
12. Grosof, B.N.: Prioritized conflict handling for logic programs. In: Proc. Int. Logic Programming Symposium, pp. 197–211 (1996)
13. Kakas, A.C., Moraitis, P.: Argumentation based decision making for autonomous agents. In: AAMAS, pp. 883–890 (2003)
14. Kowalski, R.A., Toni, F.: Abstract argumentation. Journal of Artificial Intelligence and Law, Special Issue on Logical Models of Argumentation 4(3-4), 275–296 (1996)
15. Nute, D.: Defeasible reasoning. In: Fetzer, J.H. (ed.) Aspects of Artificial Intelligence, pp. 251–288. Kluwer Academic Publishers, Dordrecht (1987)
16. Nute, D.: Apparent obligation. In: Nute, D. (ed.) Defeasible Deontic Logic, pp. 287–316. Kluwer Academic Publishers, Dordrecht (1997)
17. Pollock, J.: Defeasible reasoning. Cognitive Science 11(4), 481–518 (1987)
18. Prakken, H., Sartor, G.: Argument-based extended logic programming with defeasible priorities. Journal of Applied Non-Classical Logics 7(1), 25–75 (1997)
19. Prakken, H., Sartor, G.: The role of logic in computational models of legal argument: A critical survey. In: Kakas, A.C., Sadri, F. (eds.) Computational Logic: Logic Programming and Beyond. LNCS (LNAI), vol. 2408, pp. 342–381. Springer, Heidelberg (2002)
20. Toni, F., Kowalski, R.A.: Reduction of abductive logic programs to normal logic programs. In: ICLP, pp. 367–381 (1995)

Author Index