



Kobe Shoin Women's University Repository

Title	A Computational Treatment of V-V Compounds in Japanese 日本語におけるV-V 複合語の計算言語学的取扱い
Author(s)	橋本 力
Citation	博士学位論文 内容の要旨および審査結果の要旨, 第1号: 1-11
Issue Date	2005
Resource Type	Others/その他
Resource Version	
URL	
Right	
Additional Information	学位記番号: 甲第1号

博 士 学 位 論 文

内容の要旨および審査結果の要旨

第 1 号

平成 17 年 6 月

神戸松蔭女子学院大学

この冊子は、学位規則（昭和28年文部省令第9号）第8条および神戸松蔭女子学院大学大学院学位規程第11条の規定により、本学において、平成17年3月10日に授与した博士（言語科学）の学位について、その論文の内容の要旨および論文審査結果の要旨を公表するためのものである。

博士論文要旨

氏名	橋本 力
学位の種類	博士（言語科学）
学位記番号	甲第 1 号
学位授与年月日	平成 17 年 3 月 10 日
学位授与の条件	神戸松蔭女子学院大学大学院学位規程第 3 条 2 項による
学位論文題目	A Computational Treatment of V-V Compounds in Japanese (日本語における V-V 複合語の計算言語学的取扱い)

論文の要旨

本論文の目的は、日本語複合動詞の言語学的分析に基づいた計算機処理方法を開発することと、それを通じて、自然言語処理がなぜ、そして、いかにして理論言語学の成果を利用すべきなのかを論じることである。

1 章では理論言語学と自然言語処理の目的や方法論の違いを述べ、その上でなお、自然言語処理が言語学的手法を積極的に活用すべきことを論じている。その根拠は、言語学的に「深く」解析することにより意味表示が得られることと、従来の表層的自然言語処理技術が大きく依存していた統計情報への依存度を減らすことができることである。複合動詞は近年自然言語処理において重要課題として認識されつつある **Multiword Expressions (MWEs)** (Sag, Baldwin, Bond, Copestake, & Flickinger, 2002) の 1 つである。全ての複合動詞を一様に合成的に扱う方法や、逆に全てを区別無く一語として辞書に登録するような単純な方法では、理論的にも工学的にも満足な結果は得られない。前者の方法では非文法的な複合動詞を排除することも、ある複合動詞のイディオムの側面を捉えることもできない。Sag et al. (2002) はこれらを **overgeneration problem**、**idiomaticity problem** を呼んでいる。後者の方法は **flexibility problem** と **lexical proliferation problem** に直面することが知られている。つまり、一部の複合動詞は柔軟、かつ、生産性合成性が高いため、全ての可能な組み合わせを辞書に列挙することは現実的ではなく、同時に、それらの意味や統語的振る舞いを適切に処理するのが困難になる。これらは複合動詞の計算機処理に言語学的知見が欠かせないことの根拠となる。

2 章では、橋田 (1997) による、言語理論が工学に応用され得るための 4 つの基準、現象の重要性 (理論が対象としている問題が工学的にも重要であるか)、設計の単純性 (理論がシステムの設計を単純にするか)、計算の容易性 (理論の予測を導くための計算が効率的に行えるか)、入力の利用可能性 (理論の参照する情報がシステムにも容易に入手可能であるか) についてまず述べている。次に、本論文の複合動詞の分析の基礎をなす、影山 (1993) と Matsumoto (1996) による分析を、橋田 (1997) の基準に照らしながら批判的に概観している。

影山 (1993) は統率・束縛理論に基づき、複合動詞を統語的複合動詞と語彙的複

合動詞に区別し、さらに統語的複合動詞を、繰り上げ型、コントロール型、 \bar{V} 補文型の3つに分類している。語彙的複合動詞に関して項構造に基づく原則を提案し、いくつかの例外に対しては逆形成に基づく分析と語彙概念構造に基づく分析を与えている。影山(1993)の分析は複合動詞のMWEs的側面を計算機上で扱うための有効な理論的基盤を与えてくれているが、いくつかの工学的問題を抱えている。統語的複合動詞の分析では、統率・束縛理論によって立つことから計算の容易性に違反する。つまり、統率・束縛理論、特に移動分析や空範疇分析には、厳密で形式的な数学的基盤が欠けており、効率的な処理技術が確立されていない。

語彙的複合動詞の分析では、例外の扱いが設計の単純性と現象の重要性に違反する。つまり、少数の例外に対して工学的コストが高い説明を与えている。

Matsumoto(1996)は語彙的複合動詞に関して、項構造に基づき、網羅的で示唆に富む観察を示している。Matsumoto(1996)は語彙的複合動詞をPair compounds、Cause compounds、Manner compounds、Means compounds、Compounds exhibiting other relations、Compounds with semantically deverbilized V_2 、Compounds with semantically deverbilized V_1 の7つに分類し、前項動詞と後項動詞の意味関係に基づいて詳細な意味論的分析を試みている。しかし、前項動詞と後項動詞の意味関係の認識には文脈や世界知識が大きく関わっており、計算機にそれを行わせることは容易ではない。つまりMatsumoto(1996)の分析は計算機に入手困難な情報を参照している点で橋田(1997)の入力の利用可能性に違反する。また、Matsumoto(1996)が用いている緻密、かつ、個人間で判断が一致しにくいと思われる意味概念は、大規模な計算機用文法や辞書を構築する際に障害となり、設計の単純性に違反する。

一方、2章を通してMWEsとしての複合動詞の扱いの難しさを明らかにしており、計算機処理においても言語学的手法が不可欠であることを示している。

3章では本論文の分析を示している。まず、橋田(1997)の基準に沿って文法の設計方針を述べている。本分析では、現象の重要性の基準を満たすため、例外的現象や個人間で文法性判断が揺れるような現象の扱いにより規則が複雑化するようなことは避けた。設計の単純性を満たすため、理論的に高度であるより保守的な記述に努めた。また、計算機に与えることが可能な情報のみを参照するように規則を設計することで、入力の利用可能性を満たした。計算の容易性に関しては、文法の記述に \mathcal{TDL} 言語(Krieger & Schafer, 1994)を用いることで対応した。

次に本分析の枠組について述べている。本分析は計算機用大規模日本語文法JACY(Siegel & Bender, 2002)の上に実装されている。そのため統語論的枠組は**Head-driven Phrase Structure Grammar (HPSG)**(Sag and Wasow (1999))、意味論的枠組は**Minimal Recursion Semantics (MRS)**(Copestake et al. (1999))に依存している。文法開発には**LKB**(Copestake, 2002)を用いた。

本論文の統語的複合動詞の扱いは影山(1993)に概ね従っており、A、B、Cの3タイプに分類している(橋本, 2003)。特に、A、Bタイプには、従来のほとんどの計算機用日本語文法では語順の問題のため扱われてこなかったが理論的には重要な、VP埋め込み構造を採用している。言語学的な正確さを重視した結果、高精度

な自然言語処理で重要になる緻密な意味表示を得ることに成功した。さらに本分析は、移動も空範疇も用いない簡潔で明示的な句構造分析であり、影山 (1993) と同等の理論的な説明力を持ちつつも、計算の容易性を満たしている。しかし一方で、VP 埋め込み構造はかき混ぜ現象の扱いを難しくする。JACY の枠組の中でかき混ぜ現象を適切に扱うために、本分析ではãによる分析を採用している。この分析によりかき混ぜを正しく扱え、かつ、適切に制限した結果、他の分析では得難い処理効率を実現している。

語彙的複合動詞は、Matsumoto (1996) を参考に、**Right headed V_1 - V_2** 、**Argument mixing V_1 - V_2** 、 **V_1 - V_2 with semantically deverbilized V_1** 、 **V_1 - V_2 with semantically deverbilized V_2** 、**Non-compositional V_1 - V_2** に分類している。Matsumoto (1996) の Pair、Cause、Manner、Means の 4 種は Right headed V_1 - V_2 と Argument mixing V_1 - V_2 によりカバーされ、その 4 つの意味関係は文法においては未指定のままにしている。これは入力の利用可能性の基準から支持される。本分析は今泉・郡司 (2000) の項構造に基づく簡潔なものであり、さらに高いカバー率を持つ。項構造は従来の大規模計算機用日本語文法では、複雑さを避けるため扱われて来なかったが、VP 埋め込み構造と同様、理論的な正確さを保つ上で不可欠である。これにより Matsumoto (1996) の分析では満たし難い設計の単純性を満たしつつも、同等の理論的説明力を持つことができる。さらに、統語的・意味的な基準に基づく分類により、統語的な生産性と意味的な合成性を適切に捉えることに成功している。意味表示は、統語的複合動詞と同様、緻密なものになっている。

3 章を通して、本分析が、橋田 (1997) の基準を満たしつつも、複合動詞の理論的あるいは MWEs 的な側面を適切に捉えていることを明らかにしている。特に、従来 of 計算機用日本語文法では扱えていなかった VP 埋め込み構造と項構造の導入が大きな役割を果たしている。

4 章では、本論文の実装の工学的観点からの評価実験について述べている。具体的には、本文法のカバー率、曖昧性の多寡、処理効率を、文法評価システム [incr tsdb()] (Oepen & Carroll, 2000) と Lexeed コーパス (笠原他, 2004) を用いて調べた。実験の際には本分析が実装されている JACY-vv とオリジナルのままの JACY-plain を用意して、カバー率、曖昧性の多寡、処理効率の差を調べた。JACY-vv には複合動詞の語彙項目は (non-compositional V_1 - V_2 を除いて) 全く与えられていないが、複合動詞規則が実装されている。JACY-plain は 1,325 の複合動詞語彙項目が与えられているが、規則は実装されていない。結果として、JACY-vv は JACY-plain よりも高いカバー率と少ない曖昧性が得られることが判明した。カバー率の向上は主に、複合動詞の高度な生産性によるもの、言い替えれば、MWEs に関わる問題の一つ、lexical proliferation problem によるものだった。つまり、JACY-vv は未知の複合動詞が入力されても適切に解析することが可能だが、JACY-plain は膨大な数の複合動詞が辞書に登録されているにも関わらず、その生産性の高さには追いつけなかった。曖昧性の減少は VP 埋め込み構造に関わるかき混ぜ現象の扱いの違いによるものだった。つまり、本分析のãに基づく分析が正確、かつ、より適切に制

限されていることが要因であった。また、JACY-vv は言語学的分析により生産性・合成性の高いものとそうでないものを区別して扱っているため、規則による複合動詞の扱いが直面し得る MWEs に関わる問題、overgeneration problem と idiomaticity problem にも適切に対応できている。一方、処理効率では JACY-plain よりも劣っている。一般に、規則が増えれば処理効率は落ちてしまうのだが、本章では、規則の形式化を工夫することで改善し得ることを議論している。

5 章では、今後の課題と、言語学と自然言語処理のあるべき関係について議論している。今後の課題として、本分析では捉え切れなかった現象をいかに扱うか、non-compositional V_1 - V_2 をどうやって自動的に検出、収集するか、日本語複合動詞はどのように翻訳されるべきなのか、の 3 点を挙げている。本分析で扱い切れなかった複合動詞については、それらがどの程度生産的なのかをコーパスなどを通して見極め、新たな規則を追加するか、それらを 1 語として辞書に登録するかを検討すべきであることを述べている。non-compositional V_1 - V_2 の自動検出については、英語の句動詞の自動検出の研究を例に挙げ、それらの日本語複合動詞に対する適用可能性について議論している。言語学と自然言語処理の今後のあるべき関係については、自然言語処理の応用において言語学的手法では賅えない技術には何があるのかをまず述べている。最後に、自然言語処理が言語学の究極の目標にいかに関与するかについて議論し、言語学的手法に基づく自然言語処理を提唱している。

参考文献

- Copestake, A. (2002). *Implementing Typed Feature Structure Grammars*. CSLI Publications.
- Copestake, A., Flickinger, D. P., & Sag, I. A. (1999). *Minimal Recursion Semantics: An Introduction*. Manuscript, Stanford University: CSLI.
- Krieger, H.-U. & Schafer, U. (1994). *\mathcal{TDL} — A type description language for constraint-based grammars*. In *Proceedings of the 15th International Conference on Computational Linguistics*.
- Matsumoto, Y. (1996). *Complex Predicates in Japanese: A Syntactic and Semantic Study of the Notion 'Word'*. CSLI Publications.
- Oepen, S. & Carroll, J. (2000). Performance profiling for grammar engineering. *Natural Language Engineering*, 81–97.
- Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002). Multiword expressions: A pain in the neck for NLP. In *Computational Linguistics and In-*

telligent Text Processing: Third International Conference, pp. 1–15 Mexico City, Mexico.

Sag, I. A. & Wasow, T. (1999). *Syntactic Theory: A Formal Introduction*. Center for the Study of Language and Information, Stanford. Japanese edition (two volumes) – translated and edited by Takao Gunji and Yasunari Harada, appeared in 2001.

Siegel, M. & Bender, E. M. (2002). Efficient Deep Processing of Japanese. In *Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization* Taipei, Taiwan.

橋田浩一 (1997). 「言語への情報科学的アプローチ」. 大津由紀雄, 郡司隆男, 田窪行則, 長尾真, 橋田浩一, 益岡隆志, 松本裕治 (編), 『言語の科学入門』, 3章. 岩波書店.

今泉志奈子・郡司隆男 (2000). 「語彙的複合における複合事象」. 伊藤たかね・矢田部修一 (編), 『レキシコンと統語』, pp. 33–59. ひつじ書房.

影山太郎 (1993). 『文法と語形成』. ひつじ書房.

笠原要, 佐藤浩史, Francis Bond, 田中貴秋, 藤田早苗, 金杉友子, 天野成昭 (2004). 「基本語意味データベース:Lexeed」の構築」. 『情報処理学会自然言語処理研究会報告 2004-NL-159』, pp. 75–82 東京.

橋本力 (2003). 「長距離受け身の HPSG に基づく分析」. 『日本言語学会第 126 回研究大会』, pp. 256–261 東京.

ABSTRACT OF THE THESIS

The purposes of the thesis are to implement the linguistic analyses of Japanese verbal compounds in a computational grammar of Japanese and to discuss why and how Natural Language Processing (NLP) should benefit from theoretical linguistics.

In chapter 1, I describe the difference between theoretical linguistics and NLP, and then I argue that NLP should make use of linguistics because we can acquire a fine-grained semantic representation by means of a deep linguistic treatment and because a linguistic treatment of NLP does not have to rely as heavily on statistical information, as long as a grammar describes a language precisely. Japanese verbal compounds (V_1 - V_2 compounds) are one kind of **Multiword Expressions (MWEs)** (Sag et al., 2002), which NLP researchers have recently acknowledged as a troublesome problem. As such, V_1 - V_2 compounds resist simple solutions. If we regard all MWEs as totally compositional, and derive all of them by means of some sort of rule, we would face the **overgeneration problem** and the **idiomaticity problem**; that is, we would overgenerate unattested V_1 - V_2 s and cannot treat V_1 - V_2 's idiomaticity. On the other hand, if we regard all MWEs as single words, and register all of them in the lexicon, then we would suffer from the **flexibility problem** and the **lexical proliferation problem**; namely we would fail to express V_1 - V_2 's productivity. These constitute the evidences that we definitely need a sophisticated linguistic analysis to deal with V_1 - V_2 compounds.

In chapter 2, I first describe the criteria of Hasida (1997) by which a linguistic theory is judged to be suitable for NLP. The criteria include **Importance of Phenomena**, whether the problem that a linguistic theory tries to account for is also important for NLP, **Simplicity of Design**, whether a theory makes an NLP system simple, **Efficiency of Computation**, whether computation posited by a theory is executed by computer efficiently, and **Availability of Input**, whether inputs that a theory makes reference to are easily available for NLP systems. Next I move on to a critique of Kageyama (1993) and Matsumoto (1996) in light of Hasida (1997), although my analysis owes much to them.

Based on the Government and Binding (GB) theory, Kageyama (1993) distinguishes **syntactic V_1 - V_2 compounds** and **lexical V_1 - V_2 compounds**. He further divides syntactic V_1 - V_2 s into Raising, Control, and \bar{V} complementation types. Regarding lexical V_1 - V_2 s, he proposes the Transitivity Harmony Principle, and posits a back formation analysis and an LCS analysis for some exceptions to the principle. Although Kageyama's analysis gives us a theoretical basis for computational implementation of V_1 - V_2 compounds, it has several defects in terms of Hasida (1997); the GB analyses, especially the movement analysis and the empty category analysis, lack a mathematical foundation, and thus lack efficient processing techniques, resulting in a violation of Efficiency of Computation. In addition, his analysis of lexical V_1 - V_2 s violates Simplicity of Design and Importance of Phenomena, since he posits computationally expensive machinery to

account for a few exceptions.

Matsumoto (1996) presents comprehensive and suggestive observations about lexical V_1 - V_2 s based on argument structure. He classifies lexical V_1 - V_2 s into Pair compounds, Cause compounds, Manner compounds, Means compounds, Compounds exhibiting other relations, Compounds with semantically deverbilized V_2 , and Compounds with semantically deverbilized V_1 , and tries to analyze them in terms of a semantic relation between V_1 and V_2 . However, recognizing such a semantic relation involves pragmatics or world knowledge, which means that it would be difficult for computers to do such a job. In other words, the analysis of Matsumoto (1996) violates Availability of Input in that it refers to information that a computer cannot easily obtain. As well, semantic notions that his lexical analysis makes use of are too fine-grained for us to develop an appropriate large-scale grammar and lexicon, resulting in a violation of Simplicity of Design.

Through this chapter, it is shown that a sophisticated linguistic analysis is indispensable for a computational treatment of V_1 - V_2 compounds, since they show complicated MWE characteristics.

In chapter 3, I present my analysis of V_1 - V_2 compounds. I first describe my policy of grammar development that observes the criteria of Hasida (1997). In order to satisfy Importance of Phenomena, I avoid complicating my analysis to account for exceptional cases and linguistic phenomena where people's judgments are not stable or consistent. Also, to satisfy Simplicity of Design, I make my analysis descriptively adequate rather than theoretically advanced, even though it is not maximally parsimonious. Availability of Input is met by restricting information that is referred to by my analysis to material that is computationally available. As for Efficiency of Computation, I adopt the \mathcal{TDL} language (Krieger & Schafer, 1994) as a grammar description language so that my analysis can be executed efficiently.

I also describe the framework of my analysis. I implement my analysis in the existing computational grammar of Japanese, JACY (Siegel & Bender, 2002), which adopts **Head-driven Phrase Structure Grammar (HPSG)**, Sag and Wasow (1999) as a syntactic framework and **Minimal Recursion Semantics (MRS)**, Copestake et al. (1999) as a semantic framework. In the implementation, I use the **LKB** system (Copestake, 2002).

My analysis of syntactic V_1 - V_2 s roughly follows Kageyama (1993), and I classify syntactic V_1 - V_2 s into **A type**, **B type**, and **C type** (Hashimoto, 2003). In particular, I posit VP embedding structures for A and B type. The structure is indispensable for a theoretically precise analysis for them, although almost all of the previous computational grammars of Japanese have avoided it because of a difficulty involving scrambling. As a result, I can utilize a fine-grained semantic representation, which is essential to a precise NLP. Besides, my analysis is a simple phrase structure analysis without movement

or empty categories, and still it is theoretically precise. This way, my analysis satisfies Efficiency of Computation. However, the VP embedding structures cause a problem concerning scrambling. To get around the problem, I posit \hat{a} , which is precise and properly restricted. The approach is more efficient than alternative approaches thanks to its restrictive nature.

Roughly following Matsumoto (1996), I classify lexical V_1 - V_2 s into **Right headed V_1 - V_2** , **Argument mixing V_1 - V_2** , **V_1 - V_2 with semantically deverbalized V_1** , **V_1 - V_2 with semantically deverbalized V_2** , and **Non-compositional V_1 - V_2** . Right headed V_1 - V_2 and Argument mixing V_1 - V_2 cover the Pair, Cause, Manner, and Means compounds of Matsumoto (1996), but I underspecify these four semantic relations. This strategy is justifiable on the ground of Availability of Input. My analysis of lexical V_1 - V_2 s is simple and is based on argument structure of Imaizumi and Gunji (2000). Previous computational grammars of Japanese have avoided adopting argument structure, but it is also essential to theoretical preciseness. Thanks to the conciseness and the argument structure, my analysis better satisfies Simplicity of Design. In addition, it successfully accounts for lexical V_1 - V_2 's syntactic and semantic properties. Especially, we can acquire the correct semantic representation of lexical V_1 - V_2 s, as well as that of syntactic V_1 - V_2 s.

Through chapter 3, it is shown that my analysis captures the MWEs properties of V_1 - V_2 compounds while observing the criteria of Hasida (1997). Notably, the VP embedding structures and argument structure play a important role.

In chapter 4, I describe the evaluation experiment through which I illustrate the coverage, the number of ambiguity and the efficiency of my implementation. In the evaluation, I used the [incr tsdb()] system (Oepen & Carroll, 2000) and the Lexeed corpus (Kasahara et al., 2004). I also prepared two versions of JACY: JACY-vv and JACY-plain. JACY-vv is equipped with my implementation, but is not given lexical entries for V_1 - V_2 s except for those of non-compositional V_1 - V_2 s. On the other hand, JACY-plain, which is the original one, has no rule for V_1 - V_2 s, but contains 1,325 lexical entries for V_1 - V_2 s in the lexicon. Consequently, JACY-vv outperformed JACY-plain in terms of coverage and the number of ambiguity. The more coverage was gained because of the remarkably high productivity of V_1 - V_2 s. JACY-vv, but not JACY-plain, could deal with it. In other words, JACY-vv could get around the lexical proliferation problem; it can handle the unknown V_1 - V_2 s by means of appropriate rules. On the other hand, the 1,325 entries of JACY-plain, which was not quite small, could not deal with the productivity. The reason for the less ambiguity involves the difference of the treatment of scrambling from an embedded VP. To be more precise, the restrictive nature of my \hat{a} approach made us get less ambiguity. Also, since JACY-vv distinguishes productive V_1 - V_2 s from non-productive ones and compositional V_1 - V_2 s from non-compositional ones, it can get around the overgeneration problem and the idiomaticity problem. However,

as for performance, JACY-vv turned out to be working less efficiently than JACY-plain. Generally, more rules lead to less efficiency, but I discuss the possibility that changing grammatical representations would make the grammar more efficient.

In chapter 5, I discuss future works and the prospect of the relationship between theoretical linguistics and NLP. The future works include how we treat V_1 - V_2 s that the current analysis cannot deal with, how we automatically detect non-compositional V_1 - V_2 s, and how we make computers translate Japanese V_1 - V_2 s into English expressions. Regarding the treatment of problematic V_1 - V_2 s, I claim that, first of all, we should find how productive they are through a corpus study. If they are really productive, we should add new rules to deal with them. Otherwise, we should enter them in the lexicon as single words. As for the automatic detection of non-compositional V_1 - V_2 s, I take up the studies on the automatic detection of English phrasal verbs, and discuss the applicability of the studies to Japanese V_1 - V_2 compounds. Finally, I discuss the prospect of the two studies of language: theoretical linguistics and NLP. I mention several NLP problems that theoretical linguistics cannot help. Then I discuss how NLP contributes to the resolution of the biggest issues of linguistics, and advocate a deep linguistic NLP.

References

- Copestake, A. (2002). *Implementing Typed Feature Structure Grammars*. CSLI Publications.
- Copestake, A., Flickinger, D. P., & Sag, I. A. (1999). *Minimal Recursion Semantics: An Introduction*. Manuscript, Stanford University: CSLI.
- Hashimoto, C. (2003). HPSG analysis of Long Distance Passives (in Japanese). In *Proceedings of the 126th annual meeting of Linguistic Society of Japan*, pp. 256–261 Tokyo, Japan.
- Hasida, K. (1997). Information Science Approach to Language. In Ôtsu, Y., Gunji, T., Takubo, Y., Nagao, M., Hasida, K., Masuoka, T., & Matsumoto, Y. (Eds.), *An Introduction to Language Science* (in Japanese), chap. 3. Iwanami.
- Imaizumi, S. & Gunji, T. (2000). Complex Events in Lexical Compounds. In Itou, T. & Yatabe, S. (Eds.), *Lexicon and Syntax* (in Japanese), pp. 33–59. Hituzi Syobo.
- Kageyama, T. (1993). *Grammar and Word Formation* (in Japanese). Hituzi Syobo.
- Kasahara, K., Sato, H., Bond, F., Tanaka, T., Fujita, S., Kanasugi, Y., & Amano, S. (2004). Construction of a Japanese Semantic Lexicon: Lexeed. In *Information Processing Society of Japan, 2004-NL-159*, pp. 75–82 Tokyo, Japan.

- Krieger, H.-U. & Schafer, U. (1994). *\mathcal{TDL} — A type description language for constraint-based grammars*. In *Proceedings of the 15th International Conference on Computational Linguistics*.
- Matsumoto, Y. (1996). *Complex Predicates in Japanese: A Syntactic and Semantic Study of the Notion ‘Word’*. CSLI Publications.
- Oepen, S. & Carroll, J. (2000). Performance profiling for grammar engineering. *Natural Language Engineering*, 81–97.
- Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002). Multiword expressions: A pain in the neck for NLP. In *Computational Linguistics and Intelligent Text Processing: Third International Conference*, pp. 1–15 Mexico City, Mexico.
- Sag, I. A. & Wasow, T. (1999). *Syntactic Theory: A Formal Introduction*. Center for the Study of Language and Information, Stanford. Japanese edition (two volumes) – translated and edited by Takao Gunji and Yasunari Harada, appeared in 2001.
- Siegel, M. & Bender, E. M. (2002). Efficient Deep Processing of Japanese. In *Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization* Taipei, Taiwan.

論文審査結果の要旨

橋本氏は、博士論文において、日本語複合動詞の言語学的分析に基づいた計算機処理方法を開発するとともに、自然言語処理と理論言語学との関係について論じた。

同氏の研究方法は、言語理論に基づいた計算機システムを構成するに際して、理論的成果を単に無批判に再現するのではなく、機械による効率のよい言語処理という観点から、システムの設計をするということである。また、理論的成果を論じるときにも、異なる説が並存している場合には、それぞれの長所短所を理論的見地から論じている。その意味で、氏の博士論文、および、氏の研究は、理論言語学と計算言語学の成果が理想的な形で取り込まれたものであると言える。

本論文は、日本語複合動詞の言語学的分析に基づいた計算機処理方法を開発するとともに、自然言語処理と理論言語学との関係について論じたものである。

特に、既存の計算機によるアプローチが、複合語をそのままの形で分析せずに辞書に登録するという形をとっているのに対して、筆者の提案する方法で個々に分析して登録する方が、全体として効率を向上させているという議論は説得的であり、かつまた、コーパスデータでそれが実際に検証されたことは意義深い。

本論文にも数多くの残された課題がある。一例をあげれば、理論的な整合性に重点を置いた設計のために、一部効率が犠牲になっているところがある。ただし、それととも、システムが見通しのよい設計であるため、設計の見直しによって、理論的整合性を損わない形で改善が可能な問題であり、大きな障害となることは考えにくい。

最終試験においては、論文の概要の報告とともに、残された問題、今後の課題などについても質疑をおこなったが、本研究の基本的なアプローチは保ったまま、発展させていく余地のあることが説得的に示された。

以上より、本論文は本学大学院博士論文として十分に価値のあるものであると判断される。

主査	教授	郡司 隆男
副査	教授	Joseph Emonds
副査	教授	西垣内 泰介