2004－NL－159　(13)
2004／1／13

日本語ツリーバンク「檜」: 言語理解のためのコーパス

Francis Bond* 藤田 早苗 * 橋本 力 † 笠原 要 * 成山 重子 ‡§
Eric Nichols§ 大谷 朗 ¶ 田中 貴秋 * 天野 成昭 *
* 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
† 神戸松蔭女子学院大学大学院 ‡ メルボルン大学
§ 奈良先端科学技術大学院大学 情報科学研究科 ¶ 大阪学院大学 情報学部
{bond, sanae, takaaki}@cslab.kecl.ntt.co.jp chashi@sils.shoin.ac.jp
{eric-n, shigeko-n}@is.aist-nara.ac.jp ohtani@utc.osaka-gu.ac.jp

概要

本稿では、基本語彙知識ベース構築の一環として構築した、ツリーバンク「檜」を紹介する。「檜」は、HPSG で書かれた日本語文法 JaCY に基づいて辞書の語義文を解析したものであり、詳細な統語情報と意味情報の両方が付与されている。本稿では、「檜」構築の目的や理論的基盤などについて述べる。 また、「檜」の有効性を示す一例として、知識獲得の予備実験を行なった結果について報告する。

キーワード: ツリーバンク、意味表現、知識獲得、日本語、主辞駆動句構造文法

The Hinoki Treebank
A Treebank for Text Understanding

Francis Bond* Sanae Fujita* Chikara Hashimoto†
Kaname Kasahara* Shigeko Nariyama‡§ Eric Nichols§
Akira Ohtani¶ Takaaki Tanaka* Shigeaki Amano*
*NTT Communication Science Laboratories, NTT Corporation
†Kobe Shoin Graduate School ‡The University of Melbourne
Graduate School of Information Science, Nara Institute of §Science and Technology
¶Faculty of Informatics, Osaka Gakuin University

Abstract

In this paper we present the motivation for the construction of the Hinoki treebank. It is a rich and dynamic treebank of dictionary definition sentences parsed using a Japanese HPSG. We show how the treebank is being used to build an ontology, and outline plans for further work.

Keywords: Treebanking, Semantic representation, Knowledge acquisition, Japanese, HPSG

1　はじめに

我々は、統語情報と意味情報を統合して扱える自然言語処理を目指して、「基本語彙知識ベース」の構築を進めている [1]。究極の目標は機械に自然言語を理解させることであり、そのためにテキストを統語的に解析するだけでなく、意味的に解析し、意味情報を獲得することを目指している。

ここ十数年来、自然言語処理の分野では、コーパスや辞書などの電子化されたデータの増加に伴い、統計的自然言語処理を用いた諸問題の解決技術が飛躍的に進歩している。統計的手法の利点は、対象とするデータに対して学習データがあれば、効率的に精度とカバー率を向上できる点である。統計的手法は形態素解析、構文解析、語義曖昧性解消の問題などで、大きな成果を上げている。

しかしその一方で、性能は学習するデータに依存するため、学習データの性質に起因する限界が

*本稿は 03 年 7 月より NTT コミュニケーション科学基礎研究所で行なった「檜ツリーバンクプロジェクト」について、その進捗および研究成果の一部を報告するものである。

ある。多くの場合、学習データとして表層的な情報のみが使われている。一つの問題は、学習データ中に、低頻度でしか観測できない言語現象については、十分に学習を行うことができない場合があるということである。言語現象は多様であるため、大量のデータを使用しても、対象とする表現と表層的に同じ表現が学習データ中に現れないことも多い。もう一つの問題は、統語や意味に関する詳細な情報を用いていない点である。例えば、「条約を日中で締結した」「会談を日中で終えた」の「日中」の曖昧性は単語の共起や依存関係の情報だけでは区別できない。

我々は、学習するデータとして表層的な情報だけでなく詳細な統語情報や意味情報を利用することで、これらの問題を解決できると考えている。統計的手法を含む自然言語処理技術で、自然言語の理解といったより高度な問題を扱うためには、より詳細な統語情報や意味情報が付与されたコーパスを用いて、統計的手法と深い意味処理に基づく手法を融合していくことが望まれる [4, 5]。しかし、現時点ではこれらの情報が付与された大規模なコーパスは存在しない。文節間の依存関係を付与されたコーパスであれば、日本語では京大コーパス [2] や EDR[3] などがある。しかし、下位範疇化構造などの詳細な統語情報や、単語の意味の区別するタグなどの意味情報を持つ日本語の大規模なコーパスは存在しない。

これまで、意味情報まで付与されている大規模な日本語コーパスが存在しなかった理由は、(i) 構築が難しかったこと、(ii) 単語あるいは文節間の依存関係程度の情報で十分であると認識されてきていたこと[1]、という 2 点が挙げられる。

しかし、(i) に関しては、近年、DELPH-IN(2.4 章参照) 等により、文法理論の実装とコーパス構築を同時に、かつ、効率的に行なえるツールが整備されてきている。そのため、文法のエキスパートでなくても比較的容易に、非常に詳細な情報をもつコーパスを、統計学習に利用できる規模で構築する事が可能となってきている [6]。

(ii) に関しては、前述したように表層的な情報やその間の依存関係の情報だけでは対処できない問題がある。また、意味情報を利用すれば、分野や言語に依存しない、汎用的な自然言語処理技術を実現できるという利点がある。例えば、機械翻訳では、言語非依存な意味表現を介して他の言語に翻訳することができる。そのため、多くの翻訳規則を言語対毎に作成する必要がなくなる。

我々は、統計的手法と意味情報を融合していくための第 1 段階としてツリーバンク「檜」を構築している。「檜」には統語情報と意味情報の両方の情報が統合されて付与されている。このようなコーパスは、質問応答や要約、翻訳など、あらゆる自然言語処理分野において、統計的手法と意味情報を統合した高度な処理を実現する上で基盤的情報となり得る。

本稿ではツリーバンク「檜」の特徴として、構築に用いられた日本語文法・意味表現や対象データについて述べ、構築されたツリーバンクの有効性について報告する。特に、このような詳細な情報を付与されたツリーバンクの有効性を示す一例として、知識獲得への利用を考案し、予備実験を行なった結果を報告する。

以下、2 章は解析を支える文法理論やツリーバンクの構築に利用するツールについて、3 章はツリーバンクの構築対象データである辞書の語義文について紹介する。4 章は本ツリーバンクの有効性を示す一例として行なった知識獲得実験について報告する。5 章は「檜」ツリーバンクプロジェクトの今後の方向性について述べる。

2 文法解析の理論的基盤

2.1 HPSG に基づく日本語文法の実装

HPSG とは、統語解析と意味解析が密接に関連した解析が可能な、主辞駆動句構造文法 (Head-driven Phrase Structure Grammar: HPSG) [7] であり、言語学および隣接科学における数多くの異なった研究に基づいている。HPSG は、重要なアイデアの多くを意味理論や情報科学の成果より得ており、その枠組は統語解析などの言語処理の基礎技術だけでなく、形式意味論との親和性も高い。

HPSG は型付き素性構造 (Typed Feature Structure: TFS) を用いて語彙等に内在する統語・意味・表記等の性質を宣言的に記述し、そうした情報に関する制約を単一化として捉えることで言語現象を形式的に説明する。例として、下に示す文 (1) を解析した TFS を簡単化し、図 1 に示す。図 1 に

[1]なお、ツリーバンク「檜」から依存関係のみ抽出することもできる。そのため、「檜」は、依存関係の情報のみ付与されたコーパスに対して上位互換であるといえる。

おいて、CAT は統語、CONT は意味、ORTH は表記の記述である。

(1)　自動車を運転する人

HPSG に立脚した日本語分析と文法実装の試みは，ICOT による JPSG[8] をはじめとして既にいくつか存在している[2]。しかし、それらはいずれも日本語の一般的性質に関する理論の精緻化を指向したものか、一現象の説明における形式的妥当性の証明を目的とするものであった。そうした研究は理論的には重要である。しかし、システムの実用性の面から見れば、これまでに構築されたモデルは小規模で、処理できるデータが少なく、現実的な解析を行なうには不十分な規模であった。

2.2　JaCY

いわば実験的な取組みであった先行研究に対し、本稿で利用する JaCY[10] は、7,000 以上の語彙を備えた実用指向の大規模日本語 HPSG 文法である。この文法は、Verbmobil プロジェクト [11] に端を発し、当初は旅行企画に関する対話文の処理を意図して記述されていた。その後、対象分野を広げつつ語彙や規則の拡張をすすめてきた。また、JaCY を使った解析器は未知語に対して形態素解析システム茶筌[3]の出力を利用することで、記述文や電子メールの自動応答にも対応できる実用システムとなってきている。

更に、JaCY の利点としては、MRS(2.3 章参照) による意味記述を採用していること、DELPH-IN(2.4 章参照) による文法開発環境を利用できること、という 2 点があげられる。同様に、MRS を採用し、DELPH-IN による文法開発環境を利用している HPSG 文法としては、英語を対象とした English Resource Grammar (ERG) [12] がある。

2.3　MRS

最小再帰意味論 (Minimal Recursion Semantics: MRS) [13] は、他の形式意味論研究と同様に自然言語の意味に関する記述的な妥当性を検証する体系を構築するものであり、文法情報、特に統語情報との連携を強く意識している。

[2]Lexical Functional Grammar を理論的基盤とする ParGram プロジェクトにおいても実装環境の整備と日本語の分析 [9] がすすめられている。本稿と同様に大規模データの解析を指向した文法設計ではあるが、ツリーバンクの構築はなされていないため、本稿では比較検討をしていない。

[3]http://chasen.aist-nara.ac.jp/を参照。

MRS は、形式記述のためのメタ言語であり、目的に応じて適切な体系や必要な演算操作を取り込む柔軟さを持つ。その特長を簡単に述べると、(i) 不完全指定 (Underspecification) を許容すること、(ii) 階層の少ない意味構造を生成することがあげられる。

(i) については、辞書に完全な情報がなくても解析できるため、未知語などについて柔軟な対応ができる、という利点がある。(ii) については、階層の少ない意味構造は、自然言語処理で非常に扱いやすいという利点がある。特に、MRS は様々な言語の文法から出力でき、異なる言語でも同等な意味表現になる。このため、異なる言語同士でも対応が取りやすく、多言語処理にも有効である。

図 2 に、文 (1) を解析した場合の MRS を例示する。また、異なる言語の MRS の例として、文 (1) と同じ意味の英文 (2) を ERG によって解析した結果の MRS を、図 3 に示す。但し、図 2, 3 の MRS は簡単化してある。

(2)　*somebody who drives a car*

図 2, 3 のように、異なる言語を解析したものでありながら、ほぼ同等の MRS を構築できる。現在、MRS を獲得できる文法としては、日本語 (JaCY)、英語 (ERG)、ドイツ語 (DISCO) の他、韓国語、ノルウェー語、イタリア語、ギリシア語の文法の開発が進んでいる。

$$\begin{aligned}\langle h, x_1 \{ & h : prpstn\_rel(h_1) \\ & h_1 : hito(x_1) \\ & h_2 : jidosha(x_2) \\ & h_3 : unten(u_1, x_1, x_2) \}\rangle\end{aligned}$$

図 2:「自動車を運転する人」の解析結果 (MRS)

$$\begin{aligned}\langle h_1, x_1 \{ & h : prpstn\_rel(h_0) \\ & h_0 : person(x_1) \\ & h_1 : some(x_1, h_0, h_4) \\ & h_2 : car(x_2) \\ & h_3 : drive(u_1, x_1, x_2) \}\rangle\end{aligned}$$

図 3: *somebody who drives a car* の解析結果 (MRS)

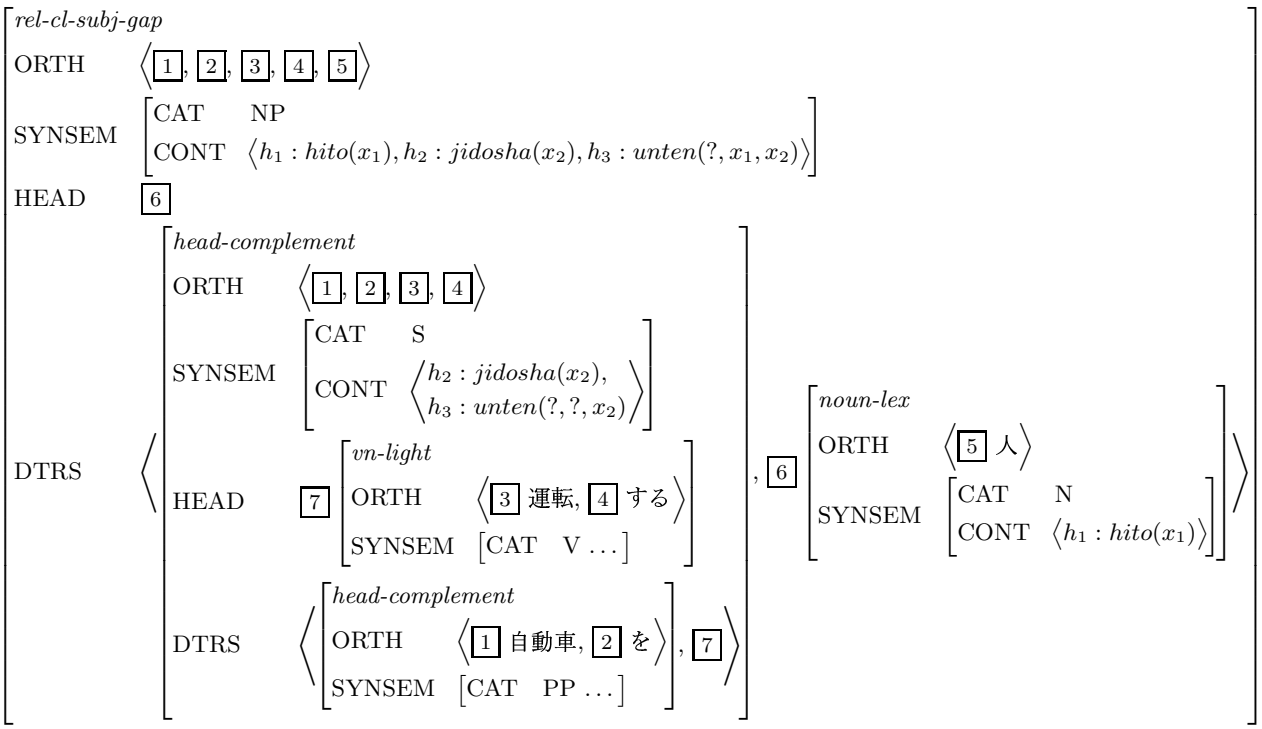

図 1: 「自動車を運転する人」の解析結果（HPSG の型付き素性構造）

2.4 DELPH-IN

HPSG は文法理論であると同時に TFS に基づく計算理論でもある。そのため言語処理との親和性が高く、その実装に関しては様々なツールが開発されている。本稿が文法開発の基盤に JaCY を採用する理由は、(i) 統語情報と意味情報を密接に関連付ける HPSG の実装が行なわれていること、(ii) 分析を即ちに処理に反映させることで効率的な文法拡張を支援する環境が、DELPH-IN (Deep Linguistic Processing with HPSG)[4]によって整備されていることである。本ツリーバンクの構築に利用する DELPH-IN のツールは、LKB(Linguistic Knowledge Builder)[14]、PET[15]、`[incr tsdb()]`[16] である。これらはお互いに連携して利用できる。

LKB とは、一般的な TFS の言語実装を支援するシステムであり、詳細な逐文解析と文法の実装のための開発環境を提供する。JaCY と ERG は、この LKB を用いて文法開発されてきている。PET とは、LKB 上に構築した文法を使って対象コーパス全体を高速に解析するツールである。`[incr tsdb()]` とは、文法の評価やツリーバンクの構築に用いるツールである。これらを用いたツリーバンク構築方法については、[6] で詳しく述べている。

[4]http://www.delph-in.net/を参照。

3 ツリーバンク「槍」

3.1 ツリーバンク「槍」の特徴

ツリーバンク「槍」は、統語情報と意味情報の両方が統合されて付与されたコーパスである。「槍」は、2 章で紹介した JaCY や DELPH-IN で開発された諸ツールを用いている。そのため、構築手法の特長として次の 2 点が挙げられる。(i) 解析結果に対して人手による修正を行なわず、自動的に解析した結果の中から、正しい解析結果を選択するだけで構築できる。(ii) 文法が修正された場合、修正された文法に合わせてツリーバンクを再構築できる。そのため、構築したツリーバンクから学習したデータに基づいて文法を改良すれば、その改良をツリーバンクに容易に反映できる。

ツリーバンク「槍」で解析対象としたのは、辞書の語義文である。辞書の語義文を利用する理由として、以下の 2 点が挙げられる。まず第 1 に、統語・意味情報が付与された語義文から、見出し語に関する知識を獲得できるという点である。更に、

「檜」で対象とした語義文は制限語彙で記述されており、語義文中で用いられている語もまた、辞書中で定義されている、という利点がある。

第 2 に、ツリーバンクの構築が比較的行ないやすいという点である。辞書の語義文は、ある程度出現する文型が決まっているため、解析のために新たに追加する文法規則が少なくて済む。実際に「檜」の構築の際にも、JaCY に対して高々数規則追加するだけで、解析率を 87%まで引き上げる事が可能であった [6]。また語義文は、新聞や小説と比べると比較的文が短く、固有表現や抽象的な表現が少ない、という特徴もある。

3.2 ツリーバンク「檜」の対象データ

実際のツリーバンク「檜」の構築対象の辞書として、基本語意味データベース Lexeed[1] を選択した。Lexeed は、日本語で一般的に使用されている語を網羅し、各語義に語義文が付けられている。Lexeed に収録されている語は、日本人の各語に対するなじみ深さの度合を表す「単語親密度」に基づいて選定されている。単語親密度は語に対するなじみ深さの度合を 1 から 7 の実数で表したものであり、7 が最もなじみ深いことを示す [17]。

このうち、単語親密度が 5 以上である 28,270 語が基本語と定義され、Lexeed に収録されている。基本語は、典型的な日本語の新聞に出現する一般語のうち延べ数で 75%以上をカバーしている [18]。

多くの語は複数の語義を持つので、総数で 46,347 の語義がある。全ての語義には、1 文以上の語義文による説明が付与されており、Lexeed 全体では、81,100 文の語義文がある。全ての語義文は、辞書の中で自己完結するように、基本語のみを用いて書き換えられている。書き換えられた語義文中で最終的に使用された基本語は、全体の 60%、16,914 語であった。

図 4 に、「ドライバー」という基本語を例に Lexeed に含まれる情報と、ツリーバンク「檜」によって新しく付与される情報を示す。「檜」によって付与される情報には 下線 を引いてある。実際には、図 4 に示した情報以外に、図 2 のような MRS や、図 5 に提示するような構文木構造の情報も付与される。

ツリーバンク「檜」の構築は、Lexeed の中でも単語親密度が高い (6 以上) 見出し語の、各語義の第 1 番目の文を優先して行なっている。2003 年 12 月現在、「檜」で解析、および、正しい解析結果の選択までが終了している語義文は、約 5,000 文である。

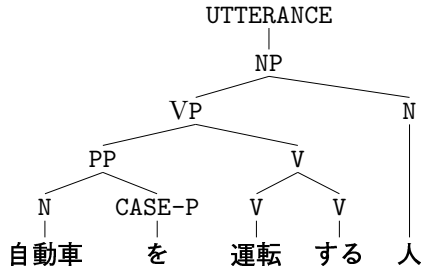

図 5: ドライバー $_2$ の解析結果（構文木）

今後も「檜」は順次拡張する。特に、Lexeed の語義文 81,100 文すべてに対して、ツリーバンクを構築する予定である。

4 知識獲得実験

ツリーバンク「檜」を用いた知識獲得の予備実験を行なった。本実験の目的は、統語情報、意味情報が付与された語義文から、見出し語間の上位下位関係をどの程度の精度で獲得できるかの検証である。また、機械翻訳、多義解消など辞書以外の分野への適用可能性を評価する目的もある。

語義文からの知識獲得の先行研究では、主に階層構造の自動生成を行なっている。例えば、鶴丸ら [19] は語義文の主節の解析に基づいたシソーラスの作成システムを開発した。上位語、下位語、同義語の関係の分類に成功したが、生成された関係の具体的な評価については述べられていない。徳永ら [20] は、機械可読辞書からオントロジを構築し既存のシソーラス [21] と組み合わせている。

本稿の提案する方法は下記の 3 点で先行研究と異なる。第 1 に、先行研究では対象が名詞に限られているのに対し、本手法では全ての品詞を扱える点。第 2 に、使用している語義文で用いられている語が制限されており、語義文に出現する全ての語が定義されている点。第 3 に、語義文から正規表現を使って知識獲得を行なうのではなく、構文解析と意味解析を行なった結果を使用する点である。

最も重要な違いは第 3 の点である。我々は詳細に定義された意味表現である MRS を使った知識獲得を行なう。MRS を用いて知識獲得を行なう理

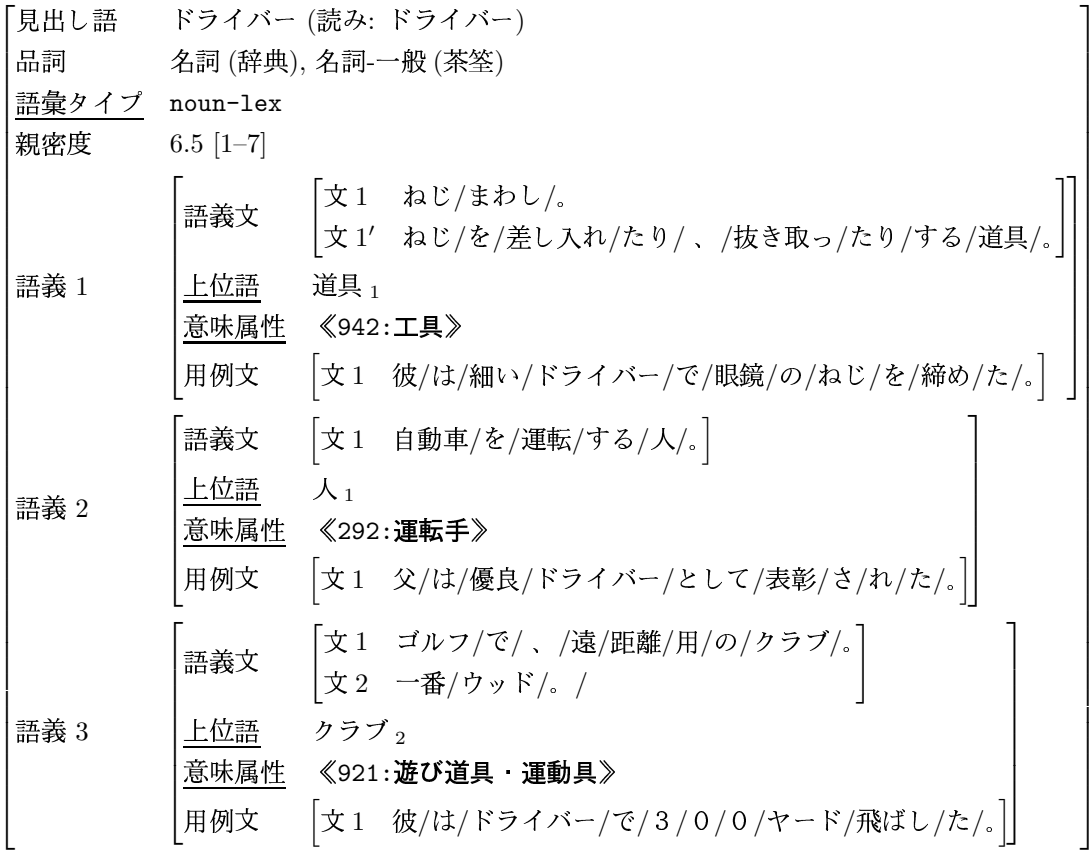

図 4: Lexeed における「ドライバー」の意味記述

由は、(i) 分野拡張が容易であること、(ii) 多言語拡張が容易であること、(iii) 獲得する知識の拡張が容易であること、という 3 点をあげることができる。(i) は、本手法では意味表現を使って知識獲得を行なうため、分野非依存であり、辞書の語義文以外の一般の文章にも適用可能ということである。これに対し、従来研究のように、正規表現によって表層的な情報から知識獲得を行なう場合には、分野毎、あるいは、文章のスタイル毎に正規表現のパターンを用意する必要がある。(ii) は、意味表現は言語非依存であるため、様々な言語に共通な枠組で知識獲得を行なえるということである。つまり、ある言語に対して MRS を出力できる解析器と辞書さえ用意できれば、アルゴリズムは容易に移植できる。(iii) は、単純な上位下位関係以外に拡張が可能ということである。例えば、土屋ら [22] が提案しているように定義表現パターンを分類することによって同義語を発見するといったこともできる。

4.1 実験方法

まず、ツリーバンク「槍」から統計モデルを学習する [6]。その統計モデルを用いて、再度、Lexeed の各語義の最初の語義文を解析する。なお、現在のところ、対象とした語義文の 87%が解析可能である。その解析結果のうち、第一候補の解析結果から MRS を得る。獲得した MRS の中で最上位のスコープを取る位置に現れる語を見出し語の上位語として獲得する。

例えば、「ドライバー」の語義 2 (文 (1)) の場合、最上位のスコープを取る位置には、「ドライバー」の上位語である「人 ($hito(x_1)$)」が現れる (図 2)。比較のため、この日本語の語義文に対応する英文の MRS を考える (文 (2)、図 3) と、最上位のスコープを取る位置には「person($person(x_1)$)」が現れている。元の文 (1)(2) では、「ドライバー」の上位語「人」と「person」は全く異なる位置に現れるが、どちらの語も MRS 中では、最上位のスコープを取っている。このように、MRS は特定の

言語に依存しないので、このような知識獲得の規則を新しい言語に対して再構築する必要はない。

語義文を利用して、更に情報を獲得することもできる。例えば、見出し語「ドライバー」(図 4 参照) の語義 3 の語義文 1、「ゴルフで、遠距離用のクラブ」を参照する。ここで、「ドライバー」の上位語は「クラブ」である、という情報以外にも、「ドライバー」という語は「ゴルフ」で用いられる用語であるというような知識獲得ができるように拡張することができる。

4.2 評価

結果の定量的な評価として、既存のシソーラスである日本語語彙大系 [21] との比較を行なった。日本語語彙大系は 2,700 の意味属性を持ち、これらの意味属性は階層化されている。本稿では、「槍」の MRS から獲得した上位語を利用し、見出し語を、日本語語彙大系と同じ意味属性に分類できるかどうかを調査した。獲得した上位語によって、異なる語義が日本語語彙大系上の正しい意味属性にマッピングできた例を表 1 に示す。

単語	上位語	意味属性
ドライバー	人	**《乗務員》**
	道具	**《道具》**
ドクター	博士	**《称号》**
	医者	**《医師》**

表 1: 上位語による語義の判別

また、本知識獲得実験では、日本語語彙大系の階層構造以上の知識を引き出せることも確かめられた。日本語語彙大系の意味属性は、本来翻訳用に構築されたものであり、統語情報と意味情報を結び付けようとする基礎研究 [23] や、質問応答のような自然言語処理システムにとって、分類が粗すぎる場合がある。特に、多くの意味属性のクラスにはクラス名と事例名が混在しており、これらを区別することが必要となる。

例えば、日本語語彙大系では「香辛料」「ソース」「スパイス」「調味料」は全て同じ意味属性《846:**調味料**》に分類されている。しかしこのような分類では、例えば質問応答システムで、「最もよく使われる調味料は何か?」という質問に対しては、「調味料」が答になり得る。また、例えば「…の料理に合うソースは何か?」といった質問に対しては、「香辛料」と「ソース」は別のクラスとして分類した方が、より適切な応答ができる。

これに対し、本知識獲得実験では、より詳細な階層構造を獲得できた (図 6)。図 6 中の語は、「材料」をのぞき、全て、日本語語彙大系では、《846:**調味料**》の意味属性を付与されている[5]。このように、本手法は、ツリーバンクからシソーラスの情報を取り出せる潜在能力を持っている。但し、「トマトケチャップ」が「ソース」に分類されているにも関わらず、「ケチャップ」は直接「調味料」配下に分類されているように、完全に整合性のとれた階層構造を獲得できているわけではない。

今後は、大規模な定量的評価を行なうことが課題である。また、全体–部分関係や同義関係など、他の関係を取り出す方法についても検討したい。

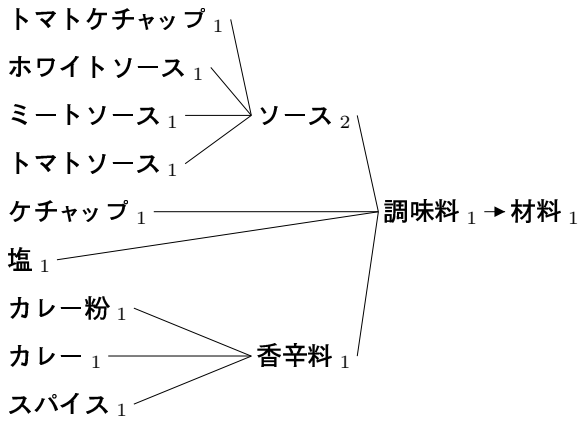

図 6: 意味属性《846:調味料》の語の階層関係

5 今後の方向性

本稿では、辞書の語義文を対象として構築し、HPSG に基づいた詳細な統語・意味情報を持つツリーバンク「槍」について述べた。また、「槍」の有効性の予備的検証として行なった知識獲得実験について報告した。今後は、第 1 段階として、Lexeed の全語義文約 8 万文のツリーバンクの構築を完了させる。ツリーバンクの構築完了後、全ての語義文を用いて統計モデルを再構成し、新しい文法を使ってオントロジの上位下位関係の獲得実験を行い、大規模な定量的評価を行う。

第 2 段階では、第 1 段階で獲得した上位語の知識を統計モデルに導入する。また、解析された語

[5]図 6 に示した語は、「槍」で解析した語のうち《846:**調味料**》に含まれる語の一部である

義文から、意味的な関連度や語彙タイプなど他の情報を獲得するシステムを構築することを目指す[24]。

第 3 段階では、第 2 段階で構築したシステムを使って、Lexeed に収録されていない語や、辞書以外の分野に統計モデルとオントロジを拡張する。また、意味内容の記述である MRS を経由した文章の生成、言い替え、翻訳、英語等他の言語との特徴比較などを行なう予定である。

参考文献

[1] 笠原要, 佐藤浩史, Francis Bond, 田中貴秋, 藤田早苗, 金杉友子, 天野昭成. 「基本語意味データベース:lexeed」の構築. In *2003-NLC-159*, 1/13-1/14 2004.

[2] Sadao Kurohashi and Makoto Nagao. Building a Japanese parsed corpus — while improving the parsing system. In Anne Abeillé, editor, *Treebanks: Building and Using Parsed Corpora*, chapter 14, pp. 249–260. Kluwer Academic Publishers, 2003.

[3] EDR. Concept dictionary. Technical report, Japan Electronic Dictionary Research Institute, Ltd, April 1990.

[4] Kristina Toutanova, Christoper D. Manning, and Stephan Oepen. Parse ranking for a rich HPSG grammar. In *Proceedings of The First Workshop on Treebanks and Linguistic Theories (TLT2002)*, Sozopol, Bulgaria, 2002.

[5] Chiori Hori, Takaaki Hori, Hideki Isozaki, Eisaku Maeda, Shigeru Katagiri, and Sadaoki Furui. Deriving disambiguous queries in a spoken interactive ODQA system. In *ICASSP-2003*, pp. 624–627, 2003.

[6] Francis Bond, 藤田早苗, 橋本力, 成山重子, Eric Nichols, 大谷朗, 田中貴秋. 精細な文法に基づいたツリーバンク「槍」の構築. In *2003-NLC-159*, 1/13-1/14 2004.

[7] Carl Pollard and Ivan A. Sag. *Head Driven Phrase Structure Grammar*. University of Chicago Press, Chicago, 1994.

[8] Takao Gunji. *Japanese Phrase Structure Grammar: A Unification-Based Approach*. D. Reidel (Kluwer), Dordrecht, 1987.

[9] 増市博, 大熊智子. Lexical functional grammar に基づく実用的な日本語解析システムの構築. 自然言語処理学会論文誌, Vol. 10, No. 2, pp. 79–109, 2003.

[10] Melanie Siegel and Emily M. Bender. Efficient deep processing of japanese. In *Procedings of the 3rd Workshop on Asian Language Resources and International Standardization at the 19th International Conference on Computational Linguistics*, Taipei, 2002.

[11] Wolfgang Wahlster, editor. *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer, Berlin, Germany, 2000.

[12] Dan Flickinger. On building a more efficient grammar by exploiting types. *Natural Language Engineering*, Vol. 6, No. 1, pp. 15–28, 2000. (Special Issue on Efficient Processing with HPSG).

[13] Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A. Sag. Minimal recursion semantics: An introduction. (manuscript `http://www-csli.stanford.edu/~aac/papers/newmrs.ps`), 1999.

[14] Ann Copestake. *Implementing Typed Feature Structure Grammars*. CSLI Publications, 2002.

[15] Ulrich Callmeier. PET - a platform for experimentation with efficient HPSG processing techniques. *Natural Language Engineering*, Vol. 6, No. 1, pp. 99–108, 2000.

[16] Stephan Oepen and John Carroll. Performance profiling for grammar engineering. *Natural Language Engineering*, Vol. 6, No. 1, pp. 81–97, 2000.

[17] 天野成昭, 近藤公久. 日本語の語彙特性. 三省堂, 東京, 1999.

[18] 金杉友子, 笠原要, 稲子希望, 天野昭成. 単語親密度に基づく基本的語彙の選定築. 第 NLC2002 巻, pp. 21–26, 2002.

[19] 鶴丸弘昭, 竹下克典, 伊丹克企, 柳川俊英, 吉田将. 国語辞典情報を用いたシソーラスの作成について. 情報処理学会自然言語処理研究会, 第 83-16 巻, pp. 121–128, 1991.

[20] Takenobu Tokunaga, Yasuhiro Syotu, Hozumi Tanaka, and Kiyoaki Shirai. Integration of heterogeneous language resources: A monolingual dictionary and a thesaurus. In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium, NLPRS2001*, pp. 135–142, Tokyo, 2001.

[21] 池原悟, 宮崎雅弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦. 日本語語彙大系. 岩波書店, 1997.

[22] Masatoshi Tsuchiya, Sadao Kurohashi, and Satoshi Sato. Discovery of definition patterns by compressing dictionary sentences. In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium, NLPRS2001*, pp. 411–418, Tokyo, 2001.

[23] Francis Bond and Caitlin Vatikiotis-Bateson. Using an ontology to determine English countability. In *19th International Conference on Computational Linguistics: COLING-2002*, Vol. 1, pp. 99–105, Taipei, 2002.

[24] Martin Hoelter. *Lexical-Semantic Information in Head-Driven Phrase Structure Grammar and Natural Language Processing: Retrieval of lexical-semantic information from Cobuild-style dictionaries*. Lincom Europa, 1999.