課題研究報告書

述語項構造解析に関する調査研究

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

山岸 博幸

2012 年 12 月

課題研究報告書

述語項構造解析に関する調査研究

指導教員　島津 明 教授

審査委員主査　島津 明 教授
審査委員　東条 敏 教授
審査委員　白井 清昭 准教授

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

0910904 山岸 博幸

提出年月 2012 年 11 月

概要

文章において述語がある事象を記述するためには、「○○が××を破壊する」「△△が美しい」などの形をとる必要があり、そこでは必然的に名詞の参画が必要となる。この述語を中心としてそれに関連する名詞が連なるという構造は、文における意味上の骨子であり、それを正しく抽出することができれば計算機が文の意味を扱う上で非常に有益となる。このような構造を述語項構造（Predicate-Argument Structure）と呼び、入力された文章から述語項構造を抽出するタスクを述語項構造解析と呼ぶ。

本稿の目的は、自然言語処理における重要なタスクである述語項構造解析についてのサーベイを行うことである。近年、述語項構造解析に対する注目は高まっており、自然言語処理の評価型ワークショップ Conference on Computational Natural Language Learning（CoNLL）において共通評価タスクに指定されるなど研究がさかんに行われている。そのこともあり、広範囲かつ大量の研究成果が報告されている。述語項構造解析は、入力文中の単語に対して適切なタギングを行うタスクであると一般化できるが、付与されるタグは大きく分けて表層格と深層格の２つに分かれる。表層格とは文中の語が文において担う文法的な役割のことを言い、深層格とは動作主格や対象格といった意味的な役割のことを表す。一般的にこれら表層格と深層格は一対一に対応しない。

深層格は意味的な情報を表現するため、これを正しく文章に付与することにより質問応答（question answering）、機械翻訳（machine translation）、文書要約（text summarization）、情報抽出（information extraction）などの処理の精度向上が期待できる。

深層格解析は、研究初期においては手動で作成したルールをベースに付与するシステム［49］や、MindNet のような言語資源を用いた研究が行われていた［114］。しかし、1998 年に大規模な意味情報が付与されたコーパスである FrameNet が登場したことと、それを用いて 2002 年に Gildea らによって報告された機械学習に基づく統計的自動解析手法が、研究史におけるエポックメイキングな出来事となっている。2005 年に Penn TreeBank に意味情報を付与した Propositon Bank（PropBank）が登場したことも、ルールベースから統計ベースへのパラダイムシフトを後押しした。このような流れの中で、深層格解析ではなく Gildea らが用いた意味役割付与（Semantic Role Labeling）という名でこのタスクは呼ばれるようになっている［92］。

一方で、動詞や形容詞以外にも事態を表す事態性名詞と呼ばれる名詞があることが知られており、この事態性名詞の述語項構造解析を扱った Gerber ら［37］の論文が ACL

ベストペーパーに選ばれるなど、近年重要なトピックとして注目されている。

このような研究史の流れを踏まえ、本稿では述語項構造解析において主要なタスクである表層格付与、深層格（意味役割）付与、および近年主要なトピックとなっている事態性名詞の解析につき研究事例を概括しながらこれらの研究の現状について整理した。

目次

第１章　はじめに

自然言語処理研究の中心にあるのは、自然言語が持つ曖昧性をいかに解消するかという問題である。例えば「橋の上で泳ぐ少女を見た」という文章においては「川で泳いでいる少女を橋で見た」という解釈と、「少女が橋の上で泳いでいて、その光景を見た」という二通りの解釈があり得る。もちろん、人間であれば、橋の上で泳ぐという行為が常識的にはあり得ないという共通認識を持っているため後者の解釈を退けることができるが、計算機は文字通り機械的に後者の解釈を選び得る。あるいは「涙を飲む」という慣用表現において、「飲む」という動詞が「水を飲む」という形とは異なる意味で用いられていることを人間は理解することができるが、計算機はそのような区別ができず、水を飲むが如く涙を飲み干すのだという文字通りの解釈をし得る。

人間は文章の表層的な構造だけからは知ることができない「意味」を理解するが、計算機はそれを理解しないということがこのような状況が生じる原因である。故に、計算機に意味を理解させるという研究は、自然言語処理の重要な研究テーマとなっている。

人間が考える文章の意味とは、おそらく頭に浮かぶ視覚的・感覚的イメージであろう。あるいはその文章によって惹起された過去の記憶かも知れない。いずれにしても、イメージという行為を行わない計算機にとっては、そのような形での意味は到底理解できない。

そのため計算機における文章の意味を定義する必要がある。

文章はある状況を記述しているため、その骨子は述語にある。そして述語がある事象を記述するためには、「○○が××を破壊する」「△△が美しい」などの形をとる必要があり、そこでは必然的に名詞の参画が必要となる。この述語を中心としてそれに関連する名詞が連なるという構造は、文における意味上の骨子を成す。

言語学者チャールズ・フィルモアが提唱した述語項構造（格文法）と呼ばれるこの構造が、計算機における意味とするに都合がよいと研究者の間でコンセンサスを得ることとなった。

計算機における文章の意味とは、その文章が持つ述語項構造のことである。

意味をこのように定義することによって、意味を解釈するという哲学的な問題を純粋な工学的問題に帰着させることが可能になった。すなわち、計算機における意味の解釈とは、入力された文章の述語項構造を判別することである。

述語項構造の情報を利用することの有効性は、機械翻訳［36, 41, 147］、質問応答［121］、含意関係認識［145］、情報抽出［129］といった自然言語処理の様々な応用領域において示されている。

本研究報告の目的は、世界中で広範に行われている述語項構造解析の研究を概観し、現時点での到達点と未解決の問題を整理することで、今後の研究の土台となるサーベイを作成することである。

本研究報告の構成は以下の通りである。

第 2 章では、述語項構造解析の概要および解析に必要な言語資源である格フレーム辞書とタグ付きコーパスについて概観する。

第 3 章では、タグ付きコーパスのうち主流となっているものを整理する。多くの述語項構造解析においては言語資源が必要となるが、それは格フレーム辞書とタグ付きコーパスに大別される。近年、深層格解析（意味役割付与）の分野においてはタグ付きコーパスを利用するものが多く、そのためか多くの種類のものが世に存在している。

第 4 章においては、格フレーム辞書の自動構築研究について整理する。言語資源は基本的に人手で構築するため、構築に要する人的・費用的コストが高い。そのため、それを自動で構築するという試みが行われており、特に格フレーム辞書については研究が多く発表されている。

第 5 章では、表層格を付与する研究について整理する。述語項構造解析は、表層格の付与を目的とするものと深層格の付与を目的とするものとに大きく分かれる。この章ではそれらのうち表層格を付与する研究について整理する。

第 6 章では、深層格を付与する研究について整理する。深層格付与は 2002 年の Gildea らのエポックメイキングな成果により、意味役割付与と呼ばれることが多い。意味役割付与においては機械学習を用いる手法が一般的である。機械学習においてはどのアルゴリズムを用いるかということも重要であるが、それ以上に素性の選択が重要となる。6.3 節においてそれらを概観する。

第 7 章では、近年研究が進められている事態性名詞の解析について概観する。動詞や形容詞などは事態を表現するが故に項構造を考慮するが、名詞の一部にも事態を表現するものがある。例えば「彼は上司の推薦で抜擢された」という文において、名詞「推薦」は「上司が彼を推薦する」という事態を表す。このような名詞は事態性名詞と呼ばれ、近年注目が集まっている。

第 8 章において、本研究報告のまとめと今後の課題を述べる。

第 2 章　述語項構造解析について

2.1　述語項構造解析概要

述語項構造解析とは、処理対象となる自然言語で書かれたテキストから述語項構造を判別する処理である。述語項構造とは、述語を中心として自然言語の文を捉える構造であり、述語とそれが持つ項、および項が持つ属性値からなる。述語項構造解析処理によって判別された述語項構造は文中の語彙にタグとして付与されることが多い。すなわち、述語項構造解析処理のインプットは生のテキストであり、アウトプットはそのテキストに述語項構造のタグが付与されたものである。直感的に言えば、述語と判定された語にはそれが述語であることを表すタグが、項と判定された語にはその項が持つ属性値がタグとして付与される。

述語がある事象を記述するためには、「○○が××を破壊する」「△△が美しい」などの形をとる必要があり、必然的に名詞の参画が必要となる。この名詞のことを述語項構造においては「項」と呼ぶ。項に種別があることは容易にわかる。例えば、

(1) 彼はコンビニで弁当を買った。

という文章においては「彼」「コンビニ」「弁当」が述語「買う」の項となるが、それぞれの意味は異なる。「彼」は動作の主体を表す項であり、「コンビニ」は場所を「弁当」は対象を表す項である。それぞれの項が文章内で担う役割を「意味役割（Semantic Role）」と呼ぶ。この意味役割は、述語項構造解析によって項に付与される重要な情報のひとつである。

上記で述べた構造は、かつて格構造（Case Structure）と呼ばれていたものと同じである。1968 年に言語学者の C.Fillmore [31, 161] によって提唱された格文法（Case Grammer）と呼ばれる言語学の枠組みの中で用いられる構造である。動詞が述語としてある事象を記述するためには、「何が何をどうする」という形をとる必要があり、必然的に名詞の参画が必要となる。格文法においては、動詞がその語彙的な意味を実現するために名詞との間に構築する関係は「格（Case）」と呼ばれる。格には表層格と深層格、必須格と任意格の区別がある。

表層格とは文の見た目の構造から決まる格である。英語の場合は統語構造から決定される。主語の位置にある語には「主格」、目的語の位置にある語には「目的格」などと

いった形で格が付与される。日本語の場合は格助詞によって格が決定され、ガ格やヲ格といったものがそれにあたる。それに対して、深層格は見た目からは決まらない格であり真の格を表現するとされている。どのような深層格を想定するかは研究者によって意見が分かれている。表 2.1 に例として Fillmore が想定した深層格を挙げる。

表 2.1: Fillmore が想定した深層格

Agent	動作主格	動作を引き起こすもの
Experiencer	経験者格	心理現象を体験するもの
Instrument	道具格	動作を起こさせるもの
Object	対象格	動作が作用する対象となるもの
Source	源泉格	移動における起点を表すもの
Goal	目標格	移動における終点を表すもの
Location	場所格	動作が起こる場所や位置を表すもの
Time	時間格	動作が起こる時間を表すもの

必須格とは述語が状況を記述する上で必要欠くべからざる格であり、任意格とはそうではない格である。例えば (1) の文においては、述語「買う」がその役割を果たす上で「彼（が)」（ガ格）や「弁当（を)」（ヲ格）は必須の格となるが、「コンビニ（で)」（デ格）という場所を指定する格は必須ではない。

なお、述語項構造解析の枠組みにおいては、深層格は「意味役割」と呼称されることが多いが、研究によっては「深層格」と呼称する例もあり、用語は必ずしも統一されていない。

述語項構造を付与されたコーパスからタグ付与の例を見る。例えば PropBank において「A year earlier the refiner earned $66 million, or $1.19 a share.」という文章には以下のようなタグが付与されている [218] 。

- [ARGM-TMP *A year earlier*], [ARG0 *the refiner*] [rel *earned*] [ARG1 *$66 million, or $1.19 a share*].

ここで「ARGM-TMP」「ARG0」「rel」「ARG1」は、PropBank で定義された意味役割を表すタグである。

日本語コーパスの例として、例えば NAIST テキストコーパスにおいては表層格レベルでのタグ付与（ガ格、ヲ格、ニ格）が行われている。

・私 *i* は彼 *j* にリンゴ *k* を食べさせる *ガ:i, ヲ:k, ニ:j*

同コーパスには述語や事態性名詞と表層格との関係や、名詞句間の共参照関係や照応関係といった情報が付与されているが、深層格レベルでのタグは付与されていない。

2.2 言語資源について

述語項構造解析を行うに当たっては言語資源が必要となる。利用する言語資源によって解析手法は変わる。言語資源は大きく分けて格フレーム辞書とタグ付きのコーパスの二つが存在する。

2.2.1 格フレーム辞書

格フレーム辞書とは、述語がどのような格を持つかということを述語単位にまとめたものである。格フレームには大きく分けて、意味素に基づくものと用例に基づくものの二種類が存在する。

2.2.1.1 意味素に基づく格フレーム

意味素（semantic primitive / feature）とは、名詞を分類ごとに包括する概念のことである。例えば「犬」「馬」といった名詞は動物であることを示す<animal>という意味素が付与され、「姉」「先生」などには<human>といった意味素が付与される。意味素に基づく格フレームには、日本語のものとしては IPA 計算機用日本語動詞辞書がある[181]。

「かける」1（希望・期待などをもつ） *<human/organization>* が　*<human>* に　*<abstract>* を （動作主格）　（目標格）　（対象格）
「かける」2（物をぶら下げる） *<human>* が　*<location/concrete>* に　*<concrete>* を （動作主格）　（目標格）　（対象格）
「かける」3（液状・粉末状の物を付着させる） *<human/animal>* が　*<concrete>* に　*<concrete>* を （動作主格）　（場所格）　（対象格）

図 2.1:「かける」の意味素に基づく格フレーム [208]

「大学が彼に期待を掛けた」という入力文を処理する場合、「大学」に<organization>と<location>、「彼」に<human>、「期待」に<abstract>という意味素が与えられていることがわかれば、上記の格フレーム群と比較して意味素の整合性から 1 番目の格フレームが入力文に対して適切であると選択される。そこから「掛ける」の意味、動詞に対して「大学」が動作主格、「彼」が目標格、「期待」が対象格であることがわかる [208]。

2.2.1.2　用例に基づく格フレーム

意味素に基づく格フレームに対し、実際の文に対して表層格・深層格を付与したデータを用いて作成した格フレームがある。

「かける」1（希望・期待などをもつ） *[企業, 彼]* が　*[社員, 息子]* に　*[希望, 期待]* を （動作主格）　（目標格）　（対象格）
「かける」2（物をぶら下げる） *[彼]* が　*[壁, ハンガー]* に　*[絵, 服]* を （動作主格）　（目標格）　（対象格）
「かける」3（液状・粉末状の物を付着させる） *[彼, 犬]* が　*[花, 街路樹]* に　*[水, おしっこ]* を （動作主格）　（場所格）　（対象格）

図 2.2:「かける」の用例に基づく格フレーム [208]

入力文を処理する際は、その入力文に対して最も類似した用例を持つ格フレームを適切な格フレームとして選択する。それぞれの格の語ごとに入力文と各用例の語の類似度を求め、すべての語の類似度の合計を入力文と格フレームの類似度と定義する。そして最も類似度の高い格フレームをその入力文に対応する格フレームとして選択する。

ここでは語と語の類似度をどのように定義するかが問題となるが、一般的に語の類似度はシソーラス中での二語の近さという形で定義される。シソーラスは木構造上に単語を分類配列したものであり、意味的に近い語ほど近くに配置されているという特徴を持つ。そこで二つの語につき、シソーラスにおけるそれらの語の共通の上位ノードの深さがそれらの語の間の類似度を示していると考えることができる。図 2.3 の場合には、「記述」と「叙述」の共通の上位ノードは深さ 6 であるのでこれらの語の類似度は 6、「記述」と「発言」の共通の上位ノードは深さ 3 であるので類似度は 3 になる。

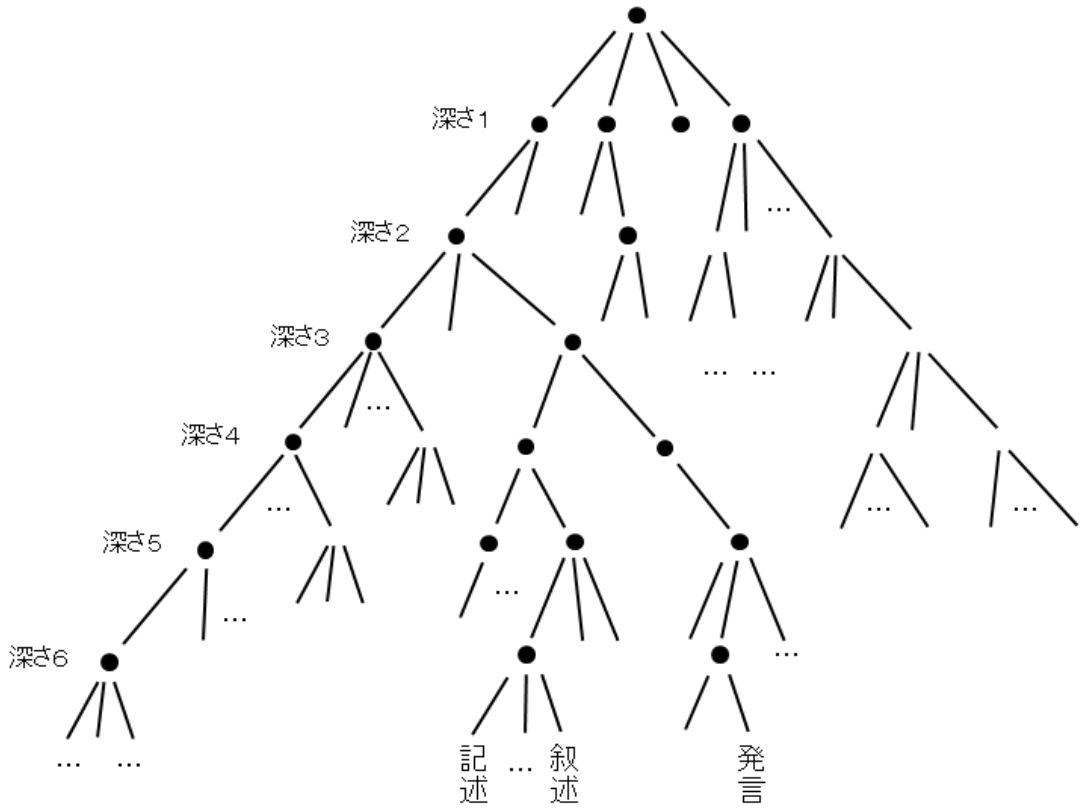


図 2.3: シソーラスの階層構造

上記の議論は木構造の葉の部分だけに実際の語が存在するタイプのシソーラスについてのものであり、シソーラスには木構造の内部ノードにも語が対応させ、語の上位下位関係を表現するタイプのものもある。そのようなシソーラスを用いた場合の語の類似度の求め方として、二つの語の深さをd_i, d_j、それらの共通の上位語の深さをd_cとしたとき、

$$\frac{d_c \times 2}{d_i + d_j}$$

という値を二語の類似度とする方法がある。基本的な考え方は先のものと同じであるが、共通の親ノードの深さが同じ場合でも語自体の深さが深い場合は類似度が低くなるよう考慮されている。

2.2.2 タグ付きコーパス

先に見た PropBank や NAIST テキストコーパスのように、すでに人手による述語項構造解析の結果としてタグが付与されたコーパスが存在する。このようなコーパスをベースにして、コーパスに掲載されていない未知の文章に対して述語項構造解析を試みる研究がある。タグ付きのコーパスについては様々なものが構築、公開されている。各コーパスの詳細については、3 章で扱う。

表 2.2: 主なタグ付きコーパスの一覧

年度	言語	名称	付与タグ	ベースのテキスト
1985	英語	Word Net [30, 85, 86]	・品詞 ・synset ごとの階層関係	（特になし）
1993	英語	Penn Treebank [78]	・品詞 ・構文情報 ・意味役割	・Wall Streat Journal（130 万語） ・Brown Corpus（100 万語） ・Switchboard Corpus（100 万語）
1997	日本語	京都テキストコーパス	・品詞 ・構文情報 ・格関係 ・照応・省略 ・共参照	・新聞記事 （毎日新聞 40,000 文）
1998	英語	Frame Net [1]	・意味役割	・British National Corpus（100 万語）
2000	英語	VerbNet [61, 63]	・構文情報 ・意味役割	・動詞クラス 274 種類 ・動詞の意味 5,257 種類
2000	日本語	日本語話し言葉コーパス [75, 183, 202, 214]	・形態素 ・品詞 ・節境界 ・文節音	・自発音声データ （約 700 万語）
2002	英語	PropBank [59, 60, 94]	・品詞 ・構文情報 ・意味役割	・Penn Treebank
2004	英語	NomBank [80, 81]	・意味役割	・Penn Treebank
2004	日本語	新聞記事 GDA コーパス 2004 [177]	・統語構造 ・語義 ・照応 ・共参照	・新聞記事 （毎日新聞 37,000 文）
2004	日本語	岩波国語辞典第五版タグ付き	・統語構造	・岩波国語辞典第五版

		コーパス 2004 [226]	・語義 ・照応 ・共参照	（56,000 項目）
2005	英語 中国語	OntoNotes [97, 104, 113]	・構文情報 ・意味役割	・英語 130 万語 ・中国語 80 万語 ・アラビア語 30 万語
2007	日本語	NAIST テキストコーパス [217]	・品詞 ・構文情報 ・格関係 ・照応・省略 ・共参照	・京都テキストコーパス
2009	日本語	KNB コーパス [180]	・形態素 ・構文 ・格関係 ・照応・省略 ・評判情報	・ブログ記事 （4,206 文）
2011	日本語	現代日本語書き言葉均衡コーパス [184, 196]	・形態素 ・品詞	・新聞、雑誌 Web 記事等から約 1 億語

タグ付きコーパスの構築には多大な人手と時間を要する上、作業者によってタグ付与の基準が統一されないなどの問題も発生し得る。特に意味役割の付与については作業者ごとの合意の低さ（IAA : low interannotator agreement）は、よく知られた問題である。Cinková ら [22] は、WordNet、FrameNet、PropBank といった主要なコーパスのタグ付与方法を比較した上で、semantic pattern recognition という手法を提案した。

Basile（2012）らはタグ付与システムそのものに着目した。タグ付与者がタグ付与の際に利用するシステムには GATE [28] 、NITE [14] 、UIMA [47] といったものがあるが、それらが多くのタグ付与者が共同で作業を行うために必要な機能を欠いていると指摘し、Wiki ベースの新しいシステムを作成した。Wiki ベースであるために編集履歴が残り、複数のタグ付与者がタグの修正を行うのに都合が良いとしている [3]。

第 3 章　コーパスについて

3.1　WordNet

英語の語（名詞、動詞、形容詞、副詞）を synsets（synonym sets）という同義語のグループに分類し、それぞれの synsets の間に hyponymy（関連付けられた語が意味的に包含関係を持つ関係）、antonymy（関連付けられた語がそれぞれ反対の意味を表す関係）、meronymy（関連付けられた語が部分と全体を成す関係）という関係を定義した概念辞書である［30, 85, 86］。所収する語は ver3.0 において約 15 万語である。

プリンストン大学のホームページにおいて、WordNet の Web インタフェースが公開されている［106］。

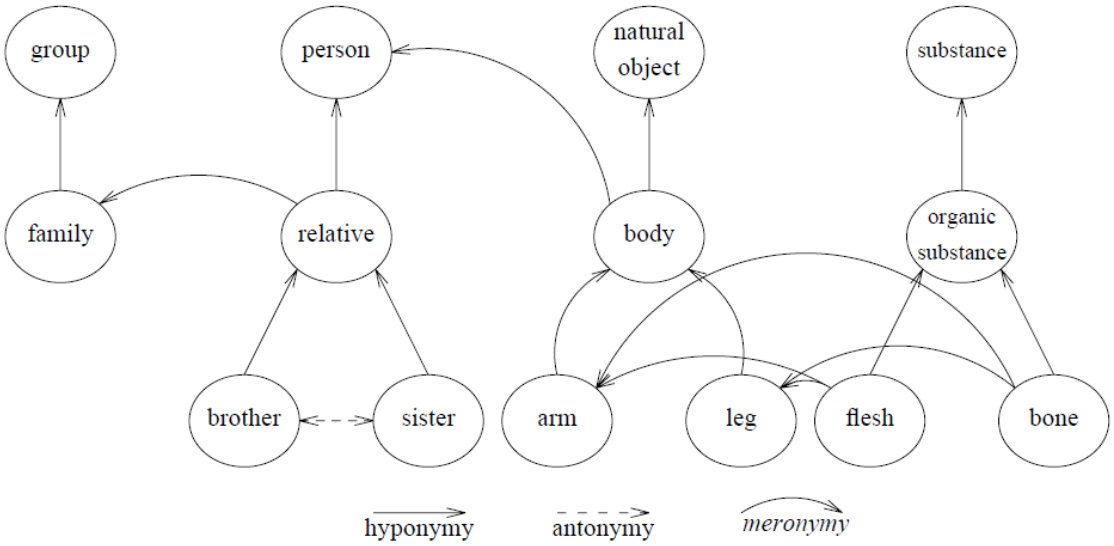


図 3.1: WordNet における hyponymy、 antonymy、 meronymy の関連付け［86］

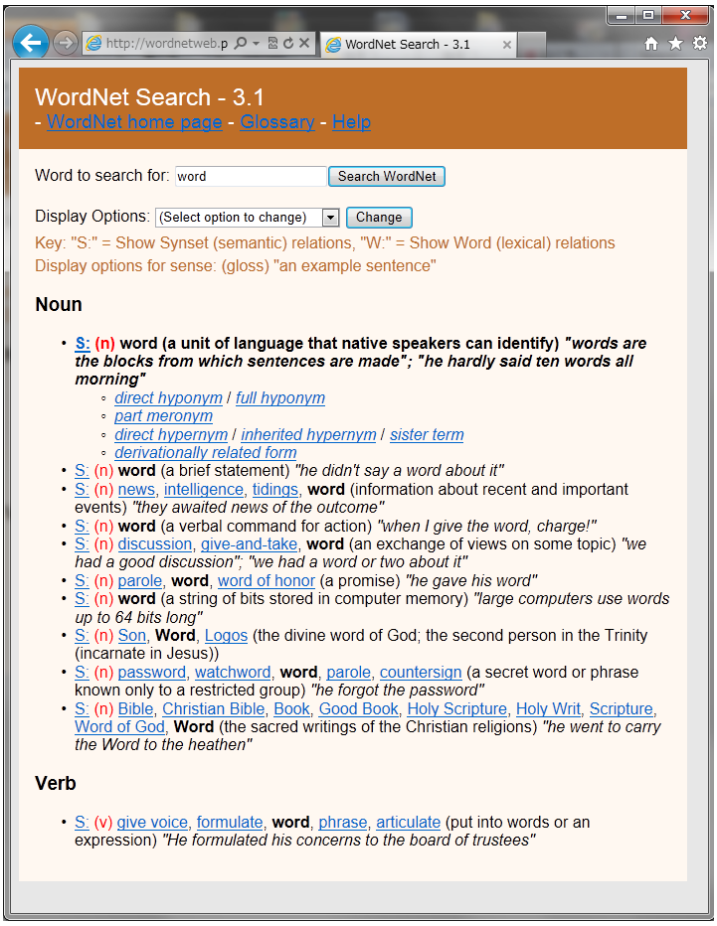


図 3.2: WordNet の Web インタフェース [106]

3.2 Penn TreeBank

Marcus ら [78] ペンシルベニア大学の研究者によって作成されたコーパス。ブラウンコーパス約 100 万語、ウォールストリートジャーナル約 130 万語、電話交換記録約 100 万語に対して品詞、構文情報、意味役割タグの付与が行われている。

タグの付与は 2 つの時期に分かれて行われており、1989 年から 1992 年までの間には、品詞と構文情報が付与された。品詞と構文情報の付与は、Hindle が作成したパーサである Fidditch によってテキストを機械的に解析した結果に対して手動で補正を加える形で行われた。構文情報は文脈自由文法で表現されるが、これは Lancaster Treebank コーパスを参考にしている。1993 年から行われた 2 期目のタグ付与においては、表 3.1 の意味役割が付与された。

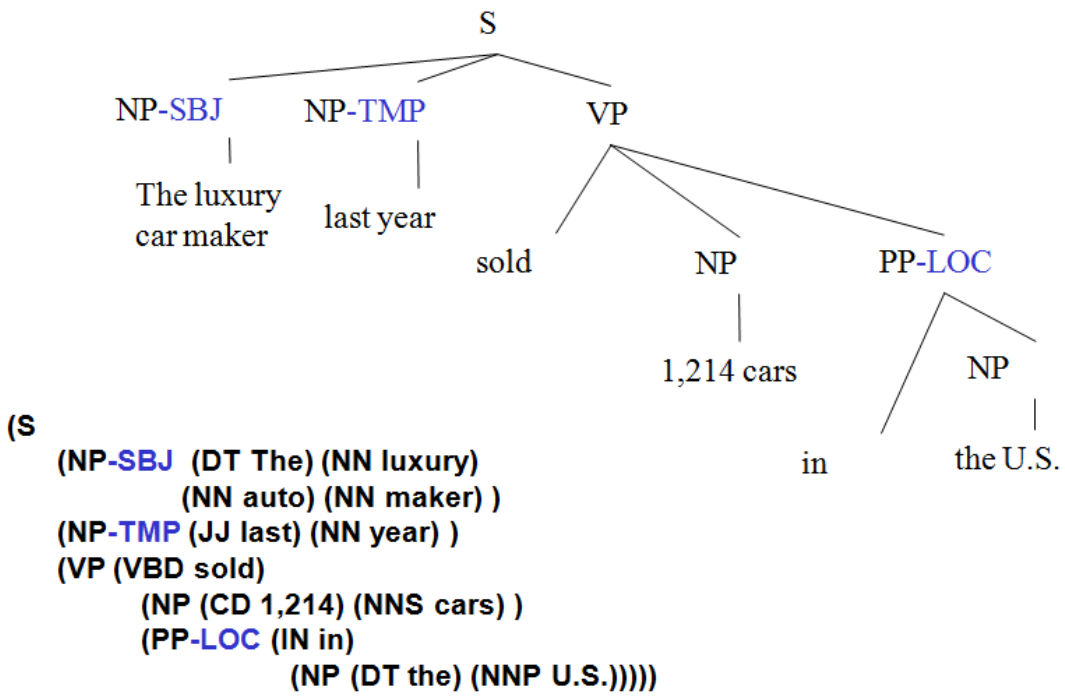


図 3.3: Penn Treebank のタグ付与イメージ [79]

表 3.1: Penn Treebank の意味役割タグ

タグ	意味
TMP	Temporal
LOC	Location
DIR	Direction
EXT	Extent
BNF	Benefactive
MNR	Manner
PRP	Purpose
NOM	Nominal
ADV	Non-specific adverbial

3.3 京都テキストコーパス

毎日新聞の記事約 40,000 文をベースにしたコーパス。作成当初は形態素・構文情報のみが付与されていたが、現在のバージョン 4.0 においては、所収されている 40,000 文のうち 5,000 文に対して、用言・サ変名詞に対する格関係、名詞間の関係、共参照タグなども付与されている [167] 。形態素・構文情報については、形態素解析システム JUMAN [175] と構文解析システム KNP [176] を使った自動解析の結果を手で修正して付与している。

```
# S-ID:950101001-001
    * 0 2D
    + 0 3D
    太郎 たろう * 名詞 人名 * *
    は は * 助詞 副助詞 * *
    * 1 2D
    + 1 2D
    東京 とうきょう * 名詞 固有名詞 * *
    + 2 3D
    大学 だいがく * 名詞 普通名詞 * *
    に に * 助詞 格助詞 * *
    * 2 -1D
    + 3 -1D <rel type="ガ" target="太郎" sid="950101001-001" tag="0"/>
            <rel type="ニ" target="大学" sid="950101001-001" tag="2"/>
    行った いった 行く 動詞 * 子音動詞カ行促音便形 基本形
    EOS
```

図 3.4: 京都テキストコーパスのデータフォーマット

3.4 FrameNet

FrameNet はコーパスに対して Fillmore のフレーム意味論に基づいた意味情報を付与する形で作成されている [1, 32] 。この枠組みにおいては、文中の単語（主として動

詞）がフレームと呼ばれる特定の項構造を持つと考えられる。図 3.1 において、例えば動詞 hit は Hit_target というフレームを持ち、それによって Agent と Target を必須の項として持つことがわかる。

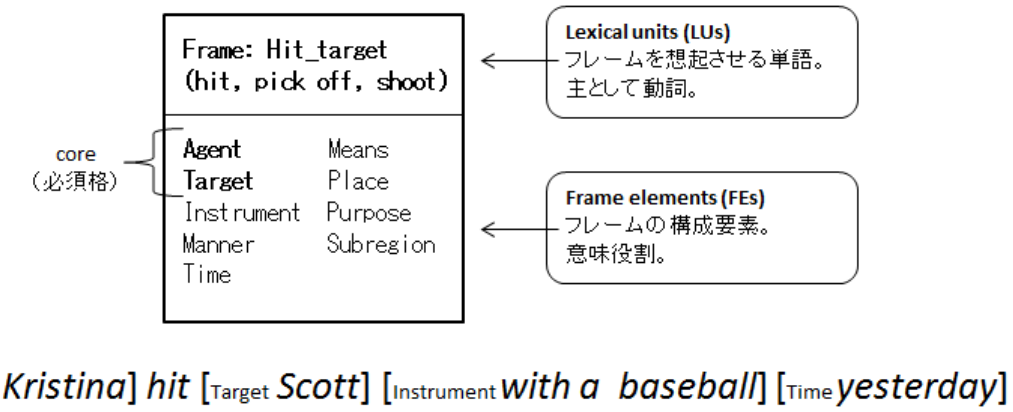


図 3.5: FrameNet の概念図とタグ付きテキストの例［156］

FrameNet I においては British National Corpus (BNC)に所収された文章のみを用いていたが、FrameNet II においては LDC North American Newswire corpora の文章も追加されている。約 8,900 個の lexical units、約 625 個のフレーム、約 135,000 文から構成される。

3.5 VerbNet

VerbNet は Levin による動詞分類から構築された辞書である［61, 63］。各動詞のクラスに対し、そのクラスに含まれる動詞、文法、文例、意味役割が記述されている。動詞クラス「Hit-18.1」の記述例を図 3.2 に示す。
PropBank のフレームセットと VerbNet のクラスとの間のマッピングを行うことで、カバレッジを向上させる研究も行われている［62］。

所収されている動詞クラスは、バージョン 1.0 時点で 191 種類、最新版であるバージョン 3.1 においては 274 種類である。所収されている動詞の意味は、バージョン 1.0 時点で 4,656 種類、バージョン 3.1 においては 5,257 種類である。これは PropBank に所収された動詞の 90.86%をカバーする［96］。

Class Hit-18.1			
Roles and Restrictions: Agent[+int_control] Patient[+concrete] Instrument[+concrete]			
Members: bang, bash, hit, kick, ...			
Frames:			
Name	Example	Syntax	Semantics
Basic Transitive	Paula hit the ball	Agent V Patient	cause(Agent, E) ∧ manner(during(E), directedmotion, Agent) ∧ ¬contact(during(E), Agent, Patient) ∧ manner(end(E),forceful, Agent) ∧ contact(end(E), Agent, Patient)
Conative	Paul hit at the window	Agent V at Patient	cause(Agent, E) ∧ manner(during(E),directedmotion,Agent) ∧ ¬contact(during(E), Agent, Patient)

図 3.6: VerbNet の動詞クラス Hit-18.1 の例 [96]

表 3.2: VerbNet の意味役割タグ [96]

Actor:	used for some communication classes (e.g., Chitchat-37.6, Marry-36.2, Meet-36.2) when both arguments can be considered symmetrical (pseudo-agents).
Agent:	generally a human or an animate subject. Used mostly as a volitional agent, but also used in VerbNet for internally controlled subjects such as forces and machines.
Asset:	used for the Sum of Money Alternation, present in classes such as Build-26.1, Get-13.5.1, and Obtain-13.5.2 with `currency' as a selectional restriction.
Attribute:	attribute of Patient/Theme refers to a quality of something that is being changed, as in (The price)att of oil soared. At the moment, we have only one class using this role Calibratable cos-45.6 to capture the Possessor Subject Possessor-Attribute Factoring Alternation. The selectional restriction `scalar' (defined as a quantity, such as mass, length, time, or temperature, which is completely specified by a number on an appropriate scale) ensures the nature of Attribute.
Beneficiary:	the entity that benefits from some action. Used by such classes asBuild-26.1, Get-13.5.1, Performance-26.7, Preparing-26.3, and Steal-10.5. Generally introduced by the preposition `for', or double object variant in the benefactive

	alternation.
Cause:	used mostly by classes involving Psychological Verbs and Verbs Involving the Body.
Location, Destination, Source:	used for spatial locations
Destination:	end point of the motion, or direction towards which the motion is directed. Used with a `to' prepositional phrase by classes of change of location, such as Banish-10.2, and Verbs of Sending and Carrying. Also used as location direct objects in classes where the concept of destination is implicit (and location could not be Source), such as Butter-9.9, or Image impression-25.1.
Source:	start point of the motion. Usually introduced by a source prepositional phrase (mostly headed by `from' or `out of'). It is also used as a direct object in such classes as Clear-10.3, Leave-51.2, and Wipe instr-10.4.2.
Location:	underspecified destination, source, or place, in general introduced by a locative or path prepositional phrase.
Experiencer:	used for a participant that is aware or experiencing something. In VerbNet it is used by classes involving Psychological Verbs, Verbs of Perception, Touch, and Verbs Involving the Body.
Extent:	used only in the Calibratable-45.6 class, to specify the range or degree of change, as in The price of oil soared (10%)ext. This role may be added to other classes.
Instrument:	used for objects (or forces) that come in contact with an object and cause some change in them. Generally introduced by a `with' prepositional phrase. Also used as a subject in the Instrument Subject Alternation and as a direct object in the Poke-19 class for the Through/With Alternation and in the Hit-18.1 class for the With/Against Alternation.
Material and Product:	used in the Build and Grow classes to capture the key semantic components of the arguments. Used by classes from Verbs of Creation and Transformation that allow for the Material/Product Alternation.
Material:	start point of transformation.
Product:	end result of transformation.
Patient:	used for participants that are undergoing a process or that have been affected in some way. Verbs that explicitly (or implicitly) express changes of state have Patient as their usual direct object. We also use Patient1 and Patient2 for some classes of Verbs of Combining and Attaching and Verbs of Separating and

	Disassembling, where there are two roles that undergo some change with no clear distinction between them.
Predicate:	used for classes with a predicative complement.
Recipient:	target of the transfer. Used by some classes of Verbs of Change of Possession, Verbs of Communication, and Verbs Involving the Body. The selection restrictions on this role always allow for animate and sometimes for organization recipients.
Stimulus:	used by Verbs of Perception for events or objects that elicit some response from an xperiencer. This role usually imposes no restrictions.
Theme:	used for participants in a location or undergoing a change of location. Also, Theme1 and Theme2 are used for a few classes where there seems to be no distinction between the arguments, such as Differ-23.4 and Exchange-13.6 classes.
Time:	class-specific role, used in Begin-55.1 class to express time.
Topic:	topic of communication verbs to handle theme/topic of the conversation or transfer of message. In some cases, like the verbs in the Say-37.7 class, it would seem better to have `Message' instead of `Topic', but we decided not to proliferate the number of roles.

3.6　日本語話し言葉コーパス（CSJ）

自発音声を自動認識できる音声認識システムの開発に寄与する目的で話し言葉のデータを収集したコーパス［75, 183, 202, 214］。CSJ（Corpus of Spontaneous Japanese）と略される。約 50 万語の情報を持つ。

句点によって明示的に文の範囲が示される書き言葉と異なり、話し言葉は文の境界が明示されていない。そのため、話し言葉のテキストを節の境界に区切るというタスクが必要となる。このタスクに対しては丸山ら［172］が開発した日本語節境界検出プログラム CBAP を使用している。品詞情報、構文情報が付与されている。

3.7 PropBank

Penn TreeBank に意味情報を付与する形で作成されたコーパス [59, 60, 94] 。意味役割として表 3.3 のタグが用意されている。ArgM は付加語的な項（adjunct-like arguments（ArgMs)）に付与されるタグである。Arg1～Arg4 は非対格動詞（unaccusative verb）を表している [76] 。

表 3.3: PropBank の意味役割タグ

タグ	意味
Arg0	agent
Arg1	direct object / theme / patient
Arg2	indirect object / benefactive / instrument / attribute / end state
Arg3	start point / benefactive / instrument / attribute
Arg4	end point
ArgM-TMP	when?
ArgM-LOC	where at?
ArgM-DIR	where to?
ArgM-MNR	how?
ArgM-PRP	why?
ArgM-REC	himself / themselves / each other
ArgM-PRD	this argument refers to or modifies another
ArgM-ADV	others
ArgM-NEG	negation
ArgM-MOD	modals
ArgM-CAU	cause
ArgM-PNC	Purpose
ArgM-EXT	Extent
Arg1	Logical subject, patient, thing rising
Arg2	EXT, amount risen
Arg3	start point
Arg4	end point

テキストは図 3.7 のようにタグ付けされる。また、PropBank は動詞の意味ごとに FrameSet と呼ばれる単位に分けられている。

・[*Arg1* Sales] rose [*Arg2* 4%] to [*Arg4* $3.28 billion] from [*Arg3* $3.16 billion].
・[*Arg1* The Nasdaq composite index] added [*Arg2* 1.01] to [*Arg4* 456.6] on paltry volume.

図 3.7: PropBank におけるタグ付きテキストの例

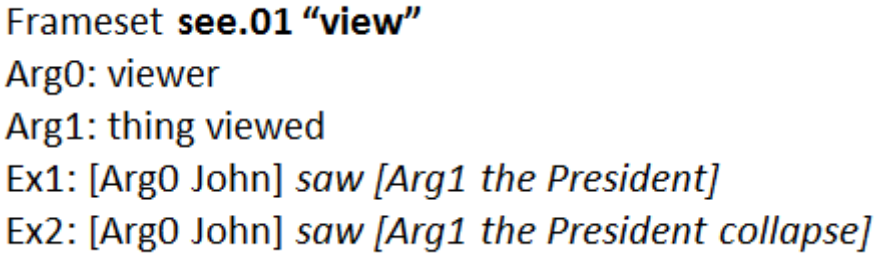


図 3.8: PropBank の FrameSet [94]

なお、Propbank の中国語版として、Chinese Treebank [148] に対して意味役割のタグを付与した Chinese Propbank [149] が公開されている。

3.8 NomBank

PropBank に出現する事態性名詞に対して意味役割の付与を行ったコーパスである [80, 81] 。タグ付与の仕様は PropBank のそれに準ずる。

1. *Her gift of a book to John [NOM]*
REL = *gift, ARG0 = her, ARG1 = a book, ARG2 = to John*

2. *his promise to make the trains run on time [NOM]*
REL = *promise, ARG0 = his, ARG2-PRD = to make the trains run on time*

3. *her husband [DEFREL RELATIONAL NOUN]*
REL = *husband, ARG0 = husband, ARG1 = her*

図 3.9: NomBank [81]

3.9　新聞記事 GDA コーパス

コーパスに対するタグ付けの仕様として Grobal Document Annotation（GDA）が提唱されている［177］。計算機と人間の双方にとって理解しやすいタグ仕様を目指したもので、XML 形式でタグを定義する。新聞記事 GDA コーパスは、毎日新聞のテキスト(3,000 記事、約 37,000 文、約 910,000 語)に対して形態素・統語構造・語義・照応と共参照の情報を付与したコーパスであり、その記述形式は GDA に準拠している［226］。

```
<su syn="f">
  <date>昨年九月二十一日</date>
  <time>午前</time><ad>の</ad><n>授業</n><n>中</n>、
  <adp><n id="Stud">三年生の男子生徒（１４）</n>が</adp>
  <adp><np id="Man">学校のそばを女性と歩いていた男</np>を</adp>
  <adp>教室内から</adp>
  <v>冷やかし</v><v>た</v>。
</su>
```

図 3.10: 新聞記事 GDA コーパスのデータフォーマット

3.10　岩波国語辞典第五版タグ付きコーパス 2004

岩波国語辞典第五版における約 5 万 6 千項目のデータに、形態素・統語構造・照応と共参照、岩波国語辞典自身に基づく語義の情報などを付与したコーパス。約 198,000 を所収する［178, 226］。

```
<entry id="e00304">
  <hd>あがなう</hd>
  <hst>あがなふ</hst>
  <sense id="e00304.0">
    <sense id="e00304.0.1">
      <orth>△購う</orth>
      <pos>五他</pos>
      <su><vp eq="use">買い求める</vp>。</su>
      ▽
      <su aen="mention"><ajp opr="in">文語的</ajp>。</su></sense>
    <sense id="e00304.0.2">
      <orth>×贖う</orth>
      <pos>五他</pos>
      <idi>『罪を<v ed="hd">贖う</v>』</idi>
      <su><vp eq="use"><np sbm="use.obj:を">つぐない</np>をする</vp>。</su>
      <su><vp eq="use"><np sbm="use.obj:を">罪ほろぼし</np>をする</vp>。</su>
      ▽
      <etym><su>その責めを免れるため金品を差し出したことから。</su></etym></sense>
  </sense>
</entry>
```

図 3.11: 岩波国語辞典に対する GDA によるアノテーション

3.11 OntoNotes

英語、中国語およびアラビア語に対してタグを付与したコーパス［113］。PennTreebank のアノテーションを元にした文法のタグ付けと、PropBank のアノテーションを元にした意味役割のタグ付け、Omega ontology [97]に基づいた語の意味情報が付与されている。

Release1.0 では、オンラインジャーナル（中国語 400,000 語、英語 300,000 語）のデータが提供されており、Release2.0 ではそれにニュース放送（中国語 274,000 語、英語 200,000 語）のデータが追加された［104］。アラビア語は Release3.0 において追加された。最新版の Release4.0 に所収されているデータについては、表 3.4 に記載している。

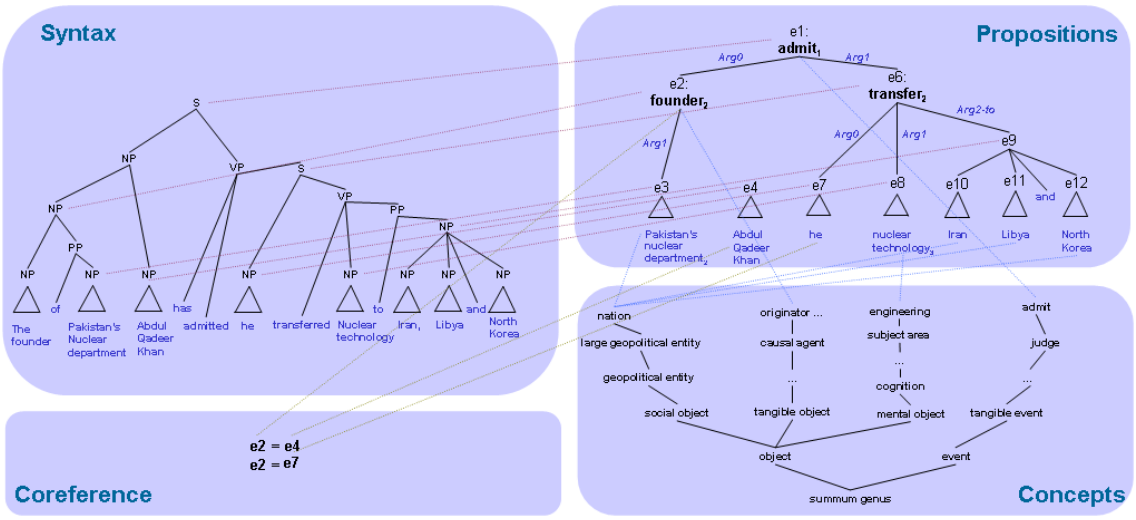


図 3.12: OntoNotes 概念図［113］

```
--------------------------------------------------------------------------------
id: 0@abc/00/abc_0001@abc@en@on
--------------------------------------------------------------------------------


sentence: Like many Heartland states , Iowa has had trouble *PRO* keeping young people down on the farm or anywhere within state lines
--------

parse tree:
---------

    (TOP (S (PP-MNR (IN Like)
                    (NP (JJ many)
                        (NNP Heartland)
                        (NNS states)))
            (, ,)
            (NP-SBJ (NNP Iowa))
            (VP (VBZ has)
                (VP (VBN had)
                    (NP (NP (NN trouble))
                        (S-NOM (NP-SBJ (-NONE- *PRO*))
                               (VP (VBG keeping)
                                   (NP (JJ young)
                                       (NNS people))
                                   (ADVP-LOC (ADVP (RB down)
                                                   (PP (IN on)
                                                       (NP (DT the)
                                                           (NN farm))))
                                             (CC or)
                                             (ADVP (RB anywhere)
                                                   (PP (IN within)
                                                       (NP (NN state)
                                                           (NNS lines)))))))))
            (. .)))

----

sense and proposition:
----------------------


Like                0
many                1
Heartland           2
states              3
,                   4
Iowa                5
has                 6
--------------------------------------------------
predicate: have;     pb_sense: 01

--------------------------------------------------
```

図 3.13: OntoNotes のデータフォーマット

表 3.4: OntoNotes 4.0 に所収されている語 [113]

	English	Chinese	Arabic
Newswire	600K	250K	300K
Broadcast News	200K	250K	—
Broadcast Conversation	200K	150K	—
Web text	300K	150K	—

3.12　NAIST テキストコーパス

京都テキストコーパスを拡張する目的で作成されたコーパス［217］。既存のコーパスにおける共参照関係のタグ付与における問題の克服に主眼が置かれている。

ある表現が同一文章内の他の表現を指す機能のことを照応といい、この場合差す側の表現を照応詞、指される側の表現を先行詞と呼ぶ。これに対し、二つまたはそれ以上の表現が世界において同一の実体を指している場合、それらの表現は共参照の関係にあるという。多くの場合において照応関係と共参照関係は同時に成立するが、そうでない場合もある。以下の(1)の文の場合、代名詞「彼」が「野田首相」を指しており、かつ同一人物を指しているため、照応関係かつ共参照関係である。それに対し(2)の文の場合、「それ」は「iPod」を指しているため照応関係となるが、同じ実体を指しているわけではないため共参照関係とはならない。

(1) 野田首相は、18 日午前の衆院予算委員会で、情報伝達の不手際を認めた。 彼の説明によると…
(2) 太郎は iPod を買った。次郎もそれを買った。

なお、Mitkov［227］は前者を identity-of-reference anaphora（IRA）、後者を identity-of-sense anaphora（ISA）と呼び区別している。

京都コーパス 4.0 においては、実体と属性の間にも共参照関係のタグが付与されている。そのため先の文(1-a)においては実体「野田」とその属性「首相」は共参照の関係であるとしてタグが付与されることになる。

3.12.1　EDT の共参照タグ付与の問題

情報抽出の主要な会議である Message Understanding Confernce（MUC）では情報抽出の部分問題として共参照解析の問題を扱っている。その過程で作成された共参照関係タグ付与コーパスでは名詞句間の共参照関係がタグ付与されているが、このコーパスの仕様では，一般に共参照関係とはみなされないような量化表現（every，most など）を伴う場合や同格表現（Julius Caesari, the/a well-known emperori, ...のような表現）も共参照関係とみなしてタグ付与されているという問題を含んでいる。

MUC の共参照解析タスクの後継に相当する Automatic Content Extraction（ACE）の Entity Detection and Tracking（EDT）のタグ付与においては、MUC の過剰な共参照関係の付与を回避するために、言及の型が人名や組織名などいくつかの固有表現に該当し、かつ総称的でない場合にのみ共参照関係のタグ付与を行うという制限を設けて

いる。しかし、この制約は、情報抽出などある種の問題に利用する場合には有利に働くが、様々な応用処理を考えた場合には好ましくない場合もある。NAIST テキストコーパスは名詞句のクラスに制約を加えずに共参照関係を認定している。

3.12.2　総称名詞間のタグ付与の問題

以下の(1)(2)の文において、「本」は共に総称名詞として用いられている。

(1) 本は、書物の一種で、印刷・製本された出版物を指す。
(2) 図書館の本は借りることができる。

だが、(1)の「本」と(2)の「本」は総称のレベルが異なる。このような関係を共参照関係として認めるかという問題があり、GDA コーパスなどは認めるという立場を取っているが NAIST テキストコーパスはこれを共参照関係とは認めていない。GDA コーパスの共参照タグは IRA と ISA の両方の関係で付与されていると考えられるが、NAIST テキストコーパスのそれは IRA の関係のみで付与される [217]。

3.13　KNB コーパス

正式名称は「Kyoto-University and NTT Blog コーパス」で、ブログ記事を解析して作成したコーパスである [180] 。京都観光、携帯電話、スポーツ、グルメの 4 テーマのブログ 249 記事、4,189 文から成る。形態素、構文、格・省略・照応、評判情報を含む。タグのフォーマットは京都テキストコーパスのそれに準拠しているが、ブログ文は話し言葉の特徴を有しているため、それに対応するための拡張が行われている。それらの拡張の中には、話し言葉を対象とした解析済みコーパスである日本語話し言葉コーパス（CSJ）の仕様と類似したものも含まれている。

```
# S-ID:KN001_Keitai_1-1-30-01 KNP:2008/02/25 SCORE:-1.89971 MOD:2009/02/01
＊2D <文節内><係:文節内><文頭><サ変><体言><名詞項候補><先行詞候補><非用言格解析:動><態:未定><正規化代表表記:プリペイド/プリペイド>
プリペイド ぷりぺいど プリペイド 名詞 6 普通名詞 1 ＊ 0 ＊ 0 "疑似代表表記 代表表記:プリペイド/プリペイド" <疑似代表表記><代表表記:プリペイド/プリペイド><正規化代表表記:プリペイド/プリペイド><文頭><品詞変更:プリペイド-プリペイド-プリペイド-15-2-0-0-"疑似代表表記 代表表記:プリペイド/プリペイド"><品曖-カタカナ><未知語><記英数カ><カタカナ><名詞相当語><サ変><自立><内容語><意味有><タグ単位始><文節始><固有キー><NE:ARTIFACT:head>
+ 2D <BGH:携帯/けいたい><文節内><係:文節内><サ変><体言><名詞項候補><先行詞候補><非用言格解析:動><態:未定><正規化代表表記:携帯/けいたい>
携帯 けいたい 携帯 名詞 6 サ変名詞 2 ＊ 0 ＊ 0 "カテゴリ:人工物-その他:抽象物 ドメイン:家庭・暮らし 代表表記:携帯/けいたい" <カテゴリ:人工物-その他:抽象物><ドメイン:家庭・暮らし><代表表記:携帯/けいたい><正規化代表表記:携帯/けいたい><漢字><かな漢字><名詞相当語><サ変><自立><複合←><内容語><意味有><タグ単位始><NE:ARTIFACT:middle>
+ 3D <BGH:電話/でんわ><文節内><係:文節内><補文ト><サ変><体言><名詞項候補><先行詞候補><非用言格解析:動><態:未定><正規化代表表記:電話/でんわ><C 用;【電話】;=;4;3;9.999:KN001_Keitai_1-1-26-01(4文前):3タグ>
電話 でんわ 電話 名詞 6 サ変名詞 2 ＊ 0 ＊ 0 "補文ト カテゴリ:人工物-その他 ドメイン:家庭・暮らし 代表表記:電話/でんわ" <補文ト><カテゴリ:人工物-その他><ドメイン:家庭・暮らし><代表表記:電話/でんわ><正規化代表表記:電話/でんわ><漢字><かな漢字><名詞相当語><サ変><自立><複合←><内容語><意味有><タグ単位始><NE:ARTIFACT:middle>
+ -1D <SM-主体><SM-人><BGH:ファン/ふぁん><文末><補文ト><体言><用言:判><体言止><レベル:C><区切:5-5><ID:(文末)><RID:112><提題受:30><主節><判定詞><名詞項候補><先行詞候補><正規化代表表記:ファン/ふぁん><用言代表表記:ファン/ふぁん><格要素-ガ:NIL><格フレーム-ガ-主体><格フレーム-ガ-主体ｏｒ主体準><C 用;【電話】;ノ？;0;2;9.999:KN001_Keitai_1-1-30-01(同一文):2タグ><NE:ARTIFACT:プリペイド携帯電話ファン>
ファン ふぁん ファン 名詞 6 普通名詞 1 ＊ 0 ＊ 0 "カテゴリ:人 ドメイン:スポーツ:無し 多義 代表表記:ファン/ふぁん" <カテゴリ:人><ドメイン:スポーツ:無し><多義><代表表記:ファン/ふぁん><正規化代表表記:ファン/ふぁん><文末><表現文末><記英数カ><カタカナ><名詞相当語><自立><複合←><内容語><意味有><タグ単位始><固有キー><文節主辞><NE:ARTIFACT:tail>
EOS
```

図 3.14: KNB コーパスのデータフォーマット

3.14　現代日本語書き言葉均衡コーパス（BCCWJ）

国立国語研究所によって作成されたコーパス［184］。既存のコーパスが新聞や文学作品など手近な資料を元に構成されているためにコーパスが備えるべき普遍性を喪失しているという問題意識から構築されたコーパスである。BCCWJ（Balanced Corpus of Contemporary Japanese）と略される。普遍性を獲得するために、新聞、雑誌、一般書籍、教科書、白書、TV 放送、Web 記事といった広い範囲の言語資源を統計的母集団とし、そこからのランダムサンプリングによってコーパスを構築している。約 1 億語を所収する。

第 4 章　格フレーム辞書の自動構築

前章で概括した言語資源であるが、これを人手で構築するには多大な工数を必要とする。そのため、計算機によってこれを自動で獲得する研究が行われている。タグ付きコーパスについては、生コーパスに対して述語項構造解析を行えばその結果として獲得できるため、ここでは格フレーム辞書の自動構築手法について概括することとする。大別して、タグが付与されていない生のコーパスから獲得する手法と、タグが付与されたコーパスから構築する手法の二種類がある。

4.1　生コーパスからの構築

下位範疇化フレーム（subcategorization frame）とは、動詞が支配する主語、目的語、前置詞などの情報を表すデータ構造のことを言う。一般的には表層格と深層格の対比情報を持たないが、その点を除けば表層格フレームと情報としての差異はない。
英語圏においても、この下位範疇化フレームを生コーパスから自動獲得する研究は様々な研究者によっておこなわれている。英語は日本語と異なり格要素が省略されることはないため、問題となるのは格要素が用言にとって必須か任意かの判定である。

4.1.1　Brent の手法

下位範疇化フレームの自動獲得の研究の最初期のものとして、Brent [11, 12] の研究が挙げられる。Brent の手法においては、まずコーパスの中から動詞を探し、次にその動詞の項となる節を探す。そして以下の 6 つの下位範疇化フレームに分類される。

1．NP only
2．tensed clause
3．infinitive
4．NP & clause
5．NP & infinitive
6．NP & NP

動詞の探索は、コーパスから-ing という接尾辞を持つ語彙と、それと同じ語幹で接尾

辞を持たない語彙とのペアを探すことで実現する。そのようにして挙げられた動詞の候補は、限定詞や to を除く前置詞の直後に現れていない限り、動詞であると判断される。次に、これらの動詞（と仮定されたもの）に対して、下位範疇化フレームの特定が行われる。それには文法上の特性を利用して行う。例えば「I want to tell him that the idea won't fly.」において that the で始まる節は、動詞 tell の項であると考えられる。何故なら代名詞である him は関係節を取ることがまれだからである。また「I hope to attend.」のように、動詞の右にある単語列が「to V」の形であった場合、その節は不定詞補文に分類される。

ただ、この文法的な手がかり(cue)を用いた下位範疇化フレームの分類には誤りも多く、最終的なアウトプットにはノイズも多く混在している（括弧内がノイズの例)。

1．NP only（arrive them）
2．tensed clause（want he'll attend）
3．infinitive（greet to attend）
4．NP & clause（yell him he's a fool）
5．NP & infinitive（hope him to attend）
6．NP & NP（shout him the story）

Brent は、二項分布に基づくフィルターを導入することでこの問題の解消を試みた。この手法は後に Manning［77］、Ersan ら［29］、Lapata［69］、Briscoe ら［13］、Sarkar ら［119］といった研究者たちも取り入れている。
動詞 j に対する cue の個数を n とし、下位範疇化フレーム i の cue の個数をmとする。動詞に対して下位範疇化フレームscf_iが選択された際にそれが誤っている確率をp^eとすると、下位範疇化フレームscf_iの cue がm回以上現れたとき、それがscf_iのメンバーでない動詞が n 回現れたときにその動詞と同時に現れる確率$P(m+, n, p^e)$は二項分布を用いて以下の式で表現される。

$$P(m+, n, p^e) = \sum_{k=m}^{n} \frac{n!}{m!\,(n-m)!} p^m (1-p)^{n-m}$$

この確率をフィルターとして用いることで 95%以上の精度が実現されるという報告がされている。

p^e の求め方は各研究者が提示しているが、Brisco ら［13］は以下の式によって定義した［65]。

$$p^e = \left(1 - \frac{|verbs\ in\ scf_i|}{|verbs|}\right) \frac{|patterns\ for\ scf_i|}{|patterns|}$$

Brent の手法は下位範疇化フレームを特定する際に文法的な手掛かり（lexical cues）を用いるが、多くの動詞と下位範疇化フレームにはそのような手掛かりは存在しない。例えば「They assist the police in the investigation.」における assist のように、いくつかの動詞は in で始まる節を下位範疇として取るが、実際には「He built a house in the woods.」のように、動詞の後に現れる in で始まる節は多くの場合名詞修飾節か下位範疇化されない位置格の節である。

Brent の手法の欠陥を克服するために、後続の研究者はインプットとなるコーパスに対して POS タギングとチャンキングを行うことを前提とした手法を提示している。Ushioda ら［141］の手法では POS タギング済みのウォールストリートジャーナルコーパスを用いる。獲得する下位範疇化フレームの種類は 6 種類で、これは Brent のものと同じである。

4.1.2　動詞の意味の多様性に対するアプローチ

一般的な格フレームの構造は、実際の文中に現れた共起関係を何らかの手法で統合したものである。直感的な獲得手法として、文中に現れた動詞と名詞の共起情報を、単純に動詞をキーに統合するという手法が考えられる。
しかし、以下の 2 つの文について考える。

(1) 車に荷物を積む
(2) 経験を積む

これらの文には「積む」という動詞が用いられているが、それらは字面については同じでも意味が異なる。それは「車に経験を積む」という表現が適切ではないことからも判断できる。つまりこれらの「積む」はそれぞれ別の格フレームとして獲得されなければならない。共起情報を、単純に動詞をキーに統合するという手法では、この問題を回避できない。

このような動詞の意味の多様性の問題に対して、春野［191］は事例間の最小汎化というアプローチを試みている。これは Hindle [48] らによって行われた、コーパスをベースとした前置詞句の曖昧性解消の手法や、李ら［70］の格フレームの一般化手法に想を得ているが、これらの手法は用例を一定次元のベクトルで表現することを前提としており、日本語の文節数は動詞の用法によって多様でありかつ省略も多いためこのような

表現ができない。そこで春野は表層格フレームの比較基準として、その格フレームを導入したことによって事例の記述をどの程度圧縮できるかという観点から格フレームの有効度を定義している。入力文のセットから二つの事例を取り出し、それらに共通の素性のみをシソーラス上の最小上界（それらの素性の直近の共通のノード）で置き換えて表層格フレームの候補を生成し、有効度に基づいて最適な格フレームを選択し、その格フレームによってカバーされる事例を入力文のセットから取り除く。同様の手続きを残りの事例についても繰り返すことで、格フレームの獲得を行う。格フレームFを評価する際の基準である格フレームの有効度$utility(F)$は以下のように定義される。

$$utility(F) = Explict\text{-}Bits(F) - \lambda\big(CL(F) + Gen(F)\big)(bits)$$

$utility(F)$は、格フレーム F によってカバーされる全ての事例を F を使わずに明示的にコーディングした場合の情報量$Explict\text{-}Bits(F)$と、F を用いて圧縮した場合に必要な情報量$\lambda\big(CL(F) + Gen(F)\big)$の差を表し、すなわち F の情報圧縮能力を示す。この定義は、事例を説明する最良の規則とはその事例を最小の長さで記述する規則であるというオッカムの原理の一形態に基づいている。

一般的に用言と名詞の共起用例を収集しただけでは格フレーム辞書とはならず、用例を何らかの基準に従ってクラスタリングする必要がある。格要素についてシソーラスを用いて意味素の汎化レベルを決定することによって用例のクラスタリングを行う手法は、用言の用法の多様性に対応しきれていない。例えば「積む」という動詞について、以下の２つの名詞との共起用例が観察されたとする。

(1) 従業員が荷物を積む
(2) 従業員が経験を積む

「荷物を積む」が物理的な積載を意味するのに対し、「経験を積む」は心理的・比喩的なものである。すなわちこれらの「積む」は異なる意味であると解釈するのが妥当であるが、従来のクラスタリング手法では格要素「従業員が」が同一であるために同一の格フレームとしてマージされてしまう可能性がある。

このような問題に対し、用言と直前の格要素を組として捉えると用言の用法がほぼ一意に特定できるとする意見がある［166］。河原らはこの観察に基づき、生コーパスから格フレームを獲得する手法を提示している。生コーパスから用言と名詞との共起用例を獲得し、それを用言と直前の格要素ごとにマージして格フレームの候補を作成する。得られた候補につき相互に類似度を算出し、類似度が閾値を超えた候補同士をマージすることによって最終的な格フレームを得る。

上記の河原の手法は、生コーパスに対して KNP を用いた構文解析を行っているが、その際に問題となるのは解析の誤りである。そのため河原らはヒューリスティックスを用いて解析結果の中から確信度の高いもののみを抽出しインプットとして用いている。そのため以下のような言語現象に対する用例を収集できていない。

・係助詞句

(1) a. 車は速い
b. 本も読んだ

・被連体修飾詞

(2) a. 速い車
b. 読んだ本

・二重主語構文

(3) この車はエンジンがよい

・外の関係

(4) 魚を焼くけむり

・格変化

(5) a. 社会党が新進党の支持を得る
b. 社会党が新進党から支持を得る

(6) a. この車のエンジンがよい
b. この車はエンジンがよい

河原らは、コーパスからこれらの用例を獲得し、それを先に自動獲得した格フレームを用いて解析することで格フレームを拡充する手法を提示している [57, 168, 170] 。このようにして獲得された格フレーム辞書は、省略解析などに応用されている[169] 。なお、河原らは同手法によるシステムを高性能計算機グリッド上で実行し、5 億日本語文を解析し、約 9 万用言からなる格フレームを構築した実績についても報告している [171] 。

4.1.3 語順を考慮した格フレームの獲得

日本語においては語順の変化は文全体の意味に大きな変化を与えない。しかし統計的に自然な語順というものはあり、動詞のよって語順変化のパターンが異なるならば、それは動詞の意味分類を詳細化する上で役立つ情報になると考えられる。大竹ら [204] はこのような推測に基づき、語順を考慮した格フレームの獲得手法を提示している。語

順をグラフで表現した格遷移ネットワークモデルを用いる。ネットワークにおいては、弧に重みが設定されるが、それは bi-gram によって与えられる。名詞には意味素性が割り当てられる。意味素性としては、計算機用日本語基本動詞辞書 IPAL で採用されている 18 種類の素性を用いている。コーパスは、日本経済新聞の CD-ROM 版を使用している。

4.1.4 遺伝的アルゴリズム

係り受け関係に格を割り当てるルールの学習に、帰納学習法 C4.5 と遺伝的プログラミング（generic programming：GP）を組み合わせた研究がある［198］。帰納学習法は、訓練データの集合からそのデータをクラスに分類する決定木を学習する手法だが、訓練データにノイズが含まれている場合や属性値が連続の場合などに対応できないという欠点を持つ。GP はそれを補完するものとして導入されている。

4.2 タグ付きコーパスからの獲得

東ら［212］は、EDR 電子化辞書に含まれる格フレームと例文との類似度を計算し、それに従って動詞を語義ごとに分類する手法を提示している。分類される動詞のカテゴリは、EDR 電子化辞書の格フレームにおける動詞概念の分類に一致する。EDR コーパスから取得した名詞と動詞の共起関係と EDR 電子化辞書に含まれる格フレームとの類似度を計算し、類似度が最大の格フレームの動詞概念を分類の結果とする。なお、名詞と動詞の共起関係は EDR 格フレームと同じ構造をしている。すなわち、求めるのは格フレーム間の類似度である。

格フレーム間の類似度は、スロット間の類似度の和として定義される。「に」と「へ」のように関連の深い助詞であれば異なる助詞であっても類似度を 0 とするのは適切ではないが、ここでは単純に助詞が異なれば類似度は 0 になる。すなわち、同じ助詞を持つ名詞間の類似度の和が、格フレーム間の類似度である。名詞間の類似度$sim(N_i, N_j)$は、EDR 概念体系辞書を利用して、以下のように定義される。

$$sim(N_i, N_j) = \frac{1}{k} \times Dmax_{ijk} \times \left(\alpha \frac{CS_{ijk}}{S_{ik}} + \beta \frac{CS_{ijk}}{S_{jk}} \right)$$

$$(0 \leq \alpha \leq 1,\ \beta = 1 - \alpha)$$

S_{ik}はN_iの上位概念のうち、概念体系上の深さがk以下のものの数を、CS_{ijk}はN_iとN_jの

共通上位概念のうち、深さがk以下のものの数を、$Dmax_{ijk}$はCS_{ijk}のうちの最大の深さをそれぞれ表す。

4.3　対訳コーパスからの獲得

宇津呂ら［162］は、二言語対訳コーパスから動詞の表層格フレームを獲得する手法を提示している。これは統語的曖昧性を克服するために、二言語間の翻訳例を構文解析して、その結果を二言語間で比較するというアプローチに基づいている。構文解析された二言語は、対訳組成構造と呼ばれるデータ構造で表現される。実験においては、16の動詞に対して対訳素性構造を集め、表層格フレームの獲得を行っている。
ただ、動詞の多義性の問題については、システムから発せられる質問に回答するという形で人間が介在する必要があり、完全な自動獲得処理ではない。また、講談社学術文庫の和英辞典から約 4 万対訳例を取り出して作成されたコーパスもタグ付きでないとはいえ、対訳コーパスであることからタグ付きコーパスが持つ用例の量の問題は抱えている。

4.4　既存の格フレーム辞書からの獲得

既存の格フレームから別の新たな格フレームを生成するというアプローチも行われている。

Mu プロジェクト［200］で構築された格フレーム辞書を用いて、サ変動詞の意味特性とこのサ変動詞と共起する表層格および深層格情報のセットから名詞の意味情報を含む格フレームを生成する手法が考案されている［201］。これは元々獲得済みの格フレーム情報を「サ変動詞構文情報」「サ変動詞意味情報」「表層格」「深層格」「名詞意味情報」からなるデータ構造に変換し、それに対して「新規サ変動詞」「サ変動詞構文情報」「サ変動詞意味情報」「表層格」「深層格」というデータ構造で新規のサ変動詞を入力し、マッチングさせることによって入力されたデータ構造に対して名詞意味情報を付与するというものである。ここで言う格フレームとは、表層格と深層格に加え名詞の意味情報を含むものである。ただ、この手法は、その多くを人手による作業に負っており、特にマッチングしなかった場合については名詞意味情報を含む格フレームの分析を手動で行わなければならない。実験においては未知語 975 語に対して格フレームの不一致が 187 パターン生じており、手動での解析に少なからず手間を要すると考えられる。

橋本ら［179］は、EDR 日本語辞書を元にした動詞格フレームからサ変動詞の格フレ

ームを獲得する手法を提示している。

第 5 章　表層格付与

5.1　下位範疇化の確率モデル

コーパス中の共起データにおける複数の格の共起を観測したときに、これを格フレーム中の必須格の共起に相当する強い依存関係ととらえるのか、あるいは独立事象に相当する任意格的な格が偶然共起したととらえるのかという問題がある。例えば、

(1) 子供が公園でジュースを飲む。

という文について考える。用言「飲む」の格要素としては、ガ格の「子供」、デ格の「公園」、ヲ格の「ジュース」が考えられるが、これらの間の依存関係には

1. ガ、デ、ヲの 3 つが依存関係にある
2. ガとデが依存関係にある
3. ガとヲが依存関係にある
4. デとヲが依存関係にある
5. いずれも独立であり、依存関係にはない

の 5 パターンが考えられる。

また、意味素に基づく格フレームを獲得する場合、コーパスから格要素の名詞を観測したときに、これを概念階層のどのレベルの概念クラスの共起知識としてとらえるのかという問題がある。(1)の文において、用言「飲む」と共起している「子供」「公園」「ジュース」について、それを包含するクラスをそれぞれ「人間」「場所」「飲料」とする。ここで、「人間」「飲料」の上位クラスとして「動物」「液体」を考える。こう考えた場合、共起の用例を生成し得る下位範疇化フレームの可能性として、以下の 4 種類が考えられる。

1. ガ格は「子供」、ヲ格は「ジュース」
2. ガ格は「動物」、ヲ格は「ジュース」
3. ガ格は「子供」、ヲ格は「液体」
4. ガ格は「動物」、ヲ格は「液体」

宇津呂らは、これらの問題に対し、動詞と複数の格の格要素の共起データから、動詞

の下位範疇化の確率モデルを学習する手法を提示している。分下位範疇化フレームの組を値にとる隠れ変数を用いた EM アルゴリズムに基づく手法［142, 143, 163］と、を用いる手法と、最大エントロピー法に基づく手法［164］を提示している。後者の場合においては 4 種類の素性を検討する。格の依存関係を考慮しない「部分フレームモデル」、全ての格が依存していると仮定する「1 フレームモデル」、全ての格が独立であると仮定する「独立格モデル」、コーパスから格の依存関係を統計的に算出した結果を反映した「独立フレームモデル」の 4 つである。それらのモデルを比較し、結果として「独立フレームモデル」が最も高い成績を収めている。「独立フレームモデル」は、パラメータによって部分下位範疇化フレームの格の独立性の判定基準を調整できるが、この独立性条件が厳しい方が良い結果になることが報告されている。

動詞の下位範疇化の確率モデルとして Bayesian Network を用いる手法も提案されている［174］。一般的に、ある動詞に関して助詞と名詞の組合せが N 通りあるとすると、その動詞と共起する可能な格のパターンは2^N種類に上るが、実際に観測されるのはそれらのうちの非常に少数である。つまり実際には同一種類の格は共起しにくい、格の種類と名詞のクラスには依存関係がある、名詞のクラス同士に依存関係がある、といった様々な制約がある。宮田らの手法［174］では、これらの 3 つの制約を下位範疇化フレームの包摂関係として表現し、それらに対応する確率モデルを推定している。

5.2 決定リスト

ノードごとに判断条件が与えられている二分岐を決定木という。これを if-then 形式のリストとして表現したものを決定リストという［117］。これは以下の式で形式化される。

$$r_i: E_i \rightarrow d_i$$

ここで規則r_iは、ある事象が証拠E_iを満たすときに、分類対象をd_iに分類するという判定をすることを表す［216］。

規則の適用の優先順序を適切なものとすることが決定リストにおける重要な問題である。これは通常最尤推定によって学習されるが、事例の数が少ないときに生じる問題の克服としてベイズ推定による手法［210］や、判定の根拠となる証拠によって規則をタイプごとに分類してから優先順位を決定するといった手法［216］が提案されている。

自然言語処理における問題の多くはクラス分類の問題として捉えられることから、アクセント記号復元［152］、単語のわかち書き［199］、形容詞の修飾先の決定［215］、語の多義性解消［216］、スペルミス検出［122］、固有表現抽出［225］、文節の係り受

け解析など、様々な問題に決定リストを適用した事例が報告されている。

平ら［221］は、決定リストを応用した述語項構造解析を行っている。NAIST テキストコーパスを対象とし、述語の基本形、係り受けタイプ、汎化レベル、機能語、態の五種類の属性を設定し、SVM に基づき、格ごとに独立した分類器を作成する。SVM によって各属性の重みを学習し、学習された重みの順に決定リストの優先度を設定する。動詞だけでなく事態性名詞も解析の対象となっている。

5.3 Markov Logic

日本語述語項構造解析においては、平ら［221］の研究のように SVN のような分類器を用いて格ごとに独立して同定を行う手法が主に研究されてきた。
しかし、同じ述語に属する項の間には依存関係があると考えられる。

(1) ライオンがシマウマを食べた。
(2) ライオンに追いかけられたシマウマが崖から落ちた。

(1)の文において「食べた」のガ格とヲ格が共にライオンになるとは考えられないが、格ごとに独立した分類器を用いる手法では、そのような誤った判断を行う可能性がある。（２）においては「ライオン」が項として同定され、述語「落ちた」の項は「シマウマ」だけであったとすると、「ライオン」はもう一つの述語である「追いかけられた」の格になることが分かる。つまり文内にある他の述語との関係が同定の手掛かりになることがあり、格ごとに独立した分類器を用いる手法ではこのような依存関係を反映させることが難しい［173］。

Markov Logic は Richardson によって考案された機械学習の手法である［115]。
Markov Logic は、与えられた複数の一階述語論理式をなるべく多く満たすような形で推論を行う学習手法である。各論理式に重みが設定されている場合は、満たすことのできる論理式の重みの総和が大きくなるように学習する手法となる。実際には各論理式をノードにマッピングした Markov Network を形成することで学習を行う。同手法を意味解析（semantic parsing）に適用した論文［99］が EMNLP-2009 の best paper awards に選ばれるなど、近年注目されている手法である。

意味役割付与のタスクを述語同定（predicate identification）、フレームの曖昧性（frame disambiguation）、項同定（argument identification）、項分類（argument classification）の複合タスクであると捉え、Markov Logic を適用した報告例がある

[83] 。また Meza-Ruiz らは同様の手法を、カタロニア語、中国語、チェコ語、英語、ドイツ語、日本語、スペイン語の各国語の意味役割付与に適用している [84]。なお、日本語の処理においては、京都テキストコーパス [167] を用いている。

吉川ら [173] は、NAIST テキストコーパスをインプットとして、MarkovLogic による述語項構造解析のモデルを提示している。述語項構造解析における主な部分問題である述語同定、述語語義曖昧性解消、項同定、意味役割付与の 4 つのうち、述語同定と述語語義曖昧性解消を除く 2 つを対象としている。述語同定については、NAIST テキストコーパスのアノテーションに従う。NAIST テキストコーパスは格フレームを持たないため、述語語義曖昧性解消を明示的に行うことができない。

5.4 係り受け構造からの構造変換

述語項構造解析は係り受け構造からの構造変換の問題であると捉えることができる。平ら [220] は、構造変換アルゴリズムのひとつであるSVM^{struct} [140] を述語項構造解析に応用した例を報告している。

X, Y をそれぞれ取り得る可能性のある構造の集合とする。このとき、入力構造$x \in X$から出力構造への写像を考え、訓練用正解データとして以下のようにm個の入出力ペア

$$(x^1, \mathrm{y}^1), \ldots, (x^m, \mathrm{y}^m) \in X \times Y$$

が与えられたとする。入力構造と出力構造の組み合わせに対し、それが正しい変換の組み合わせであった場合に大きな値を取る関数 $h : X \times Y \to \mathbb{R}$ を考え、与えられた訓練用正解データを用いて最適な h を求めることを学習の目的とする。入力構造および出力構造はベクトル化される。関数 h は入力構造ベクトル $\mathbf{x}$ と出力構造ベクトル $\mathbf{y}$ の組に対して一つの実数値を返す関数と重みベクトル $\mathbf{w}$ との内積として

$$h(\mathbf{x}, \mathbf{y}; \mathbf{w}) = \langle \mathrm{w}, \Psi(\mathbf{x}, \mathbf{y}) \rangle$$

のように定義した上で、重みベクトル $\mathbf{w}$ の学習をマージン最大化の考えに基づいて実施する。

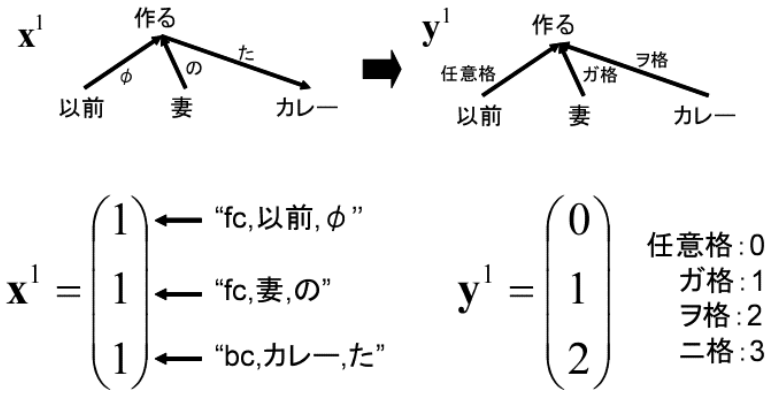


図 5.1: 係り受け構造からの変換とベクトル化の例 [220]

実験に際しては、NAIST テキストコーパスからランダムに生成した訓練用とテスト用それぞれ 500 サンプルをインプットとしている。ベースラインとして、項の述語・事態性名詞に対する係り受け関係それぞれの組み合わせに対し、訓練データに現れた正解の格で最も頻度の高い格を出力とするモデルを用意している。

実験に際しては、NAIST テキストコーパスからランダムに生成した訓練用とテスト用それぞれ 500 サンプルをインプットとしている。ベースラインとして、項の述語・事態性名詞に対する係り受け関係それぞれの組み合わせに対し、訓練データに現れた正解の格で最も頻度の高い格を出力とするモデルを用意している。

格の同定が難しいとされるヲ格とニ格に対して実験を行った結果、ベースラインと比較して 2%程度の精度向上が見られたと報告している。

表 5.1: 平らの評価実験結果 [220]

	訓練データ		テストデータ	
	ヲ格	ニ格	ヲ格	ニ格
ベースライン手法	96.36	98.51	45.12	46.62
提案手法	99.51	99.48	46.15	47.97

第 6 章　深層格付与（意味役割付与）

6.1　ルールベースの手法

統計的手法が導入される以前は、手動で作成したルールをベースに意味情報を解析する手法が主に研究されていた。このタスクは semantic interpretation と呼ばれており、パーサによって統語的に解析されたアウトプットを意味表現に対応させることによって意味解析を実現していた。Woods ら［146］は LUNAR というシステムを作成した。このシステムにおいては意味表現の記述を procedural semantics という記法で扱っていた。これは以下のように、入力文を関数によって表現するものである。

(1) AA-57 is nonstop from Boston to Chicago.
(a) equal(numstops(AA-57, boston, chicago), 0)

関数(a)は文(1)を procedural semantics で関数表現したものである。ここでは numstops が航空機、出発地、目的地の 3 つを引数にとり停止時間を返す関数として与えられており、この関数の値を equal 関数によって 0 と同じであることを表現することで文の意味を表現している。

Hirst [49] は、procedural semantics には入力文から意味表現への対応規則がアドホックになりがちであることなどの問題点があることを指摘し、Montague semantics をベースにした ABSITY というシステムを作成している。

Finin [33] は、複合名詞（nominal compound）の意味記述を行っている。複合名詞における意味解釈とは、複合名詞を構成する名詞の間の意味的繋がりを判別することであり、例えば「aircraft engine」という複合名詞に対しては「part_of」という意味的繋がりを、「meeting room」には「location」を、「salt water」には「dissolved in」という意味的繋がりをタグ付けする。Finin(1980)らは、表 6.1 に示すような複合名詞の意味解釈ルールを定義し、これに基づいて複合名詞の意味記述をすることを提案した。

Sondheimer ら［123］は、KL-ONE [10, 120] と呼ばれる知識表現言語（knowledge representation language）を用いて、意味記述をする方法を提案している。

Charniak ら［16］は、新しい論理的枠組みを提案している。これは意味表現と語用

論の間に存在するギャップを、それらに対して同じ論理表現と推論規則を用いることで緩和することを目指している。また、4 つの簡潔なルールのみで語用論上の様々な現象を記述できるため、意味解釈の簡潔な理論を提供できると主張している。この枠組みに基づいた推論システムである Wimp2 を作成している。

Pustejovsky ら［111］は、Lexical Conceptual Paradigm(LCP)という記法を提案している。

Moore [88] は、unification-based な意味記述を提案している。これを用いることで、ラムダ計算を用いた従来的手法と比較して簡潔な記法を提供できるとしている。しかし、現時点ではラムダ計算で表現できる意味記述のすべてを表現するには至っていない。

Bos［7］は、従来の意味記述に用いられてきた一階述語論理に代わって、Discourse Representation Theory（DRT）[56] を用い、推論規則には Combinatory Categorial Grammar（CCG）を使うことを提案した。

表 6.1: Finin が定義する複合名詞の意味解釈ルール [33]

RoleValue + Concept	engine repair (a to-repair with object = (an engine)) January flight (a to-fly with time = (a January)) F4 flight (a to-fly with vehicle = (an F4)) engine housing (a housing with superpart = (an engine)) iron wheel (a wheel with raw-material= (a iron))
Concept + RoleValue	drinking water (a water which is (an object of (a to-drink))) washing machine (a machine which is (an instrument of (a to-wash))) maintenance crew (a crew which is (an agent of (a to-maintain)))
Concept + RoleNominal	cat food (an object of (a to-eat with agent = (a cat))) oil pump (an instrument of (a to-pump with object = (an oil))) dog house (a location of (a to-dwell with agent = (a dog)))
RoleNominal + Concept	

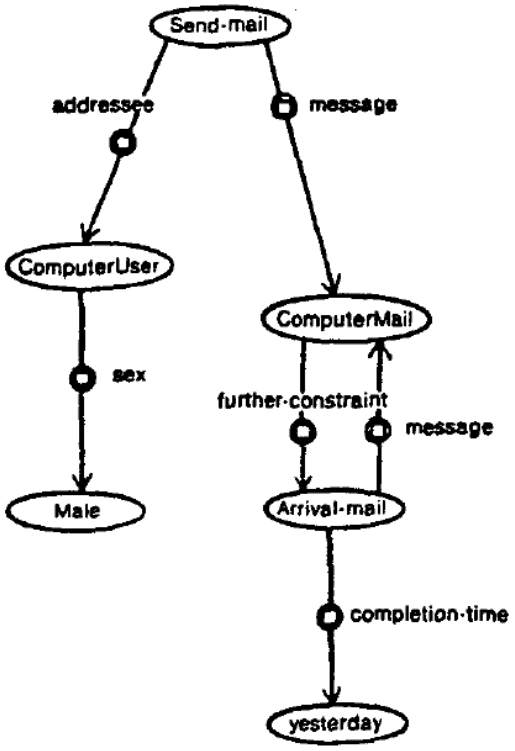


図 6.1: KL-ONE による「Send him the message that arrived yesterday.」の表現［123］

主辞駆動句構造文法（Head-Driven Phrase Structure Grammar：HPSG）[98] のような単一化文法（unification-based grammars）の実装を行ったものも現れるが、それらは人手で構築した文法に依存するため、構築に時間がかかってしまう。また概してカバレージも低い。

データ駆動の手法は、テンプレートをベースとした意味解析を行う。複雑な素性構造を無視し浅い統語解析を行う。例えば、Air Traveler Information System（ATIS）は、「Atlanta」のような単語がテンプレートの中の Destination のスロットを埋める確率を計算した［87］。

Riloff ら［116］は、特定のドメインにおいてスロットを埋めるパターンの辞書を構築する手法を、格フレームの自動構築に拡張した。しかしこれらの手法も、自動生成された仮説が適切なものであるか否かを判別するために人手が必要となる。

Blaheta and Charniak (2000)は、Penn Treebank に含まれる Manner や Temporal といったタグを学習するシステムを作成した［5]。これらのタグのうちのいくつかは

FrameNet の意味役割と一致する。しかし Treebank のタグは多くの述語についてそのすべての項を含んでいるわけではない。

本木ら [222] は、CANAL と呼ばれる階層型ニューラルネットワークを構築し、それを用いて英語文を解析した結果を報告している。入力層は 4 種類の表層格（主語、動詞、目的語、with 前置詞句）に対応した 4 つのスロットで構成される。「格出力ユニット」と呼ばれる出力層のユニットには 4 つの深層格（agent, patient, modifier, instrument）に対応する 4 つのスロットが設けられている。入力文は単語表現ベクトルが 4 つつながった 1 つのベクトルで表現され、ネットワークに入力され、動詞を除く各単語がそれぞれ格出力ユニットのうちの 1 つを発火させる。単語が発火させた格出力ユニットの深層格が、その単語の深層格となる。

実験においては、計算機で生成された、名詞 24 語、動詞 4 語からなる、選択制限性を満たす文（深層格が付与可能な文）である正文とそうでない逸脱文の合計 57701 文をコーパスとして用いる。それぞれを学習用データとテスト用データに分割している。格解析結果では、学習データに対しては正答率 99.75%、テストデータに対しては 96.31%という結果となっている。CANAL の理論的基礎となっている Forming Global Representations with Extended Backpropagation（FGREP）法と従来の誤差逆伝播法との比較も行っている。中間ユニット数を同数にした両者のネットワークで学習を行った後の格解析比較テストでは、学習データに対しては前者が正答率 99.94%、後者が 99.87%、テストデータに対しては前者が 95.81%、後者が 64.80%という結果となっている。

6.2 意味役割付与（SRL：semantic role labeling）

FrameNet や PropBank といった意味情報のタグ付けがされた大規模なコーパスが整備されたことにより、高い精度の統計的な深層格解析手法が構築し得るようになった。Gildea [40] らは、英語の述語に対する深層格解析のタスクを意味役割付与（semantic role labeling：SRL）と呼び、機械学習を用いた手法を提示した。これらの手法は Conference on Computational Natural Language Learning（CoNLL）の共通評価タスク [15] によって取り上げられ（2004-2005, 2008-2009）、以後様々な解析手法が検討されている [92]。

6.2.1 項同定と項分類

ほとんどの SRL システムは、統語パーサーによって解析された結果である解析木を

インプットとして使用する。統語パーサーは Collins ら [25] によるものや Charniak [17] によるものが多く使用されている。

項は語の連なりとして現れる。これは入力された解析木に含まれる全てのノード（すなわち句）が項の候補となり得ることを意味する。意味役割付与のタスクは、これらの候補から述語の項を特定するタスクと、特定された項に対して正しい意味役割を付与するタスクの２つに大きく分類される。一般的にこれらのタスクは、それぞれ項同定（argument identification）、項分類（arugument classification）と呼ばれる。

入力文sに含まれるm個の語に$1,\cdots,m$のラベル付けを行う。このとき、候補はこれらの語の組合せとして表現される。これらの語の全ての組合せからラベルへの関数として形式化される。Lを全ての意味役割の集合とする。$NONE$を項でないことを表すラベル、ARGを項であることを表すラベルとする。項同定は以下のように形式化される。

$$2^{1,2,\cdots,m} \rightarrow \{NONE, ARG\}$$

項分類は以下のように形式化される。

$$2^{1,2,\cdots,m} \rightarrow L \setminus \{NONE\}$$

6.2.2 枝刈り（pruning）

全てのノードが項の候補となることに加えて、文に含まれる述語によってどの候補が項になるか否かが異なるため文中の述語の数だけ解析木を処理することとなる。これらの理由により、探索空間は膨大となり処理が困難になる。加えて、ほとんどの候補が負例であるために正例負例の数がアンバランスとなり過学習が引き起こされるために機械学習が有効に機能しない。

この問題に対応するために、Xue ら [150] はシンプルなヒューリスティックによって候補を枝刈り（pruning）する手法を提案した。

この手法は、ある述語に対してその項が存在する範囲は and や or といった接続詞を超えない範囲内に限定されるという観察に基づいている。処理としては、解析木上の調査対象となる述語にポインタを合わせ、ポインタが示す要素の兄弟ノードを項の候補のリストに加える。この処理をポインタを解析木を一段ずつ登りながら繰り返し、ポインタがルートノードに達したら処理を終了することで実現される。

これらの処理によって探索空間が狭められることにより処理の効率が向上する。また、

負例が刈り込まれることによって正例と負例のバランスが取られ機械学習の有効性が向上する。解析木が誤っている場合には精度は低くなるものの、その場合においても意味役割付与の精度が向上することが報告されている。

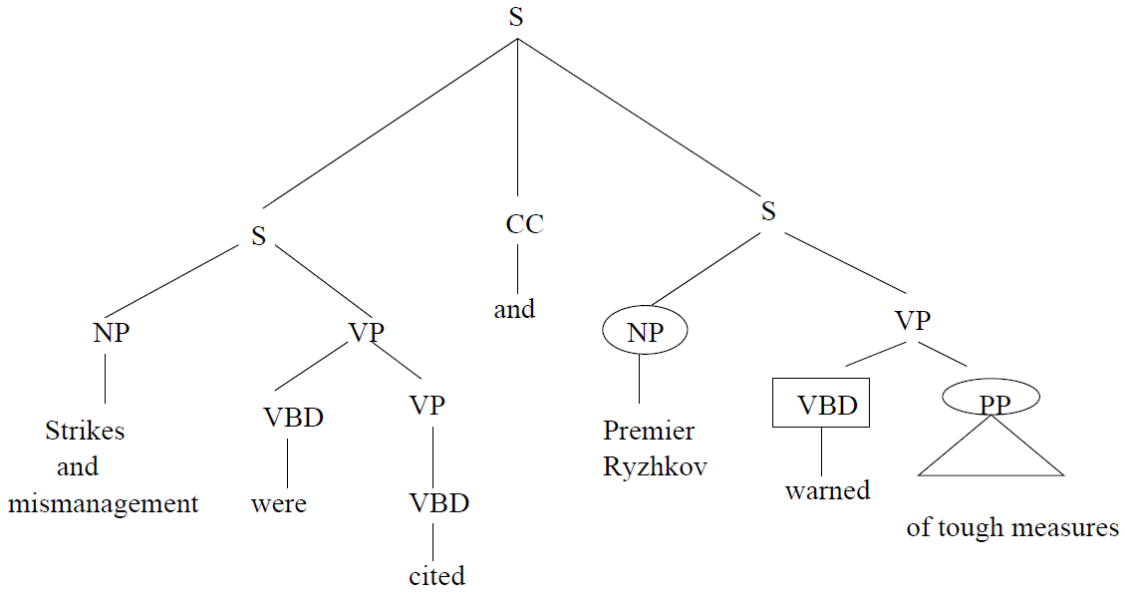


図 6.2: Xue らによる枝刈りの例。項の探索範囲は and を超えない [150]

Punyakanok ら [110] は、統語解析の結果（full syntactic parsing information）は SRL の最初期の処理である枝刈りのプロセスでこそ活きることを示している。Punyakanok らは、項同定、項識別、枝刈りのそれぞれのプロセスにおいて構文木を含む full parsing の情報を用いる場合と、品詞、チャンク、節情報を含む shallow parsing の情報を用いる場合との性能を比較した。その結果、項同定と項識別においては、full parsing の情報を用いた場合も shallow parsing の情報を用いた場合もさほど性能に差が見られなかったのに対し、枝刈りにおいては full parsing の情報を用いた場合の性能は、shallow parsing の情報を用いた場合に比べて F1 値で 4 ポイント程度高かったと報告している。

6.2.3　Joint Model

以下の文章を考える。

(1) Final-hour trading accelerated to 108.1 million shares yesterday.

(1)の文章において、他動詞 accelerated の最初の項である名詞句 Final-hour trading は、動詞以下の文章を考慮しなければ AGENT の意味役割を付与するのが適切に思われる。しかし、後続する項である「to 108.1 million shares」と「yesterday」は、それぞれ TARGET、ARGM-TMP の意味役割を付与されるに相応しく、そうであるとすると他動詞 accelerated が必要とする THEME の意味役割を受ける項が文中に存在しないことになる。それ故に、Final-hour trading は AGENT ではなく THEME の意味役割を受けるほうが適切であると判断される。

このように項の間には意味役割上強い依存関係があり、それを考慮することによって意味役割付与の精度を高めることができる。この項の間の依存関係を統計的モデルで表現したものは、一般に joint model と呼ばれる [136]。

joint model は、SRL の研究の初期から導入されていた。しかしこれらは、初期の Guildea(2002)らの研究成果を大きく超えなかった。それは高々N-gram のマルコフ連鎖（Markov assumption）にとどまっていた。

Punyakanok ら [108] は、文全体が満たすべき制約を導入することで、長距離依存を SRL に導入した。それは例えば、あるノードに non-NONE というラベルが付与された場合は、その子ノードには non-NONE を付与してはならないというものである。具体的には、Punyakanok らは、表 6.2 に示す 10 の制約を利用した。

表 6.2: Punyakanok(2004)らの制約条件 [108]

1	Arguments cannot cover the predicate except those that contain only the verb or the verb and the following word.
2	Arguments cannot overlapwith the clauses (they can be embedded in one another).
3	If a predicate is outside a clause, its arguments cannot be embedded in that clause.
4	No overlapping or embedding phrases.
5	No duplicate argument classes for A0-A5,V.
6	Exactly one V argument per sentence.
7	If there is C-V, then there has to be a V-A1-CV pattern.
8	If there is a R-XXX argument, then there has to be a XXX argument.
9	If there is a C-XXX argument, then there has to be a XXX argument; in addition, the C-XXX argument must occur after XXX.
10	Given the predicate, some argument classes are illegal (e.g. predicate 'stalk' can take only A0 or A1).

これらの制約条件は、それぞれ線形制約条件として定式化することができる。例えば、5 番の制約については、以下のように定式化することができる（S は節の集合を、c は意味役割のクラスをそれぞれ表す）。

$$\sum_{i=1}^{M}[S^i = c] \leq 1$$

(where $[x]$ is 1 if x is true and 0 otherwise.)

制約条件を線形制約条件として定式化することによって、整数線形計画法（integer linear programming：ILP）を用いて依存関係を計算することが可能になる。Punyakanok(2004b)らの報告によれば、この joint model を使用しなかった場合と使用した場合とでは、表 6.3 に示す結果の差異が生じた。

表 6.3: joint model の使用の有無による実験結果の差異［108］

	Precision	Recall	F1
Without Inference	86.95%	87.24%	87.10
With Inference	88.03%	88.23%	88.13

Toutanova ら［136］は、対数線形モデル（log linear model）を利用し、長距離依存のモデル化を行うことで、SRL の精度を向上させた。

表 6.4: Toutanova らの実験結果［136］

Model	Features	CORE		COARSE ARGM		ALL	
		F1	A cc.	F1	Acc.	F1	Acc.
LOCAL	5,201K	90.5	81.4	90.6	76.2	88.4	72.3
JOINT LOCAL	2,193K	90.9	82.6	91.1	78.3	88.9	74.3
LABEL SEQ	2,357K	92.9	86.1	92.6	81.4	90.4	77.0
ALL JOINT	2,811K	94.0	87.6	93.4	82.7	91.2	78.3

Sun ら［126］は、項分類のタスクに、意味役割が持つ序列関係（thematic hierarchy）を利用した自動予測の考え方を導入した。
Fillmore の格文法の枠組みには「文中に Agent があればそれが主語となり、Agent がない場合は Instrument が主語となり、Instrument もない場合は Object（Patient あるいは Theme）が主語となる」という観察がある。これは項の間には順序関係を見出すことができるということである。この観察を定式化すると以下のようになる。

$$Agent \succ Instrument \succ Patient/Theme$$

この視点に基づくと、例えば項 a_i のランクが a_j よりも高い場合、a_i が Patient で a_j が Agent だという付与は間違いだと判断できる。なぜならば、Agent は最も高いロールであり、この付与はその制約に反しているからである。

表 6.5: Sun らが導入した二項関係［126］

1	$a_i \succ a_j$	a_iはa_jより高い。
2	$a_i \prec a_j$	a_iはa_jより低い。
3	$a_i\ AR\ a_j$	a_jはa_iの referenced argument である。
4	$a_i\ RA\ a_j$	a_iはa_jの referenced argument である。
5	$a_i\ AC\ a_j$	a_jはa_iの continuation argument である。
6	$a_i\ CA\ a_j$	a_iはa_jの continuation argument である。
7	$a_i = a_j$	a_iとa_jは同一の意味役割が付与されている。
8	$a_i \sim a_j$	a_iはa_jはArg2-5のいずれからのラベルが付与されているが、同一のラベルではない。

Sun らは、これらの順序関係をリランキング手法と組み合わせて項分類タスクの性能向上を試みた。二つの項の間の順序関係は、文法的な手がかり（syntax clues）から推測される。彼らは、どの序列関係を用いるのが有効かを報告している。A は Agent のランクが最も高いことを、P↑は Patient のランクが最も高いことを、P↓は Patient のランクが最も低いことをそれぞれ表している。SRL(S)は緩い制約に基づくリランキング手法に基づく結果を、SRL(G)は順序関係について正解データ（gold hierarchies）を用いた場合の結果を示している。

表 6.6 は、順序関係の推測実験の結果を示している。全体としては約 96%の精度で順序関係を推測することができている。

表 6.6: 序列関係の有効性の比較［126］

	Detection	SRL(S)	SRL(G)
Baseline	−	94.77%	−
A	94.65%	95.44%	96.89%
A & P↑	95.62%	95.07%	96.39%
A & P↓	94.09%	95.13%	97.22%

表 6.7: 序列関係の推測実験の結果 [126]

Rel	Freq.	BL	P(%)	R(%)	F
$\succ$	57.40	94.79	97.13	98.33	97.73
$\prec$	9.70	51.23	98.52	97.24	97.88
$\sim$	23.05	13.41	94.49	93.59	94.04
$=$	0.33	19.57	93.75	71.43	81.08
AR	5.55	95.43	99.15	99.72	99.44
AC	3.85	78.40	87.77	82.04	84.81
CA	0.16	30.77	83.33	50.00	62.50
All	−	75.75	96.42		

次の文章について考える。

(1) The field goal by Brien changed the game in the fourth quarter.

(1)の文章において、動詞 changed の項としては以下の 3 つが選択される。

(a) The field goal by Brien (A0, the causer of the change).
(b) the game (A1, the thing changing).
(c) in the fourth quarter (temporal modifier).

しかし、この選択によって field goal を行ったのが Brien であることはわからない。これを知るためには前置詞 by に着目しなければならない。また前置詞 in は時間的関係（temporal relation）を示しており、これは動詞を中心とした SRL の結果と一致している。

従来の SRL システムは動詞と名詞の関係を中心に考慮していたが、この例はそこに前置詞を加えることでより潤沢な意味解析ができることを示唆している。この観点から、Srikumar ら [124] は、前置詞に意味役割（preposition role）を与えることで SRL システムを拡張することを提案した。ここで inference problems は整数線形計画法（integer linear programming）として定式化されている。

システムは Punyakanok et al. (2008)らのそれを拡張する形で作られており、実験においてはベースラインとしても用いている。入力データとして Penn Treebank の先頭 500 文を選んで使用している。

実験結果は表 6.9 の通りである。Prep.→SRL は前置詞に付与した意味役割を利用した際の述語の意味役割付与結果を、SRL→Prep.は述語の意味役割付与結果を利用した際の前置詞の意味役割付与結果をそれぞれ表している。

表 6.8: PennTreeBank における前置詞意味役割の統計 [124]

Role	Train	Test
ACTIVITY	57	23
ATTRIBUTE	119	51
BENEFICIARY	78	17
CAUSE	255	116
CONCOMITANT	156	74
ENDCONDITION	88	66
EXPERIENCER	88	42
INSTRUMENT	37	19
LOCATION	1141	414
MEDIUM OF COMMUNICATION	39	30
NUMERIC / LEVEL	301	174
OBJECT OF VERB	365	112
OTHER	65	49
PARTWHOLE	485	133
PARTICIPANT / ACCOMPANIER	122	58
PHYSICAL SUPPORT	32	18
POSSESSOR	195	56
PROFESSIONAL ASPECT	24	10
RECIPIENT	150	70
SPECIES	240	58
TEMPORAL	582	270
TOPIC	148	54

表 6.9: Srikumar らの実験結果［124］

Setting	**SRL** (F1)	**Preposition Role** (Accuracy)
Baseline SRL	76.22	－
Baseline Prep.	－	67.82
Prep. → SRL	76.84	－
SRL → Prep.	－	68.55
Joint inference	77.07	68.39

6.2.4　SRL における語の意味（word sence）の利用

意味役割を推測するうえで、単語の意味は重要な役割を果たす。
我々は「cat」や「dog」が述語「eat」の AGENT となり得ることを容易に推測できるが、これは我々が語の意味を理解しており、それに基づいて意味役割の推測を行うことができるからである。

CoNLL-2008 Shared Task において、Surdeanu ら［131］は、述語の語義曖昧性解消（predicate sense disambiguation）の考え方を導入することで、述語の意味を考慮する SRL システムを提示した。Meza-Ruiz [82］らも、述語の意味を考慮することは最終的な SRL の性能の向上に寄与することを報告している。

しかしこれらの研究においては、述語の局所的な意味にのみ着目するだけで、述語周辺の語の意味も含めた包括的な処理は考慮されていない。それは、文中の語に対して包括的に意味が付与されたコーパスがまだ存在しないためだった。
しかし OntoNotes コーパスが登場したことにより、述語の周辺の語の意味も考慮することが可能になった。

Che ら［19］は、OntoNotes コーパスから語彙素性（word sense feature）を抽出し、それが SRL の性能を向上させることを報告した。

表 6.10: Che らが用いた素性 [19]

Feature	Description	Category
FirstwordLemma	The lemma of the first word in a subtree	Subtree-word related
HeadwordLemma	The lemma of the head word in a subtree	
HeadwordPOS	The POS of the head word in a subtree	
LastwordLemma	The lemma of the last word in a subtree	
PredicateLemma	The lemma of a predicate	Predicate related
POSPath	The POS path from a word to a predicate	Word and predicate related
PathLength	The length of a path	
Position	The relative position of a word with a predicate	
RelationPath	The dependency relation path from a word to a predicate	

表 6.11: Che らが用いた意味抽出の 3 つの手法 [19]

Lemma + Sense	It is the original word sense representation in OntoNotes, such as "dog.n.1". In fact, This is a specialization of the lemma.
Hypernym(*n*)	It is the hypernym of a word sense, e.g. the hypernym of "dog.n.1" is "canine.n.1". The *n* means the level of the hypernym. With the increasing of *n*, the sense becomes more and more general. In theory, however, this strategy may result in inconsistent sense, e.g. word "dog" and "canine" have different hypernyms. The same problem occurs with Basic Concepts method (Izquierdo et al., 2007).
Root Hyper(*n*)	In order to extract more consistent

	sense, we use the hypernym of a word sense counting from the root of a sense tree, e.g. the root hypernym of “dog.n.1” is “entity. n.1”. The *n* means the level of the root hypernym. With the increasing of *n*, the sense becomes more and more special. Thus, word “dog” and “canine” have the same Root Hyper: “entity”, “physical entity”, and “object” with *n* = 1, 2, and 3 respectively.

表 6.10: 素性分類ごとに 3 つの手法を適用した結果 [19]

		Subtree-word related sense	Predicate sense	Sense path
Lemma+Sense		85.34%	86.16%	85.69%
Hypernym(*n*)	1	85.41%	86.12%	85.74%
	2	85.48%	86.10%	85.74%
	3	85.38%	86.10%	85.69%
Root_Hyper(*n*)	1	85.35%	86.07%	85.96%
	2	85.45%	86.13%	85.86%
	3	85.46%	86.05%	85.91%

表 6.11: OntoNotes コーパスを利用した WSJ テストデータにおける実験結果 [19]

	Precision	Recall	F1
without sense	88.38	82.93	85.57
first sense	88.72	83.29	85.92
word sense	89.25	84.00	86.54

表 6.10 から、以下の 3 つのことを見ることができる。

1. 述語の意味素性（predicate sense feature）と意味経路素性（sense path feature）は、どちらも性能の向上に寄与する。
2. Subtree-word related sense は、ほとんど性能向上に寄与していない。これは lemma と POS がすでに十分に Subtree-word related sense に相当する情報を保持しているためだと考えられる。
3. 各素性に対し、手法の違いは性能にほとんど影響を与えていない。これは、語の意味の曖昧性が解消されてしまえば、意味の表現方法は SRL において重要なものではないことを示している。

OntoNotes に所収されている 7 つのニュースソース(ABC, CNN, MNB, NBC, PRI, VOA, WSJ)に対し、実験を行った結果の平均値が表 6.11 である。word sense の使用により性能が向上していることがわかる。

6.2.5 Combinatory Categorial Grammar（CCG）の利用

Boxwell ら [9] は、Combinatory Categorial Grammar（CCG）によって表現された素性を利用した SRL システム「Brutus」を発表した。
CCG は、語彙レベルでの関連性を表現することに重点を置いた形式文法の一種である。CCG は、局所的、大域的な依存につき単一の表現形式で表現できるのが特徴で、動詞と項の間の同一の関連性を示すにも関わらずそれを表現する経路（tree path）が複数存在する場合があり、従来の手法ではこれはデータスパースネスの問題を引き起こしていたが、CCG はこれら複数の tree path を単一の仕方で表現できるためこの問題を回避できる。

テキストは C&C CCG parser [23] によってパースされる。最大エントロピー法において、表 6.X に示す素性を使用した。表 6.12 のうち 1～11 の素性は項同定と項分類に、12～16 の素性は項分類にのみ使用している。また、異なるパーサは異なるパースエラーを出すという考えに基づき、頑健性を向上させるために Charniak's parser [18] による解析結果から抽出した素性も別途加えている。

実験に際しては、同じく CCG を使ったシステムである Gildea and Hockenmaier ら [38] のシステムとの比較実験と、当時の gold standard であった Punyakanok ら [110] のシステムとの比較実験を行っている。それぞれの比較において、素性を treebank から抽出した結果と、C&C CCG parser によって抽出した結果を報告している。

表 6.12: Boxwell らが使用した素性 [9]

1	Words	Words drawn from a 3 word window around the target word,[4] with each word associated with a binary indicator feature.
2	Part of Speech	Part of Speech tags drawn from a 3 word window around the target word, with each associated with a binary indicator feature.
3	CCG Categories.	CCG categories drawn from a 3 word window around the target word, with each associated with a binary indicator feature.
4	Predicate.	The lemma of the predicate we are tagging. E.g. *fix* is the lemma of *fixed*.
5	Result Category Detail	The grammatical feature on the category of the predicate (indicating declarative, passive, progressive, etc). This can be read off the verb category: declarative for *eats*: *(s[dcl]\np)/np* or progressive for *running*: *s[ng]* \np.
6	Before/After.	A binary indicator variable indicating whether the target word is before or after the verb.
7	Treepath	The sequence of CCG categories representing the path through the derivation from the predicate to the target word. For the relationship between *fixed* and *car* in the first sentence of figure 3, the treepath is (s[dcl]\np)/np>s[dcl]\np<np<n, with > and < indicating movement up and down the tree, respectively.
8	Short Treepath	Similar to the above treepath feature, except the path stops at the highest node under the least common subsumer that is headed by the target word (this is the *constituent* that the role would be marked on if we identified this terminal as a role-bearing word). Again, for the relationship between *fixed* and *car* in the first sentence of figure 3, the short treepath is (s[dcl]\np)/np>s[dcl]\np<np.
9	NP Modified	A binary indicator feature indicating whether the target word is modified by an NP modifier.
10	Subcategorization	A sequence of the categories that the verb combines with in the CCG derivation tree. For the first sentence in figure 3, the correct subcategorization would be *np,np*. Notice that this is not necessarily a restatement of the verbal category – in the second sentence of figure 3, the correct subcategorization is *s/(s\np),(np\np)/(s[dcl]/np),np*.

11	PARG feature	We follow a previous CCGbased approach (Gildea and Hockenmaier, 2003) in using a feature to describe the PARG relationship between the two words, if one exists. If there is a dependency in the PARG structure between the two words, then this feature is defined as the conjunction of (1) the category of the functor, (2) the argument slot that is being filled in the functor category, and (3) an indication as to whether the functor (→) or the argument (←) is the lexical head. For example, to indicate the relationship between *car* and *fixed* in both sentences of figure 3, the feature is (s\np)/np.2.→.
12	Headship	A binary indicator feature as to whether the functor or the argument is the lexical head of the dependency between the two words, if one exists.
13	Predicate and Before/After	The conjunction of two earlier features: the predicate lemma and the Before/After feature.
14	Rel Clause	Whether the path from predicate to target word passes through a relative clause (e.g., marked by the word *'that'* or any other word with a relativizer category).
15	PP features	When the target word is a preposition, we define binary indicator features for the word, POS, and CCG category of the head of the topmost NP in the prepositional phrase headed by a preposition (a.k.a. the '*lexical head*' of the PP). So, if *on* heads the phrase *'on the third Friday'*, then we extract features relating to *Friday* for the preposition *on*. This is null when the target word is not a preposition.
16	Argument Mappings.	If there is a PARG relation between the predicate and the target word, the argument mapping is the most likely predicted role to go with that argument. These mappings are predicted using a separate classifier that is trained primarily on lexical information of the verb, its immediate string-level context, and its observed arguments in the training data. This feature is null when there is no PARG relation between the predicate and the target word. The Argument Mapping feature can be viewed as a simple prediction about some of the non-modifier semantic roles that a verb is likely to express. We use this information as a feature and not a hard constraint to allow other features to overrule the recommendation made by the

		argument mapping classifier. The features used in the argument mapping classifier are described in detail in section 7

表 6.13: Boxwell らのシステムと Gildea and Hockenmaier(2003)らのシステムの比較実験 [9]

	Precision	Recall	F1
G&H(treebank)	67.50%	60.00%	63.50%
Brutus(treebank)	88.18%	85.00%	86.56%
G&H(automatic)	55.70%	49.50%	52.40%
Brutus(automatic)	76.06%	70.15%	72.99%

表 6.14: Boxwell らのシステムと Punyakanok et al.,(2008)らのシステムの比較実験 [9]

	Precision	Recall	F1
P. et al(treebank)	86.22%	87.40%	86.81%
Brutus(treebank)	88.29%	86.39%	87.33%
P. et al(automatic)	77.09%	75.51%	76.29%
Brutus(automatic)	76.73%	70.45%	73.45%

6.2.6 ドメイン外（out-of-domain）データへの対応

一般的な SRL システムにおいては教師有り学習の手法が用いられており、タグ付きコーパスを教師データとして学習を行い、多くの場合同じコーパスからラベルを除去したものをテスト用データセットとしてその性能を検証していた。

しかし、各々のコーパスはデータソースとして新聞記事などを用いており、それ故に特定のドメイン（domain）に偏ったものとなっている。このことは、過学習の問題を引き起こし、あるドメインのデータを用いて学習したシステムをドメイン外（out-of-domain）のデータに対して用いると性能が低下するという事例が報告されている。

CoNLL 2005 においては､例えば訓練用コーパスとして Wall Street Jounal を用い、テスト用として Brown コーパスを用いるといったように、訓練用のコーパスとは異なるコーパスをテスト用に用いた場合に深刻なパフォーマンスの低下を見せた。
Johansson ら［54］は、複数のシステムにおいて訓練用のコーパスとは異なるコーパスをテスト用に用いた場合に、項分類タスクにおいて 20%近くの精度の低下が見られたことを報告している。

ドメイン外のデータに対してもドメイン内のデータと遜色ない性能を達成するシステムを、汎ドメイン（open-domain）なシステムと言う。

Pradhan ら［105］は、彼らが作成した SRL システムである ASSERT に対し、Wall Stream Journal にアノテーションした PropBank で学習を行い、そのシステムでもって Brown Corpus にアノテーションした PropBank を解析した。実験結果は、項同定の結果は良いが、項分類の結果は芳しくないというものだった。この結果から、構文的素性が項同定に寄与し、語彙的意味的素性が項分類に有効であることがわかった。

Zapirain ら［159］は、選択選好を素性として用いることで項分類タスクの精度が向上することを報告した。実験に際しては、当時の最新のシステムであった SwiRL [130] を拡張する形で選択選好に関わる素性を追加した。WordNet を利用した選択選好モデルとして Resnik（1993）が提唱した類似度測定法に基づいた選好モデルに基づく素性を追加したものをSP_{Res} 、Zapirain ら［158］の選好モデルに基づく素性を追加したものをSP_{wn} 、また距離に基づく類似度（distributional similarity）を利用したモデルとしてコサイン類似度に基づく素性を追加したものを$SP_{sim_{cos}}$、Padó らが提唱した optimal dependency-based model [93] に基づく素性を追加したものを$+SP_{sim_{Jac}}$として性能比較を行った。結果として、選択素性を用いない場合と比較して精度の向上が見られた。

out-of-domain データに対する実験も行っており、学習用データである PropBank とは別に BrownCorpus に対しても実験を行っている。その結果、ドメイン内データおよびドメイン外データの双方において性能の向上が見られたことを報告している。

表 6.15: Zapirain らの実験結果 [159]

	WSJ-test			Brown		
	Core	Adj	All	Core	Adj	All
SwiRL	93.25	81.31	90.83	84.42	57.76	79.52
$+SP_{Res}$	93.17	81.08	90.76	84.52	59.24	79.86
$+SP_{wn}$	92.88	81.11	90.56	84.26	59.69	79.73
$+SP_{sim_{Jac}}$	93.37	80.30	90.86	84.43	59.54	79.83
$+SP_{sim_{cos}}$	93.33	80.92	90.87	85.14	60.16	80.50
$+SP_{sim^2_{Jac}}$	93.03	82.75	90.95	85.62	59.63	80.75
$+SP_{sim^2_{cos}}$	93.78	80.56	91.23	84.95	61.01	80.48
Meta	94.37	83.40	92.12	86.20	63.40	81.91

一般的な SRL システムは、タグ付きコーパスから学習を行う。フレーム意味論に基づく（Frame-based）システムは通常 FrameNet を用いるが、ドメイン外データに対する一般化能力に欠けている。 Croce ら [26] は、意味的類似性を確率モデルとして捉えることを通して最新のフレーム意味論に基づくシステムをドメイン外データに対する一般化能力を持つように拡張した。FrameNet の Frame elements が意味役割として付与された文において、意味役割が付与された節の主辞（head）を k 近傍法（k-nearest neighbor）によってクラスタリングすることで一般化を得ようとしている。

FrameNet のタグが付与された British National Corpus（BNC）を学習用データとして用い、out-of-domain のテスト用データとして Nuclear Threat Initiative（NTI）と American National Corpus（ANC）を使用した。パーサには LTH parser (Johansson and Nugues, 2008a)を使用している。

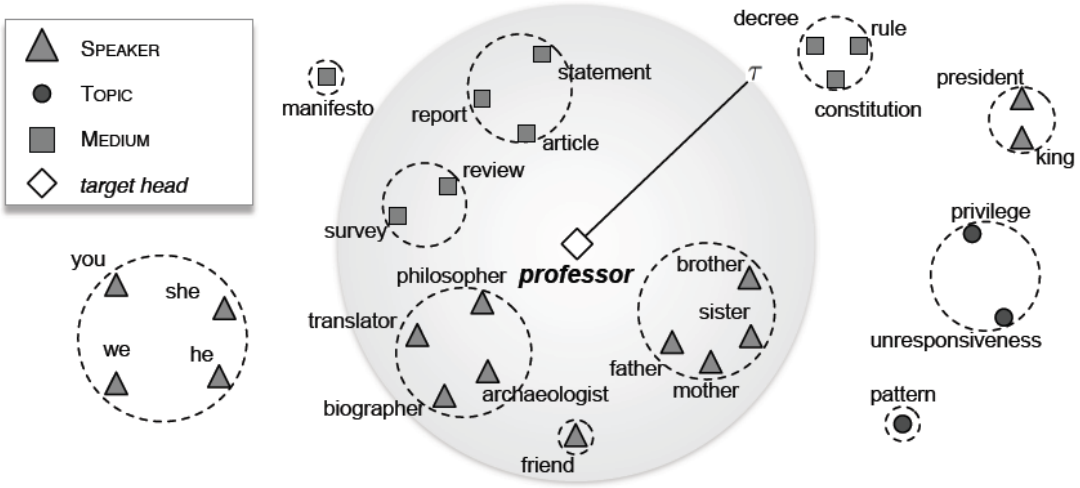


図 6.3 主辞「professor」を k 近傍法で分類する [26]

表 6.16: Croce らの実験結果 [26]

Model	FN-BNC	NTI	ANC
Local Prior	43.9	50.9	50.4
Global Prior	67.7 (+54.2%)	75.9 (+49.0%)	68.8 (+36.4%)
Distributional	81.1 (+19.8%)	82.3 (+8.4%)	69.7 (+1.3%)
Backoff	84.6 (+4.3%)	87.2 (+6.0%)	76.2 (+9.3%)
Backoff + HMMRR	86.3 (+2.0%)	90.5 (+3.8%)	79.9 (+5.0%)
(Johansson&Nugues, 2008)	89.9	71.1	—

Huang ら [51] は、従来の意味役割付与システムが汎ドメインでない原因を、素性の表現方法に問題があるからだとした。例えば「bank」「CEO」といった項と述語の関係を素性として表現した場合、新聞記事などには適合するシステムになるだろうが、それらの語が出現しない生命科学の文章などには適用できない。Huang(2010)らは、隠れマルコフモデル（HMM : Hidden Markov Model）によって表現された素性を用いることにより、一回の学習で色々なジャンルの文章を解析できる汎ドメインな意味役割付与システムの構築を試みた。彼らが使用したモデルは Multi-Span-HMM と呼ばれる。

彼らは、ラベルなしコーパスを用いて、教師なし学習を行った隠れマルコフモデルに

基づいた素性と、path 素性と呼ばれる素性を用いた。path 素性とは、2 つの語の間に含まれるものを表したもので、「The HIV infection rate is expected to peak in 2010」という文章において、項候補「rate」と述語「peak」の word path は「is expected to」となり、POS path は「VBZ VBD TO」となる。

ただ、この path 素性はデータスパースネスに陥りやすい。そこで彼らは Span-HMM 素性を導入して、この問題の解決を図った。

実験に際しては、ドメイン内コーパスとして WSJ を、ドメイン外コーパスとして BrownCorpus を使用している。表 6.X は、その他の最新システムとの F1 値の比較である。ドメイン外テキストである Brown Corpus を用いた際の性能の低下が、その他のシステムよりも低いことがわかる。

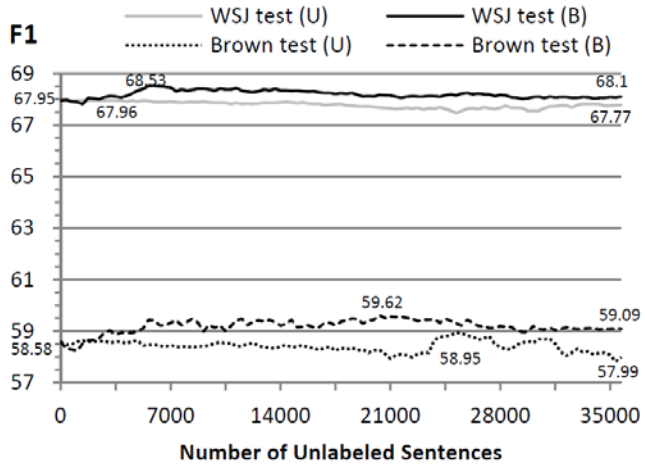
表 6.17: Multi-Span-HMM とその他の最新のシステムとの比較実験［51］

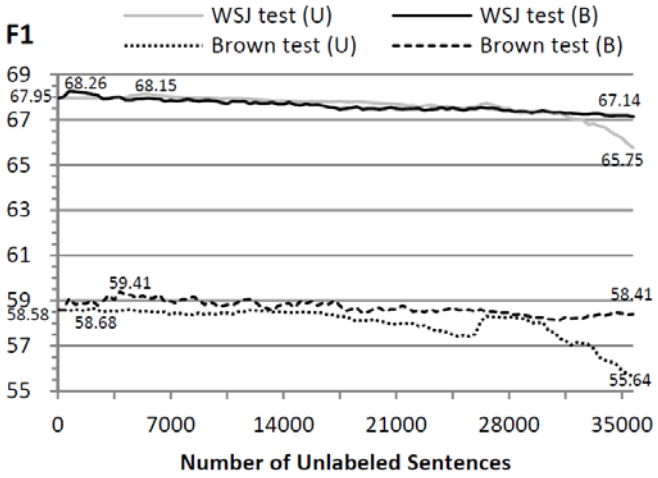
System	WSJ	Brown	Diff
Multi-Span-HMM	79.2	73.8	5.4
Toutanova et al. (2008)	80.8	68.8	12.0
Pradhan et al. (2005)	78.6	68.4	10.2
Punyakanok et al. (2008)	79.4	67.8	11.6

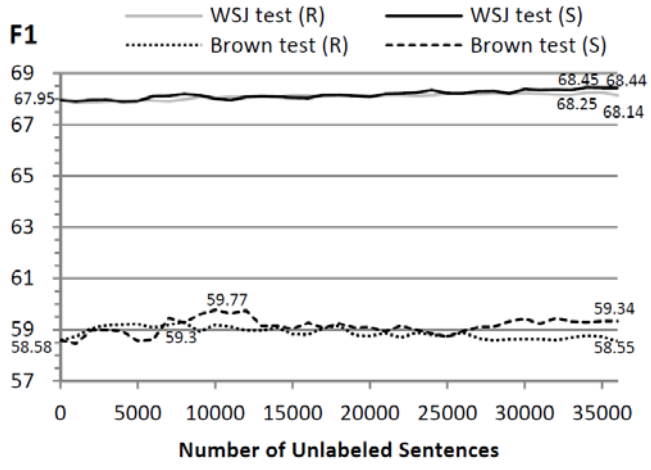
Samad ら［118］は、半教師有り学習（semi-supervised learning methods）を用いることで、この問題の解決を試みた。

学習には、Self-training アルゴリズム［153］を用いる。

実験には CoNLL 2005 shared task で用意された PropBank のデータを用いる。39,832 文から 4,000 文をランダムにシードデータとして選ぶ。ドメイン外データとして Open American National Corpus（OANC）を用いる（ただし、ドメインごとのデータ量のバランスを保つために biomed section は除外している）。


WSJ test (U)
WSJ test (B)
Brown test (U)
Brown test (B)
F1
69
67
65
63
61
59
57
68.53
68.1
67.95
67.96
67.77
59.62
59.09
58.58
58.95
57.99
0
7000
14000
21000
28000
35000
Number of Unlabeled Sentences


WSJ test (U)
WSJ test (B)
Brown test (U)
Brown test (B)
F1
69
67
65
63
61
59
57
55
68.26
68.15
67.95
67.14
65.75
59.41
58.41
58.58
58.68
55.64
0
7000
14000
21000
28000
35000
Number of Unlabeled Sentences


WSJ test (R)
WSJ test (S)
Brown test (R)
Brown test (S)
F1
69
67
65
63
61
59
57
68.45
68.44
67.95
68.25
68.14
59.77
59.34
58.58
59.3
58.55
0
5000
10000
15000
20000
25000
30000
35000
Number of Unlabeled Sentences

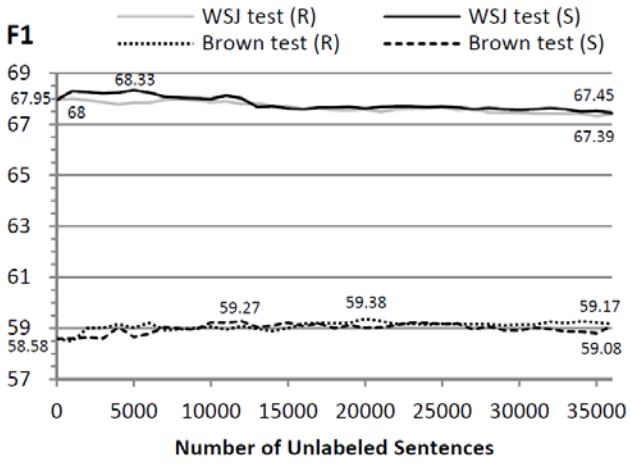


図 6.4: Samad らの実験結果［118］

あらゆるドメインに対応するタグ付きコーパスがあればドメイン外データの問題は解消できるかもしれない。しかし、タグ付きのコーパスを作成するには多大な時間がかかる。それ故に、タグのない生コーパスから教師なし学習の手法を用いる意味役割付与システムが提案されている。

Lang ら［66, 67］は、項分類タスクをグラフの分割問題として定式化し、その最適解を教師なし学習を用いて求める手法を提案した。

頂点を動詞の項とし、辺の重みはそれらの項の類似度とする。生コーパスを MaltParser で解析した結果の依存構造木をインプットとする。

項識別においては、Lang ら［68］の手法を用いて、項の候補を選別する。この性能は、88.1%の precision と、87.9%の recall である。

項の類似度は以下の 3 つの規則に基づいて求められる。

(1) whether the instances are lexically similar.
(2) whether the instances occur in the same syntactic position.
(3) whether the instances occur in the same frame.
(i.e., are arguments in the same clause).

実験に際しては、PropBank から Wall Street Journal の部分を抜き出したものを使用している。PennTreebank が持つ依存構造は、比較のための gold standard とする。

表 6.18: Lang らの実験結果 [67]

	Syntactic Function			Latent Logistic			Split-Merge			Graph Partitioning		
	PU	CO	F1	PU	CO	F1	PU	CO	F1	PU	CO	F1
auto/auto	72.9	73.9	73.4	73.2	76.0	74.6	81.9	71.2	76.2	82.5	68.8	75.0
gold/auto	77.7	80.1	78.9	75.6	79.4	77.4	84.0	74.4	78.9	84.0	73.5	78.4
auto/gold	77.0	71.0	73.9	77.9	74.4	76.2	86.5	69.8	77.3	87.4	65.9	75.2
gold/gold	81.6	77.5	79.5	79.5	76.5	78.0	88.7	73.0	80.1	88..6	70.7	78.6

Titov ら [133] は、項の統語的特徴（syntactic signatures）を、意味役割と対応したクラスタに分類するクラスタリング問題として捉えた上で、2 つのベイジアンモデルを提案した。

ひとつ目のモデルは、中華料理店過程（CRP : Chinese Restaurant Process）を利用して、それぞれの述語において独立にクラスタリングを行う。2 つ目のモデルは、2 つの統語的特徴（syntactic signature）間の距離を距離依存中華料理店過程（distance-dependent CRP）を用いてモデル化したものである。

表 6.19: Titov らの実験結果 [133]

	gold parses			auto parses		
	PU	CO	F1	PU	CO	F1
Factored + Br	87.8	82.9	85.3	85.8	81.1	83.4
Coupled + Br	89.2	82.6	85.8	87.4	80.7	83.9
SyntF	83.5	81.4	82.4	81.4	79.1	80.2

6.2.7 ドメインを固定した意味役割付与

より広範なドメインに対応する汎用性を備えた SRL システムの研究が進む一方で、あえて特定のドメインに限定することで実用的なシステムを作成しようとする研究もある。

近年増え続ける生体医学分野のテキストに対して、機械的に情報抽出を行う処理に関

して研究が行われている。生体医学分野において主要となる抽出タスクは、タンパク質や遺伝子間の相互作用を抽出することだ。このような関連を抽出する手法としては、共起ベース、パターンベース、機械学習ベースの手法があったが、そのいずれもが複雑なテキストからの情報抽出において課題を抱えている。それらのシステムは「proteins」「genes」といった単語とそれらを関連付ける動詞しか抽出できない。それ故に、副詞や前置詞が表現している場所や時間といった情報を取りこぼしている。

このような情報を取得するために、Tsai ら [138, 139] は生体医学分野に特化した SRL システムを開発した。

PennTreebank 形式のタグが付与された生体医学分野のコーパスである、GENIA corpus [58] をベースとして、PropBank 形式のタグを付与したコーパス BioProp を作成した。

項分類のプロセスにおいては、最大エントロピーモデルを使用している。

表 6.20: Tsai らの実験結果 [138]

Configuration	Training	Test	P	R	F
Exp 1a	PropBank	PropBank	90.47	82.48	86.29
Exp 1b	PropBank	BioProp	75.28	56.64	64.64
Exp 2a	PropBank	BioProp	74.78	56.25	64.20
Exp 2b	BioProp	BioProp	88.65	85.61	87.10
Exp 3a	BioProp	BioProp	88.67	85.59	87.11
Exp 3b	BioProp	BioProp	89.13	86.07	87.57

生体医学分野のタグ付きコーパスは不足している。そしてそれを手で作成するには人的・費用的コストがかかる。Dahlmeier ら [27] は、ニュースドメインのタグ付きコーパスを元に、生体医学分野に特化した意味役割付与システムを設計する手法を開発した。

特徴は、ドメイン適合（domain adaptation）と彼らが呼んでいる手法で、ニュースドメインのタグ付きコーパスを用いた教師あり学習と、生体医学分野のドメインのテキストを用いた教師なし学習を組み合わせることで、ドメインを超えた意味役割付与を実現している。

ドメイン適合は、以下の 3 つの手法から構成される。

1. Instance weighting（INSTWEIGHT）: あるドメインのデータで学習された素性とクラスの同時分布は、別のドメインにおいては異なるが、これを調整する。

2. Augment method（AUGMENT）: 素性ベクトルをより高次元の素性空間にマップすることで、素性空間の拡張（feature space augmentation）を行う。マッピングの方法は、対象となる素性ベクトルが source domain で学習されたものか target domain で学習されたものかに依存する。
3. Instance pruning（INSTPRUNE）: target domain で学習された分類器を用いて source domain のテキストに対してラベル分類を行う。ラベル識別に失敗した文章の上位 N 個につき、source domain から除外する。これは、この失敗した文章は target domain と source domain で著しく異なるものであるため分類器の学習に悪影響を与えるという推測に基づいている。文章を除外した後の source domain のテキストを用いて、分類器を学習する。

実験に際しては、以下のベースラインとなる 3 種類のアルゴリズムを比較実験のために追加で用いる。

1. Source only（SRCONLY）: source domain のデータだけで分類器を学習する。
2. Target only（TRGTONLY）: target domain のデータだけで分類器を学習する。
3. Source and Target（ALL）: source domain と target domain のデータの両方を用いて分類器を学習する。

表 6.21: Dahlmeier らの実験結果 [27]

Algorithm	0	8	16	24	32	40	60	80	160	240	320	356
SRCONLY	76.46	–	–	–	–	–	–	–	–	–	–	–
TRGTONLY	–	57.04	68.03	71.84	73.24	74.69	77.77	78.65	82.03	83.11	84.15	84.44
ALL	–	77.65	78.83	79.57	79.93	80.72	81.00	81.51	81.66	83.28	83.74	84.12
INSTWEIGHT	–	67.62	74.04	75.50	76.47	78.18	78.99	80.49	82.44	83.33	84.06	83.57
AUGMENT	–	76.98	78.86	79.04	78.34	80.72	81.60	82.04	82.73	84.01	85.07	85.03
INSTPRUNE	–	75.49	78.99	79.30	80.05	80.96	81.92	82.47	83.55	84.37	85.02	85.38

マイクロブログサービスである Twitter には、1 日当たり 5,000 万以上のツイートが投稿されるが、そのうち 4%がニュース関連のものだという調査がある。それらニュースに関連するツイートは、ニュース文を引用しているもの（news excerpt）とそうでないもの（news tweet）に分けられる。ニュース関連ツイートを 1000 件ランダムにサンプリングすると、そのうちの 865 件が news tweet であった。

news tweet は、現在世界で何が起こっているかをリアルタイムで知ることができる

重要な情報である。そのため、ツイートからの情報抽出を目的として SRL を行うことを考える。

Liu ら [73] は、既存の意味役割付与システムを news tweet に対して用いたが、結果が芳しくなかった。そのため、彼らは、自己教師あり学習（self-supervised learning）アプローチを用いて、Twitter のツイートを解析するためのドメイン固有の SRL システムを作成した。

実験に際しては、既存の意味役割付与システムを用いてタグを付けた 10,000 件の news tweet をトレーニングデータとして、1,100 件の news tweet に手動でタグを付けたデータを gold-standard として使用している。Meza-Ruiz [83] らのシステムをベースライン（SRL-BS）として、彼らのシステム（SRL-TS）と性能を比較している。

表 6.22: Liu らの実験結果 [73]

	Precision	Recall	F-Score
SRL-BS	36.0%	54.5%	43.3%
SRL-TS	78.0%	57.1%	66.0%

6.2.8 英語以外の言語における意味役割付与

意味役割付与の研究は英語を対象としたものが多いが、近年は英語以外の言語を対象とした研究報告も増えている。その中でも中国語を対象とした研究が特に増えている。

Xue [151] は、Chinese PropBank と Chinese Nombank を使用して中国語のテキストに対して意味役割付与を行った。
Chinese PropBank は量の面で英語の PropBank よりも小さいにも関わらず、人手でパースを行ったアウトプットに対して実験を行った結果は最新の英語の意味役割付与システムの結果と拮抗するという報告をしている。しかしながら、パーサーによるアウトプットを使用した場合の性能は、英語の意味役割付与システムに比べて著しく低かったとも報告している。

中国語構文パーサーの精度の向上は、意味役割付与システムの精度の向上に必須である。Zhuang ら [160] は、同一のインプットを複数のパーサーで解析し、それらのアウトプットをマージすることで SRL システムの精度を向上させようと試みた。結果は、Xue [151] のものと比較して F1 値で 11.55%高かった。

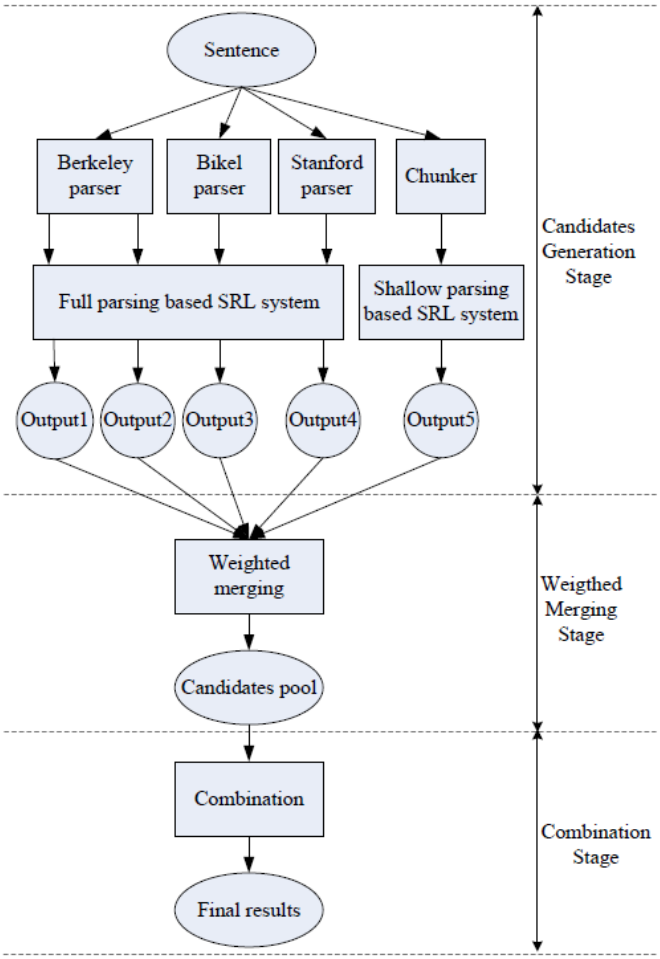


図 6.5: Zhuang らのシステムのアーキテクチャ [160]

表 6.23: Zhuang らの実験結果 [160]

	POS	P(%)	R(%)	F1
(Xue, 2008) [151]	auto	76.80	62.50	68.90
(Sun et al., 2009) [125]	gold	79.25	69.61	74.12
(Zhuang, 2010) [160]	auto	83.02	78.03	80.45

既存の多くの中国語の意味役割付与システムは、全体的な統語解析（full syntactic parses）を利用しているが、浅い統語解析（shallow syntactic parses）を利用する研究もある。Sun ら［125, 127］は、Chen ら［21］の中国語における shallow parsing の手法を土台とした意味役割付与システムを提案した。彼らの枠組みの中で意味役割付与タスクは、IOB2 形式［134］で表現されたチャンク（syntactic chunks）を意味役割に対応付けるタスクとして捉えられる。

実験では 93.42%の精度を出しており、これは gold parse をベースとした結果である 94.68%に迫っている。これは、中国語においては、異なる意味タイプを持つ項はパースをせずともそれを取り出せる程度に特徴的な単語境界を持っているためだと考えられる。

Li ら［71］は、統語解析と意味役割付与を組み合わせることで、中国語意味役割付与タスクの精度向上を試みた。これは 2 つのレベルで実現される。ひとつは、統語解析を統合するアプローチを意味役割付与に適用すること。もうひとつは、意味役割に関連した素性（semantic role related features）を統語解析モデルに組み合わせることで意味情報をより良く捉えられる統語解析モデルを作るというものだ。

中国語意味役割付与システムは基本的に英語のそれを土台として設計されているため、素性についても英語のものを踏襲している。Sun [128] は、中国語固有の事情を考慮した素性を追加で用いることで、当時の最新であった Xue［151］のシステムの 92.00%という F 値に対して 1.49%の性能向上を実現した。

中国語には Chinese PropBank のような意味役割が付与されたコーパスが存在するが、世界にはそのようなコーパスが存在しない言語の方が多い。そのような言語のひとつであるウルドゥー語に対して、Mukund［90］らは、英語とウルドゥー語の対訳コーパスを利用した意味役割付与のアプローチを提案している。これは対訳コーパスに所収された英文に対して意味役割付与を行い、その意味ラベルを対応するウルドゥー語に付与することでウルドゥー語にの意味役割付与コーパスを構築するというアプローチである。得られたコーパスを用いた意味役割付与においては、短い文章に対しては 92%の精度を実現すると報告している。

英語を対象として作成された FrameNet を参考に、スウェーデン語の FrameNet が作成されているが、そのサイズはまだ小さい。そのため、スウェーデン語の FrameNet を用いて意味役割付与システムを作成するのは難しい。Johansson［55］は、英語の FrameNet とスウェーデン語のそれの間に対応関係を定義し、その対応関係を元に英語の FrameNet をスウェーデン語の解析に利用する手法を提案している。

ベースラインとなるのは、彼らが作成したオンライン学習に基づく線形分類器を用いたスウェーデン語用の意味役割付与システムであり、英語の FrameNet との対応関係に基づく今回の手法による実験結果は、ベースラインの 54.4%に対して、5.2%の向上を見せた。

6.2.9 日本語における意味役割付与

日本語における意味役割付与の研究は、PropBank のような意味役割が付与された大規模なコーパスが不足していることもあり、英語のそれにおけるほど活発ではなく、また機械学習の手法を用いたものよりも規則ベースのものが多い。

意味解析を行う場合に重要な役割を果たすのが付属語と意味素（名詞が持つ概念を木構造で階層的に表現したデータで、意味素はそれぞれ対応する一つ以上の深層格の情報を持っている）であるが、これらは単独でいくつもの深層格と対応するため、これらだけで深層格を一意に抽出するのは困難である。浪岡ら [224] は、これらの情報を結合価パターン（用言ごとにそれに係る格要素の組合せを定義したデータのことで、格フレームとほぼ同義）を用いてランク付けして競合解消を行う意味解析手法を提案している。結合価パターンにおいてはひとつの用言に対して複数のパターンが掲載されている場合があるため、それらの候補を一つに絞り込むのがこの研究の核である。判定においては、候補として挙がった結合価パターンそれぞれに対して評価表を作成し、その値が最も高かったものをその文における用言の結合価パターンとする。

評価は処理対象の文が持つ格助詞と意味素に関する規則的判定によって行われる。結合価パターンは、手動によって今回の実験用に作成されている。深層格のパターンは 30 種類である。実験においては、NHK のニュースの読み原稿から「非交差の原理」を満足している 105 文をインプットとし 94.0%の解析精度を得ている。

大石ら [203] は、従来の研究が個々の動詞の取り得る格を一つ一つ見ていくというアプローチでは、動詞と格が取り得るすべての組み合わせに対応することは難しいと指摘した。例えば深層格の種類を 10 個、動詞が取る格要素の数を 3 とした場合、動詞と格の組合せは一文一格の原理に従うと 10×9×8=720 通りにもなる。

そのため、大石らは、処理対象となるデータの全体としてどのような格のパターンがあるかを事前に分類しておき、動詞句としての意味がそのうちのどれに属するかを考える手法を提示している。処理対象となるデータは EDR コーパスより抽出した 858 種類の動詞のデータで、データ構造は、動詞が持つ格要素の表層格、深層格（概念関係子）、品詞の情報からなる。このデータに出現する格のパターンを事前に取得する。

実験においては、最初に意味マーカによる制限によってのみ（すなわち名詞シソーラスと助詞パターンのみを使用して）深層格解析を行い 80.32%の項が EDR 共起辞書の概念関係子と一致した。この処理で不一致となったデータを元に「動詞の意味分類と深層格の対応表」を作成し（すなわち動詞シソーラスを使用し）、これを利用して実験を行った結果、一致率が 81.97%に向上した。動詞シソーラスを用いても著しい一致率が見られなかったのは同一の動詞の同じ用法の同じ格に対しても、人間の判断のゆれによって異なる概念関係子が付されている例が多数あったためである。作成したシステムは格パターンごとに深層格を一意に決定してしまうため、このようなゆれには対応できない。このように同一の動詞の同じ用法に複数の深層格が付与されている場合（すなわちゆれがある場合）に、どちらかの深層格が得られれば正解とすれば一致率は 98.6%に向上する。

小山ら［192］は、入力文の動詞の情報を用いない解析方法を提案している。この手法の根底には、未知の動詞の深層格対応は既知の動詞の深層格対応と似ているか同じであるという前提がある。この手法では、名詞と表層格（格助詞）のペア（つまり格要素）と深層格との対応をあらかじめ「深層格対応データベース」にストックしておき、解析する際には入力文から名詞と表層格のペアを抜き出し、ストックされた情報から深層格を取得する。「深層格対応データベース」には、EDR 日本語動詞共起パターン辞書から抽出した動詞と「名詞と表層格」および深層格の情報が出現頻度と共に格納されている。これに対して「名詞カテゴリ」（ある動詞に係る名詞のうちで、同じ深層格に対応する全ての名詞の上位概念）と表層格の組をキーとして検索を実施し、深層格の候補を得る。その後、頻度を元に最終的な深層格を選択する。

実験におけるインプットは、EDR 日本語共起辞書から得ており、動詞 100 個を含む 648 文である（事例文の取得手法は大石ら［203］の手法に準じる）。結果としては 91.5%の確率で正解が含まれる 1 つ以上の深層格対応の候補を生成でき、70.8%の正解率で正しい深層格の対応が行えている。

幼児が単語と意味を対応付ける際には、単語に対応する意味の様々な論理的可能性を全て吟味せず、内的な制約により吟味する対象の可能性を狭めている。また、論理的には一意に決定できない事例に対しても、単なるランダム選択により決定しているわけではなく、内的な選好により決定を行っている。このような内的な制約や選好により、幼児は外部から明示的な正解を与えられることなく単語と意味を対応付けることができると考えられている。このような認知科学からの報告に基づき、渋木ら［187］は、動詞の内的な制約と選好を表層格と深層格の対応付けに応用する手法を提示している。動詞の選好は分類語彙表と EDR コーパスを元にした計算によって被下位範疇化確率 $P(c,d)$として定義される。

$$P(c,d) = \frac{F(c,d)}{\sum_{k \in D} F(c,k)}$$

ここで$P(c,d)$は概念 c が深層格 d を取る確率であり、$F(c,d)$は意味フレーム中で概念 c が深層格 d と共起した回数である。また、D は深層格の集合である。この$P(c,d)$を被下位範疇化確率と定義し、概念 c が深層格 d を伴って下位範疇化される選好の指標として用いる。

動詞の選好によって候補となった深層格に対し、フィルモアが提示した、一つの文章において格はそれぞれ一つしか出現しないという「一文一格の原理」に基づいた推測によって候補となった深層格を更新することを入力文中の全ての深層格が推測されるまで繰り返す。

この手法においては、概念の意味の特定は分類語彙表に依存している。そのため分類語彙表に登録されていない単語の解析について課題がある。渋木ら［188］は、コーパスから助詞や語順という表層的な情報を元にクラスタリングを行い、その結果を分類語彙表の代わりに使用するという手法を提案している。深層格選好が助詞の出現頻度および語順に反映されていると仮定する。この仮定に基づいて得られたクラスタリング結果を用いた深層格推測実験の結果は分類語彙表を用いたそれよりも高く、本手法の有効性を示している。この手法を渋木は DCAPR（Deep Case Analysis based on Preference of deep case and Regularization）［190］と呼び、日本語のみならず英文への適用においても有効性が見られたと報告している［189］。

下村ら［165］は、入力文から述語に対する係り受け構造を抽出し、そこに規則ベースの手法で意味役割を付与する手法を提案している。実験では京都コーパス中の 100 例文に対して、81.71%の精度を実現したと報告している。この実験を通して下村らは、意味役割付与モデルの問題の多くは、名詞の概念の類似度計算、メタファーを含む文の処理、機能語の問題に帰着するとしている。

肥塚ら［219］は、日本語フレームネットを利用した意味役割付与について報告している。日本語フレームネットは英語の FrameNet に準拠した言語資源である。肥塚らの研究では、最大エントロピー法ならびにサポートベクターマシンを用いて、日本語フレームネットから英語の意味役割付与における項識別（argument identification）に対応する「項候補獲得モデル」と、項分類（argument classification）に対応する「対応付けモデル」の二つのモデルを学習した。

2007 年当時においては、日本語フレームネットは作成途中であり、所収されている事例の数が少なかった。そのため、日本語フレームネットから抽出した事例から新しい事例を作成する処理を行っている。

実験には毎日新聞 1992～2002 年版の記事を用いている。意味役割が付与されるべき項が分かっている文に対して精度 77%、再現率 68%、意味役割が付与されるべき項がわかっていない文に対しては、最大エントロピー法を用いた場合は精度 63%、再現率 43%、サポートベクターマシンを用いた場合は精度 65%、再現率 38%の意味役割付与を実現した。

竹内ら [206] は、自身らが作成した動詞間の項構造関係のシソーラス [205] を基に、語義および意味役割付与システムを作成した。

実験においては、規則ベースのモデルを用い、入力文に対して動詞項構造シソーラス内の例文とマッチさせることで動詞語義付与と名詞意味役割の付与を行う。なお、動詞項構造シソーラスの体系を元に人手で京都大学コーパスに語義を付与したタグ付きコーパスを用いて、条件付き確率場に基づく統計モデルによる動詞語義付与と名詞意味役割付与の実験も合わせて行っている。結果は、規則ベースにおける意味役割付与の精度が 52%、統計モデルによる精度が 54%であった。

渡邉ら [211] は、英語における意味役割研究を精査した上で、述語項構造解析においては、述語と項の間の大域的な依存関係と述語の語義と項の意味役割の間の依存関係という二種類の依存関係を考慮することが重要であるとしている。渡邉らは 4 種類の因子（述語の局所的因子、項の局所的因子、述語と項のペアワイズ因子、大域的因子）を導入した線形モデルを用いて、述語語義と項の意味役割を同時に推定する構造予測モデルを設計した。

実験は CoNLL-2009 Shared Task データを用いて行っている。日本語と英語を含む 7 ヶ国語について評価実験を行っており、F1 値について英語は 84.97%、日本語は 78.69%という結果になっている。

6.3　意味役割付与と機械学習

先に見た定式化からもわかるように、項同定も項分類もある種の分類問題である。よってその枠組みとしては機械学習が主として用いられる。本節では SRL システムにおける機械学習手法について概観するとともに、機械学習において重要となる素性について整理する。

6.3.1　意味役割付与で用いられる機械学習手法について

6.3.1.1　backoff lattice

Gildea and Jurafsky ら［40］は、確率モデルを用いてラベルの分類を試みた。彼らは h（head word）、pt（phrase type）、gov（grammatical function）、position、 voice の 5 つの素性を用いた。項を t とすると、ラベルの分布は以下の条件付き確率で表される。

$$P(r \mid h, pt, gov, position, voice, t)$$

この分布は、訓練データに出現する事例を数え上げることで求めることができる。ラベルと素性の組み合わせの出現回数を、その素性の組み合わせの全体の出現回数で割ることで求めることができる。

$$P(r \mid h, pt, gov, position, voice, t) = \frac{\#(r, h, pt, gov, position, voice, t)}{\#(h, pt, gov, position, voice, t)}$$

しかし、5 つの素性の組み合わせごとに集計を行うため、訓練データ中に出現しない素性の組み合わせも多く、データスパースネスの問題に逢着する。そのため、彼らは素性の組み合わせのいくつかのサブセットについて訓練データから分布を求め、そこから全素性の組み合わせの分布を補間するという手法を用いた。

線形補間（linear interporation）は、以下の式によって補間を行う。

$$\begin{aligned} P(r \mid constituent) = {} & \lambda_1 P(r \mid t) + \lambda_2 P(r \mid pt, t) \\ & + \lambda_3 P(r \mid pt, gov, t) + \lambda_4 P(r \mid pt, position, voice) \\ & + \lambda_5 P(r \mid pt, position, voice, t) + \lambda_6 P(r \mid h) \\ & + \lambda_7 P(r \mid h, t) + \lambda_8 P(r \mid h, pt, t) \end{aligned}$$

幾何平均（geometric mean）を用いる場合は、以下の式によって補間を行う。

$$P(r \mid constituent) = \frac{1}{Z} exp\{ \lambda_1 \log P(r \mid t) + \lambda_2 \log P(r \mid pt, t) + \lambda_3 \log P(r \mid pt, gov, t) + \lambda_4 \log P(r \mid pt, position, voice) + \lambda_5 \log P(r \mid pt, position, voice, t) + \lambda_6 \log P(r \mid h) + \lambda_7 \log P(r \mid h, t) + \lambda_8 \log P(r \mid h, pt, t)$$

ここで Z は $\sum_r P(r \mid constituent) = 1$ とするための正規化定数（normalizing constant）である。

backoff と呼ばれる手法は、訓練データから調査した各素性の組み合わせごとの分布を元に、素性の多い分布を上方に、素性の少ない分布を下方に配置した lattice を構成する。素性の多い分布について該当するデータが訓練データ中にない場合は、その下方の素性の少ない分布の値から線形補間あるいは幾何平均によって補間を行う。
表 5.1 は Gildea and Jurafsky(2002)が FrameNet を元に作成したテストデータに対して分布を求めたものである。Coverage はその素性の組み合わせのデータがテストデータ中に出現した割合を表している。Accuracy は付与されたラベルが正しかった割合を示している。Performance は Coverage と Accuracy の積である。これは一般的な指標である precision に近い考え方であり、総合的な正解率を表している。
図 6.6 は、表 6.24 の分布を元に構成された backoff lattice の例である。

表 6.24: FrameNet を元に作成したテストデータに対して分布を求めた結果 [40]

Distribution	Coverage	Accuracy	Performance
$P(r \mid t)$	100.0%	40.9%	40.9%
$P(r \mid pt, t)$	92.5%	60.1%	55.6%
$P(r \mid pt, gov, t)$	92.0%	66.6%	61.3%
$P(r \mid pt, position, voice)$	98.8%	57.1%	56.4%
$P(r \mid pt, position, voice, t)$	90.8%	70.1%	63.7%
$P(r \mid h)$	80.3%	73.6%	59.1%
$P(r \mid h, t)$	56.0%	86.6%	48.5%
$P(r \mid h, pt, t)$	50.1%	87.4%	43.8%

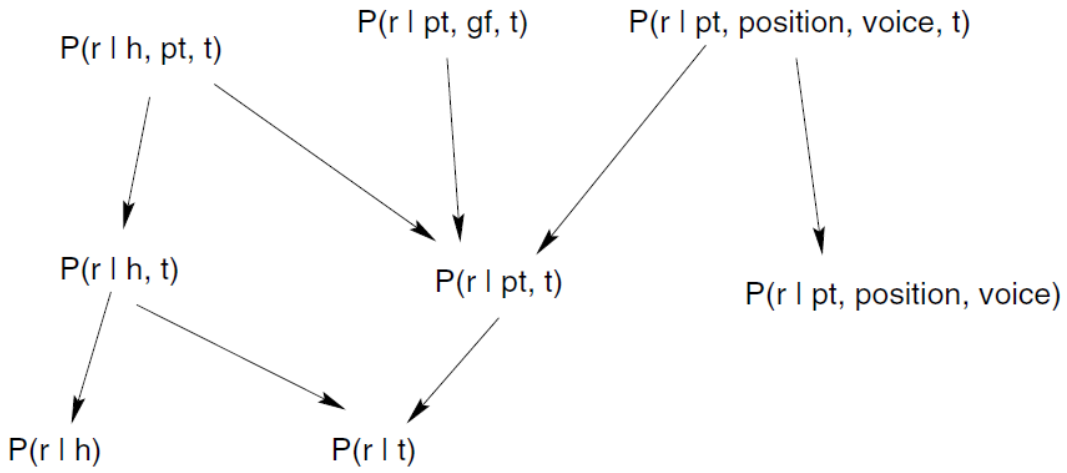


図 6.6: 表 6.24 の分布から作成された backoff lattice [40]

Gildea and Jurafsky(2002)はこれらを含めた 6 種類の補間方法を用いて、8,167 件のデータに対して試行を行った結果、表 6.25 の結果を得ている。

表 6.25: 補間方法の違いによる結果の差 [40]

Combining Method	Correct
Equal linear interpolation	79.5%
EM linear interpolation	79.3%
Geometric mean	79.6%
Backoff, linear interpolation	80.4%
Backoff, geometric mean	79.6%
Baseline: Most common role	40.9%

6.3.1.2 決定木

Gildea and Jurafsky（2002）が用いた backoff lattice には、素性を増やすと構造が複雑化するために容易に素性を増やせないという問題がある。この問題を指摘した Surdeanu ら [129] は、項の分類に決定木を用いる手法を提案した。決定木の構成には

C5.0 を用いる［112］。これは ID3、C4.5 から派生した決定木の構成手法であり、各ノードの条件としてエントロピーの期待値が最大になるものを選ぶというものである。

Surdeanu らは、Gildea and Jurafsky（2002）が用いた素性に加えて表 6.26 に記載された素性を用いた。実験に際しては PropBank を用いた。Gildea and Jurafsky（2002）の手法が精度において 82.8%だったのに対し、Surdeanu らの手法は 83.74%の精度であった。

表 6.26: Surdeanu らが追加で用いた素性［129］

CONTENT WORD(cw)	Lexicalized feature that selects an informative word from the constituent, different from the head word.
PART OF SPEECH OF HEAD WORD (hPos)	The part of speech tag of the head word.
PART OF SPEECH OF CONTENT WORD (cPos)	The part of speech tag of the content word.
NAMED ENTITY CLASS OF CONTENT WORD (cNE)	The class of the named entity that includes the content word.
BOOLEAN NAMED ENTITY FLAGS	A feature set comprising: - neOrganization: set to 1 if an organization is recognized in the phrase - neLocation: set to 1 a location is recognized in the phrase - nePerson: set to 1 if a person name is recognized in the phrase - neMoney: set to 1 if a currency expression is recognized in the phrase - nePercent: set to 1 if a percentage expression is recognized in the phrase - neTime: set to 1 if a time of day expression is recognized in the

	phrase - neDate: set to 1 if a date temporal expression is recognized in the phrase
PHRASAL VERB COLOCATIONS	Comprises two features: - pvcSum: the frequency with which a verb is immediately followed by - pvcMax: the frequency with which a verb is followed by its any preposition or particle. Predominant preposition or particle.

Chen ら [20] は、直接には決定木を用いていないが、自身の提案する手法の正当性を主張するために、既存の素性を用いて C4.5 決定木を用いた分類器による結果を Gold-Standard として用いた。

6.3.1.3 最大エントロピー法

最大エントロピー法（maximum entropy method）とは、「全ての特徴は与えられたデータの中に含まれている」という前提のもとデータからは判らない箇所の分布については一様分布（すなわちエントロピーが最大の状態）であると仮定する最大エントロピー原理（principle of maximum entropy）に基づいてデータから確率分布を求める最尤法のひとつである。Jaynes [52] によって提唱され、Berger（1996）によって自然言語処理の分野に紹介された [4]。

backoff latice や決定木のような確率に基づくモデルは、データを素性の組み合わせに分割するためにデータスパースネスの問題に晒されやすく、多くの素性を扱うのが難しい。これらのモデルが持つこの本質的な問題を、Fleischman ら [34] は、最大エントロピー法による分類器を用いることで克服した。backoff lattice を用いた手法に比べて 6.2%の性能の向上を見せた。

Toutanova ら [135] は、Pradhan(2005)が用いたのと同じ素性を用いて最大エントロピー法による実験を行った結果、SVM を用いた結果と大きな差はなかったことを報告している。

Jiang ら [53] は、NomBank ベースの意味役割付与システムの性能向上のために最大エントロピー法を利用した。

Gildea and Hockenmaier ら［38］は、Gildea and Palmer ら［39］が用いた素性をベースに、CCG 文法から抽出した素性を追加したシステムを提示した。

6.3.1.4　サポートベクターマシン

サポートベクターマシン（SVM：Support Vector Machine）は Vapnik ら［144］によって提唱された手法である。超平面で入力データ空間を分割する際にそれぞれのクラスのマージンが最大になるような超平面を選択することで、識別精度の向上を目指した手法である。発表当初は線形分類しかできなかったが、カーネル法と組み合わせると非線形分類問題を特徴空間上の写像の線形分類問題として対応可能であることが Boser ら［8］によって示されたことで、近年注目を集めるようになった。

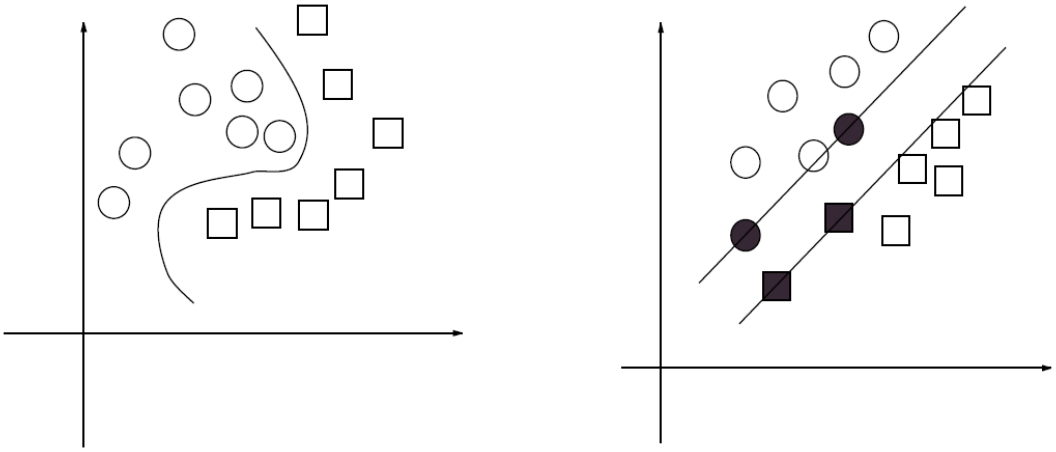


図 6.7: 入力データ空間（左）から特徴空間（右）へ写像を線形分離する［186］

Pradhan ら［101］は、SVM を用いた場合、backoff latice と比較し、項分類のタスクにおいて決定木は 2%、SVM は 10%の向上を見せたと報告している。Gildea and Palmer ら［39］の見解と同じように、項識別のタスクにおいてはパスが、項分類のタスクにおいては主辞と述語が最も有効な素性であったと報告している。

一般的に SVM や最大エントロピー法のような分類手法（discriminative apploache）に基づく手法のほうが、backoff lattice や決定木のような確率に基づくモデルよりも多くの素性を扱うことができる。これは backoff lattice や決定木がデータを素性の組み合わせに分割することにより、データスパースネスに遭遇しやすくなるためである。

6.3.1.5　部分教師付き学習

最大エントロピー法や SVM のような教師付き学習をベースにした手法においては、学習データには正解となる意味役割が付与されている必要がある。そのため、意味役割が付与されたデータを十分な量集められないこともある。この問題を解消するために、Furstenau ら［35］は、部分教師付き学習を用いた。これは入力されたラベルなしの文章に対して、文法的意味的な素性が最も近いラベルあり文章を選択し、後者のラベルを前者に反映させることで実現する。

Gordon and Swanson ら［42］は、入力されたラベルなしの文章に対して、それが持つ動詞と文法的に近い動詞の項が持つラベルを転写する。項のタイプが似ている動詞同士は似ていると判断する。

6.3.1.6　教師なし学習

教師なし学習に基づくアプローチは、Swier ら［132］や Grenager ら［43］によって提示されている。前者は VerbNet に収録されている動詞の情報を利用し、後者は文法的位置と意味役割の結び付きを探すために EM アルゴリズムを用いている。

6.3.1.7　整数線形計画法

Punyakanok [107] らは、文レベルでのラベルの割り当ての問題を、整数線形計画法（integer linear programming）の問題として捉えたアプローチを行っている。

6.3.2　意味役割付与で用いられる素性について

機械学習において分類対象をどのような素性で表現するかということは、どのような機械学習のアルゴリズムを使用するかということよりもむしろ重要なことである。

項同定と項分類では有効となる素性が異なることが知られている。一般的に項同定においては文法的な素性が有効で、項分類においては意味的な素性が有効となることが知られている［100]。

局所的素性のみを用いる分類器で解の候補を絞り込んだあとに大域的な素性も用いる分類器で最終的な解を決定する手法をリランキングという。意味役割付与においても、局所スコアリング（local scoring）と結合スコアリング（joint scoring）を組合せて用いることが多い。局所的スコアリングとは、他の項に与えられたラベルを考慮せずに行うスコアリングのことで、これのみでは以下のような矛盾を孕み得る［156]。

（1）項のバッティング（overlapping argument strings）

By [A1 working [A1 hard] , he] said , you can achieve a lot.

（2）同一の項の複数出現（repeated arguments）

By [A1 working] hard , [A1 he] said , you can achieve a lot.

（3）項の喪失（Missing argument）

[A0 By working hard , he] said , [A0 you can achieve a lot].

そのため、他の項のラベルとの間の依存関係を考慮した結合スコアリングでリランキングを行う。これらの処理はヒューリスティックによって候補を絞った後に行われる。

意味役割付与において用いられる素性には様々なものがある。統語パーサーのアウトプットから得られる文全体の統語的性質を意味役割付与の向上に利用するための研究は多く行われている。主に以下のような性質が素性として用いられる。

6.3.2.1　句の型（phrase type）

意味役割と句には相関関係があるという観察に基づいた素性。FrameNet においては意味役割「Speaker」は名詞句として、「Topic」は前置詞句あるいは名詞句として、「Medium」は前置詞句として現れる傾向がある。

6.3.2.2　統率範疇（governing category）

統率範疇（governing category）とは、チョムスキーによって 1980 年代に導入された文法理論である統率・束縛理論（GB 理論）と呼ばれる言語学の枠組みの中で用いられる概念である。ある語αにつき、そのαを含む最小の句を統率範疇という。
統率範疇が S になる名詞句は Subject であることが多く、統率範疇が VP になる名詞句は Object になることが多いという観察に基づいた素性である。

6.3.2.3　統語解析木上の経路（parth tree path）

述語から要素までの統語解析木上の経路。例えば図〇における ate から He への経路は「VB↑VP↑S↓NP」のように表現される。経路とその終着点となるノードに付与される意味役割には相関があり、例えば VB↑VP↑S↓NP という経路によって辿り着く要素は意味役割 Subject を、VB↑VP↓NP という経路で辿り着く要素は意味役割 Object を持つ傾向があることが観察されている。

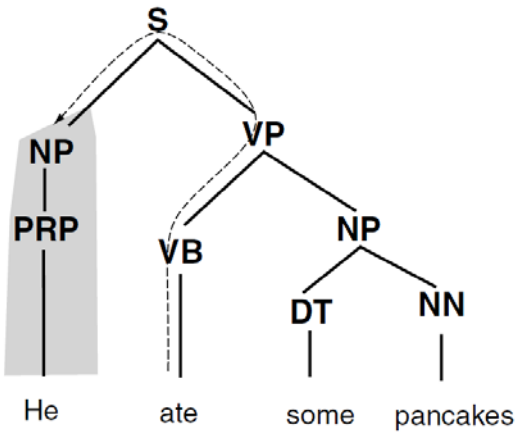


図 6.8: parth tree path [40]

6.3.2.4　候補と述語の間の位置関係（position）

候補の出現位置が述語の前か後かという情報を素性として用いる。これは、多くの場合主語は動詞の前に、目的語は動詞の後に出現するという観察に基づいている。

6.3.2.5　動詞の態（voice）

動詞が能動態か受動態かという情報を素性として用いる。能動態動詞の直接目的語はしばしば受動態目的語の目的語と意味役割の点で一致する。

6.3.2.6　主辞（head word）

名詞句の主辞は意味役割をある程度限定するのに役立つ。例えば、Bill や brother や he といった語を主辞に持つ名詞句は Speaker の意味役割を持つ確率が高く、proposal や story や question といった語を主辞に持つ名詞句は Topic の意味役割を持つ確率が高い。

6.3.2.7　下位範疇（subcategorization）

その文においてその動詞が取り得る項の種類を素性として用いる。例えば同一の動詞 close であっても「The door closed」のような自動詞的な用いられ方と「He closed the door」のような他動詞的な用いられ方では意味役割付与のパターンが異なるはずである。前者の場合の close の下位範疇素性は{subject} であるのに対し後者のそれは {subject, object}であり、区別することが可能である。

6.3.2.8　argument set

文中に出現する全ての動詞の全ての項に付与された意味役割の集合。

6.3.2.9　文における項の出現順序（argument order）

与えられた動詞の項の文の要素の位置（position of a constituent in the sequence of arugument）を指し示す整数。この素性は統語解析木を用いないため、SRL システムをパースエラーに対して頑健にすることができる [34]。

6.3.2.10　前の項に与えられた意味役割（privious role）

前の項に与えられた意味役割。異なる要素間に与えられた意味役割の依存性を利用する素性 [34]。

6.3.2.11　主辞の品詞（head word part of speech）

Penn Treebank の品詞分類においては単数系と複数形、普通名詞と固有名詞の区別がなされているため、この品詞分類によって文が解析されていれば、名詞句の主辞の品詞がその名詞句のに付与される意味役割の特定に役立つ [129]。

6.3.2.12　候補の固有表現（named entity）

人名や地名などの固有表現は無数に存在するため、データスパースネスの問題が生じやすい。入力文に対して固有表現抽出を行い、それらを「人名」「地名」といったクラスに分類することでこの問題を回避する [101]。

6.3.2.13　動詞のクラスタリング

動詞を意味的な近さでクラスタリングした結果を素性として用いる。これは意味的に近い動詞は項構造も近いという観察に基づいている。意味的に近い動詞は直接目的語が同じである場合が多いため、このことを利用してクラスタリングを行う。

6.3.2.14　前置詞句の主辞

項が前置詞句であった場合、その主辞は前置詞になる。in, across, location といった前置詞を主辞に持つ前置詞句は「場所」を示す意味役割を持つことが多い。このように

前置詞によってある程度意味役割が特定できるため、これを素性とする。例えば in のような前置詞は多様な使われ方をするが、in February は時間を、in New York は場所を表すといったように前置詞句の目的語を考慮に入れることによって意味役割を特定することができる。

6.3.2.15 First/Last word/POS in Constituent

要素の中の最初あるいは最後の語、または形態素。

6.3.2.16 文における候補の出現順序

上で紹介した argument order 素性に関連しているが、項を項でないものから識別することを目的に作られた素性。それぞれの要素につき述語との相関的な位置が計算される。項である要素はそうでない要素に比べて述語までの距離が近いという観察に基づいている［101］。

6.3.2.17 要素の文脈情報

ある要素につき、親および左右の兄弟要素それぞれの句のタイプ、主辞、主辞の品詞を素性として用いる。つまり要素の置かれた文脈情報を利用する。

6.3.2.18 時間を表す固有表現

要素があらかじめ想定している時間的手掛かり（temporal cue）を含むか否かを示す二項素性。ここで時間的手掛かりとは year、month、yesterday などの語を指す［130］。

6.3.2.19 頻出動詞

訓練データ中に複数回出現した述語かどうかを示す二項素性［130]。

なお、複数の素性を組み合わせて用いることで高い解析結果を得ることができることが知られている。Xue ら［150］は、以下の素性が有効であることを報告している。

(1) 述語と句のタイプの組み合わせ

句のタイプはそれ単体として用いた場合は小さな効果しかもたらさないが、述語

と組み合わせることで大きな効果を生む素性となる。述語それ自体は、要素が項であるか否かを判定する際の素性として対して有効ではないが、ある述語が与えられた時、ある特定のタイプの素性が項になる可能性が高まる。例えば NP 単体や give 単体は項を判定する効果を対して持たないが、give という述語が与えられた状況においては NP は項である可能性が高いと判断できる。

(2) 述語と主辞の組み合わせ
(3) 態と位置の組み合わせ
(4) 述語とその述語と要素間の距離の組み合わせ

カーネル法を用いたサポートベクターマシンなどは素性の組み合わせを取り入れることが容易いが、計算量が大きくなるという問題がある。Xue ら [150] はサポートベクターマシンより計算量が少ない最大エントロピー法を用いて、サポートベクターマシンに匹敵する解析精度を実現している。

6.4 意味役割についての議論

表層格は文の表層に現れる特徴に基づいて分類されるものであるため、その分類は有限かつ小数であり、なおかつ判断の個人差がない。それに対し深層格（あるいは意味役割）は概念のカテゴリー化であり、その分類の歴史は情報科学の範囲に留まらず、長い哲学の歴史の中にアリストテレスやカントによって行われた概念のカテゴリー化を見出すことができる。かように、概念の体系化は際限のない課題である。

科学技術庁の機械翻訳プロジェクト（Mu プロジェクト）において、名詞に対する意味マーカーの付与作業が行われた。分類体系には帰納的方法（ボトムアップ方式）と演繹的方法（トップダウン方式）があり、このうち演繹的方法による分類体系では無限に近い単語の意味分類をするには無理があるという報告がされている [200]。

辻井 [209] は、格には世界についての客観的な知識記述や人間の長期的な記憶の理論に使われる「格」と、発話者というフィルターを経た言語表現を議論する「格」が異なることを指摘している。その上で、どちらの場合における議論についても、格という意味的役割が解釈されていくプロセスはほとんど問題にされず、個々の格関係についても静的な分類基準だけを問題にしていることや、それぞれの格の範疇に絶対的な定義が与えられることを前提としていることを批判している。

意味役割の種類は項の意味上の種類であり、研究者によってその分類基準は定まっていない。Fillmore [161] や Grimshaw [44] は言語学的見地から数個程度の深層格を仮

定したが、VerbNet や PropBank は付与と処理の観点から 20 数個程度が提案されており、FrameNet においては 1,000 種類を超える意味役割が提案されている。

意味役割の付与には見方の問題が影響を及ぼす。例えば「売る」「買う」という動詞について、「太郎が古着を次郎から買う」と「次郎が古着を太郎に売る」は起こった事象は同じであるが、太郎と次郎のどちらを主体として見るかで意味役割の付与は異なる。

竹内ら [207] はこの問題を受けて、彼らが作成している動詞項構造シソーラスにおいては、実世界に対するマップを意味役割で行うのではなく、実世界で起こった実体を推論するための基本データとして意味役割を付与するとしている。動詞項構造シソーラスは現在 71 種類の意味役割を持っている。

6.4.1 フレームが持つ問題と意味役割の汎化

FrameNet や PropBank といった意味役割付与コーパスは、文中の動詞がフレームと呼ばれる特定の項構造を持つという考えに基づいて構成されている。意味役割はそれぞれのフレームに固有の役割として定義される。例えば PropBank においては sell と buy という二つの動詞は、それぞれそのフレーム内に「Seller」という意味役割を持つが、これは同じ字面であっても sell のフレームにおける「Seller」と buy のフレームにおけるそれは意味役割としては別のものとされる。

この定義により、コーパス中に事例の少ない役割が大量に存在する状況が出現し、学習時のデータスパースネスを引き起こす。SRL を行う上でのこの問題を解決するためには、類似する意味役割を何らかの指標で汎化し、共通点のある役割の事例を共有する手法が必要となる。

Moschitti ら [89] は、文中から発見された項を AX（Core Arguments）、AM（Adjuncts）、CX（Continuation Arguments）、RX（Co-referring Arguments）の 4 つのクラスに分類している。Baldewein らは、もしフレーム A の要素 A1 がフレーム B の要素 B1 と似ていたら A1 の訓練データを B1 の訓練データとして再利用する、というアプローチを採用している。相似の指標としては (1) FrameNet におけるフレームの階層関係(Frame Hierarchy)、(2) 周辺的なフレーム要素 (Perpheral frame elements)、(3) EM アルゴリズムによるクラスタ（EM-based clustering）が用いられている [2]。

松林ら [197] は、FrameNet と PropBank を対象に意味役割の汎化実験を行っている。PropBank には 4,659 個のフレーム、11,500 個以上の意味役割が存在し、フレームあたりの事例数は平均 12 個である。FrameNet では、795 個のフレーム、7,124 個の異なった意味役割が存在し、役割の約半数が 10 個以下の事例しか持たない。事例の

少なさはデータスパースネスの問題を引き起こし得るため、何らかの形で意味役割を汎化することで、事例の少なさをカバーする必要がある。

意味役割の汎化に関しては、Yi ら［155］、Loper ら［74］、Zapirain ら［157］などから様々な指標が提示されてきたが、意味役割の汎化は単一の指標に基づくのではなく複数の指標を用いるべきであると松林らは主張している。従来研究における意味役割の汎化は、コーパス中の意味役割ラベルを一対一対応の取れる汎化ラベルに置き換えることで実現されてきた。しかしこの手法においては、単一の指標に基づいてしか意味役割を汎化できない。そのため、松林らは、意味役割ラベルに対して複数の汎化ラベルを含む集合を対応させることで、意味役割がn種類の汎化指標によるラベルを同時に持つことを表現できるようにした。このような汎化によって FrameNet を利用した場合も PropBank を用いた場合も、意味役割付与の精度が向上することが報告されている。

6.5 まとめ

文章から機械的に意味を判定する研究は自然言語処理の歴史とほぼ重なると言ってもよいほど古くから行われているが、意味役割付与というアプローチによる研究は高々10 年程度しか行われていない。

手動によって作成したルールをベースとして深層格を判定する研究が中心だった中で、2002 年に Gildea and Jurafsky ら［40］が FrameNet を用いて行った研究が、意味役割付与研究の起点となった。

さらに 2005 年、意味役割を付与した大規模なコーパスである PropBank の完成が、英語を対象とした意味役割付与研究を加速させた。

PropBank による大量の正解付きデータを利用した教師あり学習を適用した研究が盛んに報告された。主な手法としてはサポートベクターマシン［100, 101, 137］、最大エントロピー法［46, 72, 150, 154］、条件付き確率場［24］などが用いられている。

完全統語パーサを用いるか［150］、浅い統語パーサを用いるか［45, 102］ということも広範に議論された問題である。意味役割付与システムは統語パーサの出力をインプットとするため、統語パーサによる構文解析結果が誤っていた場合は深刻な性能の低下を招く。Yi ら［154］は、最大エントロピー法をベースとしたパーサー（Maximum Entropy-based parser）を使うことでこの問題に対応した。Koomen ら［64］、Pradhan ら［103］、Márquez ら［91］、Tsai ら［137］は、複数のパーサの出力を用いることで対応した 。

2005 年の CoNLL の shared task において最高の性能を見せたのは Punyakanok ら [109] のシステムで、ウォールストリートジャーナルのテストデータに対して F1 値で 79.4%という数値を出した。

2008 年の CoNLL の Shared Task においては、Zhao らの 79.40%、Vickrey の 79.45%、Johansson の 86.37%（このへんのデータは、Jointly Identifying Predicates Arguments and Senses using Markov logic より）などのシステムがあったが、述語の意味を考慮することで意味役割付与の性能を改善することができるという Surdeanu らの報告が意義深かった [131]。文中の語に対して包括的に意味を付与したコーパスである OntoNotes が登場したことにより、述語のみならず周辺の語の意味も考慮した上での意味役割付与が可能になり、Che ら [19] はウォールストリートジャーナルのテストデータに対して F1 値で 89.25%の数値を出した。

実験では 90%近い性能を達成するまでに進化した意味役割付与システムであるが、その性能の向上と共に次第に過学習の問題が指摘され始めた。Johansson ら [54] は、複数のシステムにおいて訓練用のコーパスとは異なるコーパスをテスト用に用いた場合に、項分類タスクにおいて 20%近くの精度の低下が見られたことを報告している。

これらの報告を受け、学習に使用したデータセットとは異なる領域の文章に対して解析を行っても精度の低下しないシステムが研究された。HMM（Hidden Markov Model）を利用した Huang ら [51] のシステムは、学習用データにウォールストリートジャーナルのデータを用い、テスト用データに Brown コーパスを用いたが、性能の低下は 5.4%にとどまっている。

英語以外の言語に対する意味役割付与の研究も行われている。その中でも中国語に対する研究が盛んであり、中国語版の PropBank である Chinese PropBank [149] も作成されている。Xue [151] は、同コーパスを用いた意味役割付与は、統語解析を人手で行った場合は最新の英語の意味役割システムの結果と拮抗するが、中国語統語パーサの精度が低いために、それを用いた場合の意味役割の性能は英語のそれと比べて著しく低下したと報告している。

日本語における意味役割付与の研究は、PropBank のような意味役割が付与された大規模なコーパスが不足していることもあり、機械学習の手法を用いたものよりも規則ベースのものが多い。

動詞とその動詞に共起する名詞と表層格のパターンの対に着目し、入力文を手動作成した深層格フレームにマッチングさせることでニュース文の意味役割を付与する手法 [224]、EDR 辞書を用いて動詞の必須格パターンごとに動詞を分類し、入力文から抽

出した格助詞のパターンをそれにマッチングさせることで EDR 日本語共起辞書の概念関係子を付与する手法［203］、EDR 日本語共起辞書から名詞と表層格のパターンを抽出してデータベースに蓄積し、出現頻度を基に意味役割を推測する手法［192]、認知科学の報告を参考に動詞の選好をモデル化する手法［187］、係り受け構造を基に規則ベースで意味役割を付与する手法［165］などがある。

機械学習の手法を用いた研究としては、日本語フレームネットを基に最大エントロピー法とサポートベクターマシンを用いたもの［219］、動詞間の項構造関係シソーラスを基に条件付き確率場を利用したもの［206］、述語の項の間の大域的な依存関係のモデルを Passive-Aggressive アルゴリズムで学習する手法［211］がある。

NAIST テキストコーパスに対して意味役割のアノテーションを行うプロジェクトが進められており［223］、PropBank のような意味役割が付与された大規模なコーパスが整備されるとともに、日本語における意味役割付与研究も増えていくものと思われる。

第 7 章　事態性名詞の解析

7.1　事態性名詞

動詞や形容詞などは事態を表現するが故に項構造を考慮するが、名詞の一部にも事態を表現するものがある。例えば「彼は上司の推薦で抜擢された」という文において、名詞「推薦」は「上司が彼を推薦する」という事態を表す。飯田ら [217] はこのような名詞を事態性名詞と呼んだ。自然言語処理の評価型ワークショップ CoNLL 2008, 2009 では、PropBank と NomBank を用いた述語と事態性名詞に対する項構造解析の共通タスクが行われるなど、近年活発に研究されている分野である。

7.2　事態性判別と項同定

事態性とは、文中で名詞が事態を表すかどうかということであり、同じ名詞であっても文脈によって事態を表すか単にモノを表すかが変化する場合がある。例えば「レポート」などがそれにあたり、単にレポートという結果物を表すのかレポートするという行為を表すのかは文脈に依存する。すなわち事態性名詞の述語項構造解析においては、単に項構造を特定するだけではなく、対象となる名詞が事態性を持つか否かということをまず判定する必要がある。小町ら [193] は、前者のタスクを「事態性判別」、後者を「項同定」と呼んでいる。

小町ら [195] は事態性判別と項同定を別タスクとして扱っており、事態性判別のみを先に解く手法を提示している。これは、事態性判別は語義曖昧性解消の問題であり、項同定とは別の素性を用いる解析が有効だと仮定しているためである。

事態性名詞の項構造解析を述語項構造解析と比較すると、以下の 3 つの問題がある。

(1) 事態性名詞は文脈によって事態を指す場合とそうでない場合がある。
(2) 項を文節単位ではなく形態素単位で考える必要がある。
(3) 格助詞が項同定の手掛かりとならない問題。

(1) は、たとえば「公衆電話で電話をすることがめっきり減った」という文においては、電話は物体としての電話を表しており事態性を持たないが、電話は「電話をする」という行為であり「(話し手) ガ (誰か) ニ電話をする」という事態を表している。(2) は「民間支援が活性化する」という文において、述語「活性化する」の項は「民間支援

が」という文節単位で考えればよいが、事態性名詞「支援」の項は「民間」であり、これは形態素単位で考慮する必要がある。(3)は事態性名詞が形態素単位で項の候補を考慮する必要があることから必然的に浮上する問題である［194］。

事態性判別は、文中に現れる事態性名詞を事態性あり／なしの 2 クラスに分類する問題であり、小町らはこれの教師なし学習を、文の構造を木構造に変換して素性として扱うブースティングアルゴリズムである BACT [182] を用いて学習している。

NAIST テキストコーパスにおいて、述語についてはヲ格の 84%と 88%が係り受けの関係にある文節に項を持つが、事態性名詞においてはそれぞれヲ格の 31%、ニ格の 22%しか係り受け関係の文節に項がない。さらに、述語は通常格助詞を伴って出現するため、格助詞が大きな手掛かりとなるが、事態性名詞の項は格助詞を伴わないためこれを手掛かりとすることができない。

このような問題に対して、観察から導かれる以下の仮定を利用して対応する。

(1) 事態性名詞が事態性を持つときは動詞としての用法を考えたときの項構造を基本的に受け継ぐ。
(2) サ変名詞とそれに対応する動詞が意味的には共通の項を持つ。
(3) 事態性名詞のある程度のものについては支援動詞構文で用いられている。

(1)(2)の仮定を元に、動詞と格要素の共起情報を事態性名詞の項同定に用いる。(3)については、事態性名詞と述語の間で項の対応がついた辞書を作成し、支援動詞構文の認識に用いている。

小町ら［194］は、文中の名詞が事態性名詞であるかを判断するために動詞と格要素の共起モデルが利用できるかどうかを実験にて確認した。動詞と格要素の共起モデルには藤田らのモデル［213］を利用した。藤田らは、名詞nが格助詞cを介して動詞vに係っているときの共起確率$P(\langle v,c,n\rangle)$の推定に確率的潜在意味インデックス（PLSI：Probabilistic Latent Semantic Indexing）［50］を用い、$\langle v,c,n\rangle$を$\langle v,c\rangle$とnの共起と見なしてモデルを作成している。

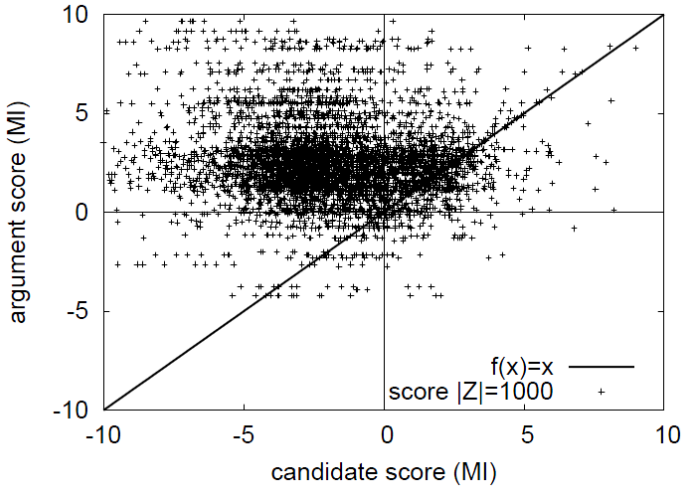


図 6.9: 項であるか否かと共起スコアの相関をプロットした図 [194]

実験に際しては、NAIST テキストコーパスから新聞記事 1 日分 137 記事（1,226 文）中、文内にガ格の項がある事態性名詞を対象に、項の名詞と文内にある他の名詞を比較し、項であるか否かと共起スコアの相関を図 6.9 にプロットしている。第二象限は他の名詞の方が共起スコアが高く項のスコアが低い事例、第四象限は項の共起スコアが高く他の名詞のスコアが低い事例である。第二象限と第四象限に含まれる項と他の名詞のペアの総数は全体の 71.2%（9,715 事例）あり、共起スコアの大きい方が正解とした場合の精度は 90.0%であった。第一象限の事例は 28.1%（3839 事例）あったが、精度は 55.8%であった。第二、四象限に含まれる事例については共起スコアが有効であるが、第一、三象限に含まれる事例については共起以外の情報を用いて判別する必要がある。

笹野ら [185] は、名詞格フレーム辞書を利用した事態性名詞の解析を提案し、未だ存在しない名詞格フレーム辞書をコーパスから自動構築することを考えた。名詞と項がノ格で接続された「A の B」の形の用例をコーパスから収集する形で格フレーム辞書を構築している。実験に際しては、毎日新聞 12 年分および日経新聞 13 年分の約 2,500 万文を用いて名詞格フレーム辞書の自動構築を行い、約 17,000 語の名詞について格フレームが構築された。適合率は 82.9%、再現率は 85.3%であった。また、自動構築した名詞格フレーム辞書を用いて関係解析を行った結果は、適合率 51.7%、再現率 67.4%の精度であった。文脈における格スロットの必須性が異なることや、コーパスに適切な用例が不足していたことなどが解析誤りの主な原因であったとしている。

NomBank においては事態性名詞（nominal predicate）に対するラベル付けは行わ

れているが、それが述語として項との間にどのような関係を持つかについては情報を持っていない。Gerber ら [37] は、NomBank に含まれる事態性名詞について述語項構造解析を行い、それらのうち 65%に対してその項（implicit argument）を判別することができたと報告している。

表 6.27 は、項の候補 c が事態性名詞pの項$iarg_n$を満たすか否かを判断するために用いた素性である。$iarg_n$は項の種類を表しており、$iarg_0$が動作の主体、$iarg_1$が動作の対象物、$iarg_2$が動作の対象先を表す。例えば investment（投資）という事態性名詞については、投資家が$iarg_0$、投資する金銭が$iarg_1$、投資される会社が$iarg_2$となる。
分類の対象となるのは、事態性名詞pと項$iarg_n$と候補cを含む照応連鎖（coreference chain）c'の三つ組$\langle p, iarg_n, c' \rangle$である。
実験に際しては、816 の事態性名詞を含むデータを用いて、素性に基づくロジスティック回帰モデル（feature-based logistic regression model）の学習を行った。テスト用データとしては 437 の事態性名詞を含むデータを使用している。比較用のベースラインとして、二つの文において事態性名詞に最も近い候補を$iarg_n$とするというヒューリスティックを用いている。実験結果は表 6.28 の通りである。ベースラインの手法と比較して F1 値で 15.8 ポイントの性能向上が達成されている。

表 6.27: 項の候補cが事態性名詞pの項$iarg_n$を満たすか否かを判断するために用いた素性 [37]

#	Feature value description
1	For every f, the VerbNet class/role of p_f/arg_f concatenated with the class/role of $p/iarg_n$.
2	Average pointwise mutual information between $\langle p, iarg_n \rangle$ and any $\langle p_f, arg_f \rangle$.
3	Percentage of all f that are definite noun phrases.
4	Minimum absolute sentence distance from any f to p.
5	Minimum pointwise mutual information between $\langle p, iarg_n \rangle$ and any $\langle p_f, arg_f \rangle$.
6	Frequency of the nominal form of p within the document that contains it.
7	Nominal form of p concatenated with $iarg_n$.
8	Nominal form of p concatenated with the sorted integer argument indexes from all arg_n of p.

9	Number of mentions in c'.
10	Head word of p's right sibling node.
11	For every f, the synset [30] for the head of f concatenated with p and $iarg_n$.
12	Part of speech of the head of p's parent node.
13	Average absolute sentence distance from any f to p.
14	Discourse relation whose two discourse units cover c (the primary filler) and p.
15	Number of left siblings of p.
16	Whether p is the head of its parent node.
17	Number of right siblings of p.

表 6.28: Gerber らによる実験結果 [37]

			Baseline			Discriminative				Oracle	
	#	Imp. #	P	R	F1	P	R	F1	p	R	F1
sale	64	60	50.0	28.3	36.2	47.2	41.7	**44.2**	0.118	80.0	88.9
price	121	53	24.0	11.3	15.4	36.0	34.6	**34.2**	0.008	88.7	94.0
investor	78	35	33.3	5.7	9.8	36.8	40.0	**38.4**	< 0.001	91.4	95.5
bid	19	26	100.0	19.2	**32.3**	23.8	19.2	21.3	0.280	57.7	73.2
plan	25	20	83.3	25.0	38.5	78.6	55.0	**64.7**	0.060	82.7	89.4
cost	25	17	66.7	23.5	34.8	61.1	64.7	**62.9**	0.024	94.1	97.0
loss	30	12	71.4	41.7	52.6	83.3	83.3	**83.3**	0.020	100.0	100.0
loan	11	9	50.0	11.1	18.2	42.9	33.3	**37.5**	0.277	88.9	94.1
investment	21	8	0.0	0.0	0.0	40.0	25.0	**30.8**	0.182	87.5	93.3
fund	43	6	0.0	0.0	0.0	14.3	16.7	**15.4**	0.576	50.0	66.7
Overall	437	246	48.4	18.3	26.5	44.5	40.4	**42.3**	< 0.001	83.1	90.7

第 8 章　おわりに

述語項構造解析の中心は、表層格付与から深層格付与にシフトしている。

現在まで意味役割付与の主流となっている手法は、PropBank などの意味役割タグが付与されたコーパス利用した教師あり学習を用いる手法である。そのため、意味役割付与研究における研究報告は、性能向上に寄与する効果的な素性の提案を行うと言った feature engineering に関するものか、新しい機械学習の手法を適用したことによる性能向上の報告が多い。

しかしそれらの手法はタグ付きコーパスに依存するため、必然的に以下の問題を抱える。

１．ドメインの問題

２．タグ付きコーパスが整備されていない問題

これらの問題に対するアプローチが、意味役割付与研究の今後の課題と重なると思われる。

Johansson ら [54] によって指摘された、訓練用のコーパスとは異なるコーパスをテスト用に用いた場合に多くの意味役割付与システムが深刻な性能の低下を見せるという問題に対しては、Huang らの隠れマルコフモデルによる手法 [51] 、Samad らの半教師有り学習を用いる手法 [118] 、Lang ら [67] の教師なし学習を用いる手法などが提案されている。

タグ付きコーパスの量の不足を補う研究としては [35, 42, 43, 132] などがある。

これらの研究に共通しているのは、従来研究の主流であった教師あり学習ではなく教師なし学習や半教師なし学習といった手法を用いていることであり、これらの手法を応用することが現在における意味役割付与研究の中で大きな流れを形成している。

一方で、文章の意味を解析するうえでの意味役割付与の問題点を指摘する視点もある。

Blanco(2011)らは、文章の意味を解析することは自然言語処理が取り組んできた重要なタスクであるとし、その分野における意味役割付与の貢献を評価する [6]。しかしその一方で意味役割は意味解析における特定の関係性に過ぎず、文章の包括的な意味を分析するにはそれを抽出するだけでは不完全だとしている。図 8.1 において、一般的な意味役割付与システムは、実線で示された統語的依存関係に対応した関係だけに着目するが、破線で示された関係も文の意味を捉える上で重要であるとしている。

Blanco(2011)らは semantic primitive という関係性を文章から抽出する手法を提案している。

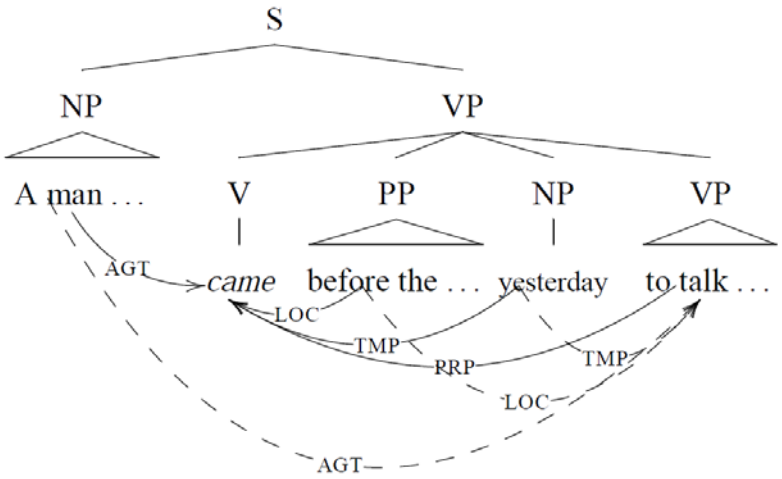


図 8.1: 語の意味的関連性（semantic relation）[6]

Blanco（2011）らの指摘は、文章の意味を解析することは入力文の項に対して意味役割を付与することであると視点を固定させてしまうことに対して警鐘を鳴らすものである。

文章の意味を理解する計算機の実現。おそらく計算機が誕生した瞬間から計算機科学が頂き続けてきた夢は、機械学習的手法の導入によって少なからず実現に近づいたようにも思われる。しかし一方で、ルールベースの手法が主流だった時代に自然言語処理が持っていた言語学との間の緊張関係が、統計的手法の導入によって薄れてしまっていることを指摘する声もある。

いずれにせよ、包括的な観点からの絶え間ない模索と問い直しこそが、この夢の実現に不可欠なものであろう。

謝辞

本課題研究を進めるにあたり、多大なご支援、ご指導を頂いた島津明教授に深くお礼申し上げます。中間審査の場において深層格付与研究に関する重要性につきご指摘を頂いた東条敏教授、白井清昭准教授に厚く感謝致します。

また、社会人に貴重な大学院教育を受ける機会を提供するとともに社会人大学院生の研究活動の全面的な支援に尽力された、北陸先端科学技術大学院大学先端領域社会人教育院関係者各位に深くお礼申し上げます。

参考文献

[1] Collin F. Baker, Charles J. Fillmore and John B. Lowe. The Berkeley FrameNet Project. ACL '98 Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1, pp.86-90, 1998.

[2] Ulrike Baldewein, Katrin Erk, Sebastian Padó and Detlef Prescher. Semantic Role Labelling with Similarity-Based Generalization Using EM-based Clustering. Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, pp.64-68, 2004.

[3] Valerio Basile, Johan Bos, Kilian Evang, Noortje Venhuizen. A platform for collaborative semantic annotation. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 2012.

[4] Adam L. Berger, Vincent J. Della Pietra, Stephen A. Della Pietra. A Maximum Entropy Approach to Natural Language Processing. Computational Linguistics, Volume 22, Issue 1, pp.39-71, 1996.

[5] Don Blaheta, Eugene Charniak. Assigning function tags to parsed text. In Proceedings of the First Annual Meeting of the North American Chapter of the ACL（NAACL）, pp.234-240, 2000.

[6] Eduardo Blanco, Dan Moldovan. Unsupervised Learning of Semantic Relation Composition. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011.

[7] Johan Bos. Towards Wide-Coverage Semantic Interpretation. Proceedings of 6th International Workshop on Computational Semantics IWCS-6, pp.42-53, 2005.

[8] Bernhard E. Boser, Isabelle M. Guyon, Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. COLT'92 Proceedings of the fifth annual workshop on Computational learning theory, pp.144-152, 1992.

[9] Stephen Boxwell, Dennis Mehay, Chris Brew. Brutus: A Semantic Role Labeling System Incorporating CCG, CFG, and Dependency Features. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009.

[10] Ronald Brachman, James Schmolze. An Overview of the KL-ONE Knowledge Representation System. Cognitive Science 9, pp.171-216, 1985.

[11] Michael R. Brent. Automatic Acquisition of subcategorization frames from untagged text. ACL '91 Proceedings of the 29th annual meeting on Association for Computational Linguistics, pp.209-214, 1991.

[12] Michael R. Brent. From Grammar to Lexicon: Unsupervised lerning of lexical syntax. Computational Linguistics - Special issue on using large corpora: II Volume 19 Issue 2, pp.243-262 , 1993.

[13] Ted Briscoe, John Carroll. Automatic extraction of subcategorization from corpora. In Proceedings of the 5th ACL Conference on Applied Natural Language Processing, pp.356-363, 1997.

[14] Jean Carletta, Stefan Evert, Ulrich Heid, Jonathan Kilgour, Judy Robertson, Holger Voormann. The NITE XML toolkit: flexible annotation for multi-modal language data. Behavior Research Methods, Instruments, and Computers, 35(3), pp.353-363, 2003.

[15] Xavier Carreras, Lluís Màrquez. Introduction to the CoNLL-2005 shared task: semantic role labeling. CONLL '05 Proceedings of the Ninth Conference on Computational Natural Language Learning, pp.152-164, 2005.

[16] Eugene Charniak, Robert Goldman. A Logic for Semantic Interpretation. ACL '88 Proceedings of the 26th annual meeting on Association for Computational Linguistics, pp.87-94, 1988.

[17] Eugene Charniak. A Maximum-Entropy-Inspired Parser. NAACL 2000 Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference, pp.132-139, 2000.

[18] Eugene Charniak. Immediate-head parsing for language models. ACL'01 Proceedings of the 39th Annual Meeting on Association for Computational Linguistics. pp.124-131. 2001.

[19] Wanxiang Che, Ting Liu, Yongqiang Li. Improving Semantic Role Labeling with Word Sense. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010.

[20] John Chen, and Owen Rambow. Use of deep linguistics features for the recognition and labeling of semantic arguments. EMNLP '03 Proceedings of the 2003 conference on Empirical methods in natural language processing, pp.41-48, 2003.

[21] Wenliang Chen, Yujie Zhang, Hitoshi Isahara. An empirical study of Chinese chunking. COLING-ACL '06 Proceedings of the COLING/ACL on Main conference poster sessions, pp.97-104, 2006.

[22] Silvie Cinková, Martin Holub, Vincent Kríž. Managing Uncertainty in Semantic Tagging. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 2012.

[23] Stephen Clark, James R. Curran. Wide-Coverage Efficient Statistical Parsing with CCG and Log-Linear Models. Computational Linguistics, Volume 33, Issue 4, pp.493-552, 2007.

[24] Trevor Cohn, Philip Blunsom. Semantic role labeling with tree conditional random fields. CONLL '05 Proceedings of the Ninth Conference on Computational Natural Language Learning, pp.169-172, 2005.

[25] Michael Collins, Terry Koo. Discriminative Reranking for Natural Language Parsing. Computational Linguistics, Volume 31, Issue 1, pp.25-70, 2000.

[26] Danilo Croce, Cristina Giannone, Paolo Annesi, Roberto Basili. Towards Open-Domain Semantic Role Labeling. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010.

[27] Daniel Dahlmeier, Hwee Tou Ng. Domain Adaptation for Semantic Role Labeling in the Biomedical Domain. Bioinformatics, vol.26, no.8, pp.1098-1104, 2010.

[28] Mike Dowman, Valentin Tablan, Hamish Cunningham, Borislav Popov. 2005. Web-assisted annotation, semantic indexing and search of television and radio news. In Proceedings of the 14th International World Wide Web Conference, pp.225-234, 2005.

[29] Murat Ersan, Eugene Charniak. A statistical syntactic disambiguation program and what it learns. In S. Wermter, E. Riloff, and G. Scheler. editors, Connectionist, Statistical and Symbolic Approaches in Learning for Natural Language Processing, volume 1040 of Lecture Notes in Artifical Intelligence, pp.146-159, 1998.

[30] Christiane Fellbaum. WordNet: An Electronic Lexical Database. The MIT Press, 1998.

[31] Charles J Fillmore. The Case for Case. Universals in Linguistic Theory, 1968.

[32] Charles J. Fillmore, Collin F. Baker. Frame semantics for text understanding. In Proceedings of NAACL WordNet and Other Lexical Resources Workshop, 2001.

[33] Timothy Wilking Finin. THE SEMANTIC INTERPRETATION OF NOMINAL COMPOUNDS. AAAI-80 Proceedings, 1980.

[34] Michael Fleischman, Namhee Kwon, Eduard Hovy. Maximum Entropy Models for FrameNet Classification. EMNLP '03 Proceedings of the 2003 conference on Empirical methods in natural language processing, pp.49-56, 2003.

[35] Hagen Furstenau, and Mirella Lapata. Semi-Supervised Semantic Role Labeling. EACL '09 Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, pp.220-228, 2009.

[36] Qin Gao, Stephan Vogel. Corpus Expansion for Statistical Machine Translation with Semantic Role Label Substitution Rules. HLT '11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Volume 2, pp.294-298, 2011.

[37] Matthew Gerber, Joyce Y. Chai. Beyond NomBank: a Study of Implicit Arguments for Nominal Predicates. ACL '10 Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp.1583-1592, 2010.

[38] Daniel Gildea, Julia Hockenmaier. Identifying semantic roles using combinatory categorial grammar. EMNLP '03 Proceedings of the 2003 conference on Empirical methods in natural language processing, pp.57-64, 2003.

[39] Daniel Gildea, Martha Palmer. The necessity of syntactic parsing for predicate argument recognition. Proceeding ACL '02 Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp.239-246, 2002.

[40] Daniel Gildea, Daniel Jurafsky. Automatic Labeling of Semantic Roles. Computational Linguistics, Volume 28 Issue 3, pp.245-288, 2002.

[41] Ding Liu, Daniel Gildea. Semantic Role Features for Machine Translation. COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics, pp.716-724, 2010.

[42] Andrew S. Gordon, Reid Swanson. Generalizing Semantic Role Annotations Across Syntactically Similar Verbs. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pp.192-199, 2007.

[43] Trond Grenager, Christopher D. Manning. Unsupervised Discovery of a Statistical Verb Lexicon. EMNLP '06 Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pp.1-8, 2006.

[44] Jane Grimshaw. Argument Structure. MIT Press, 1990.

[45] Kadri Hacioglu, Sameer Pradhan, Wayne Ward, James H. Martin, Daniel Jurafsky. Semantic role labeling by tagging syntactic chunks. Proceedings of CoNLL-2004, pp.110-113, 2004.

[46] Aria Haghighi, Kristina Toutanova, Christopher D. Manning. A joint model for semantic role labeling. CONLL '05 Proceedings of the Ninth Conference on Computational Natural Language Learning, pp.173-176, 2005.

[47] Udo Hahn, Ekaterina Buyko, Katrin Tomanek, Scott Piao, John McNaught,

Yoshimasa Tsuruoka, Sophia Ananiadou. An annotation type system for a data-driven NLP pipeline. In Proceedings of the Linguistic Annotation Workshop, pp.33-40, 2007.

[48] Hindle, D. Rooth, M. Structural ambiguity and lexical relations. Computational Linguistics, 19(1), pp.103-120, 1993.

[49] Graeme Hirst. Semantic Interpretation and Ambiguity. Artificial Intelligence, Volume 34, Issue 2, pp.131-177, 1988.

[50] Thomas Hoffman. Probabilistic latent semantic indexing. In Proceedings of ACM SIGIR 1999, pp.50-57, 1999.

[51] Fei Huang, Alexander Yates. Open-Domain Semantic Role Labeling by Modeling Word Spans. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010.

[52] Edwin Thompson Jaynes. information theory and statistical mechanics. Physical Review, vol.106, Issue4, pp.620-630, 1957.

[53] Zheng Ping Jiang, Hwee Tou Ng. Semantic Role Labeling of NomBank: a maximum entropy approach. EMNLP '06 Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pp.138-145, 2006.

[54] Richard Johansson, Pierre Nugues. The effect of syntactic representation on semantic role labeling. In Proceedings of COLING, pp.18-22, 2008.

[55] Richard Johansson. Non-atomic Classification to Improve a Semantic Role Labeler for a Low-resource Language. SemEval '12 Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation, pp.95-99, 2012.

[56] H. Kamp, U. Reyle. From Discourse to Logic; Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Springer, 1993.

[57] Daisuke Kawahara, Sadao Kurohashi. Fertilization of Case Frame Dictionary for Robust Japanese Case Analysis. COLING '02 Proceedings of the 19th international conference on Computational linguistics - Volume 1, pp.1-7, 2002.

[58] Jin-Dong Kim, Tomoko Ohta, Yuka Tateisi, Junichi Tsujii. GENIA corpus - a semantically annotated corpus for bio-textmining. ISMB Supplement of Bioinformatics, pp.180-182, 2003.

[59] Paul Kingsbury, Martha Palmer. From TreeBank to PropBank. Proceedings of

the 3rd International Conference on Language Resources and Evaluation LREC2002, pp.1989-1993, 2002.

[60] Paul Kingsbury, and Martha Palmer. PropBank: The Next Level of TreeBank. The Second Workshop on Treebanks and Linguistic Theories（TLT 2003）, 2003.

[61] Karin Kipper, Hoa Trang Dang, Martha Palmer. Class-based construction of a verb lexicon. Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence, pp.691-696, 2000.

[62] Karin Kipper, Benjamin Snyder, and Martha Palmer. Extending a verb-lexicon using a semantically annotated corpus. In the Proceedings of the 4th International Conference on Language Resources and Evaluation（LREC-04）, 2004.

[63] Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. Extending VerbNet with Novel Verb Classes. Fifth International Conference on Language Resources and Evaluation（LREC 2006）, 2006.

[64] Peter Koomen, Vasin Punyakanok, Dan Roth and Wen-tau Yih. Generalized inference with multiple semantic role labeling systems. CONLL '05 Proceedings of the Ninth Conference on Computational Natural Language Learning, pp.181-184, 2005.

[65] Anna Korhonen. Subcategorization acquisition. PhD thesis published as Technical Report UCAM-CL-TR-530, Computer Laboratory, University of Cambridge, 2002.

[66] Joel Lang, Mirella Lapata. Unsupervised Induction of Semantic Roles. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010.

[67] Joel Lang, Mirella Lapata. Unsupervised Semantic Role Induction with Graph Partitioning. Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 2011.

[68] Joel Lang, Mirella Lapata. Unsupervised Semantic Role Induction via Split-Merge Clustering. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011.

[69] Maria Lapata. Acquiring lexical generalizations from corpora: A case study for diathesis alternations. In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, pp.397-404, 1999.

[70] Hang Li, Naoki Abe. Generalizing case frames using a thesaurus and the MDL principle. Computational Linguistics, Volume 24, Issue 2, pp.217-244, 1998.

[71] Junhui Li, Guodong Zhou, Hwee Tou Ng. Joint Syntactic and Semantic Parsing of Chinese. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010.

[72] Ting Liu, Wanxiang Che, Sheng Li, Yuxuan Hu, Huaijun Liu. Semantic role lableing system using maximum entropy classifier. CONLL '05 Proceedings of the Ninth Conference on Computational Natural Language Learning, pp.189-192, 2005.

[73] Xiaohua Liu, Kuan Li, Bo Han, Ming Zhou, Long Jiang, Zhongyang Xiong, Changning Huang. Semantic Role Labeling for News Tweets. Proceedings of the 23rd International Conference on Computational Linguistics（Coling 2010）, 2010.

[74] Edward Loper, Szu-ting Yi, Martha Palmer. Combining Lexical Resources: Mapping Between PropBank and VerbNet. In Proceedings of the 7th International Workshop on Computational Semantics, pp.118-128, 2007.

[75] Kikuo Maekawa, Hanae Koiso, Sadaoki Furui and Hitoshi Isahara. Spontaneous speech corpus of Japanese. Proceedings of the Second International Conference of Language Resources and Evaluation (LREC2000), pp.947-952, 2000.

[76] Olga Babko-Malaya. PropBank Annotation Guidelines. http://verbs.colorado.edu/~mpalmer/projects/ace/PBguidelines.pdf, 2005.

[77] Christopher D. Manning. Automatic acquisition of a large subcategorization dictionary from corpora. In Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics, pp.235-242, 1993.

[78] Mitchell P. Marcus, Mary Ann Marcinkiewicz and Beatrice Santorini. Building a Large Annotated Corpus of English: The Penn Treebank. Computational Linguistics - Special issue on using large corpora: II, Volume 19, Issue 2, pp.313-330, 1993.

[79] Mitch Marcus. A Brief History of the Penn Treebank. The Center For Language and Speech Processing Seminar, 2011.

[80] Adam Meyers, Ruth Reeves, Catherine Macleod, Rachel Szekely, Veronika Zielinska, Brian Young, Ralph Grishman. Annotating Noun Argument Structure for NomBank. Proceedings of the Fourth International Conference on Language Resources and Evaluation（LREC 2004）, pp.803-806, 2004.

[81] Adam Meyers, Ruth Reeves, Catherine Macleod, Rachel Szekely, Veronika

Zielinska, Brian Young and Ralph Grishman. The NomBank Project: An Interim Report. In Proceedings of the NAACL/HLT Workshop on Frontiers in Corpus Annotation, pp.24-31, 2004.

[82] Ivan Meza-Ruiz and Sebastian Riedel. Jointly identifying predicates, arguments and senses using markov logic. In Proceedings of NAACL/HLT-2009, pp.155-163, 2009.

[83] Ivan Meza-Ruiz, Sebastian Riedel. Jointly identifying predicates, arguments and senses using Markov logic. NAACL '09 Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp.155-163, 2009.

[84] Ivan Meza-Ruiz, Sebastian Riedel. Multilingual semantic role labelling with markov logic. CoNLL '09 Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task, pp.85-90, 2009.

[85] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, Katherine Miller. Introduction to WordNet: An On-line Lexical Database. International Journal of Lexicography, Volume 3, Issue 4, pp.235-244, 1990.

[86] George A. Miller. Nouns in WordNet: A Lexical Inheritance System. International Journal of Lexicography, Volume 3, Issue 4, pp.245-264, 1990.

[87] Scott Miller, David Stallard, Robert Bobrow, Richard Schwartz. A fully statistical approach to natural language interfaces. In Proceedings of the 34th Annual Meeting of the ACL, pp.55-61, 1996.

[88] Robert C. Moore. Unification-based semantic interpretation. ACL '89 Proceedings of the 27th annual meeting on Association for Computational Linguistics, pp.33-41, 1989.

[89] Alessandro Moschitti, Ana-Maria Giuglea, Bonaventura Coppola and Roberto Basili. Hierarchical Semantic Role Labeling. CONLL '05 Proceedings of the Ninth Conference on Computational Natural Language Learning, pp.201-204, 2005.

[90] Smruthi Mukund, Debanjan Ghosh, Rohini Srihari. Using Cross-Lingual Projections to Generate Semantic Role Labeled Annotated Corpus for Urdu - A Resource Poor Language. Proceedings of the 23rd International Conference on Computational Linguistics（Coling 2010）, 2010.

[91] Lluís Márquez, Mihai Surdeanu, Pere Comas, Jordi Turmo. A robust combination strategy for semantic role labeling. HLT '05 Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp.644-651, 2005.

[92] Lluís Màrquez, Xavier Carreras, Kenneth C. Litkowski, Suzanne Stevenson. Semantic Role Labeling: An Introduction to the Special Issue. Computational Linguistics, Volume 34, Issue 2, pp.145-159, 2008.

[93] Sebastian Padó, Mirella Lapata. Dependency-based construction of semantic space models. Computational Linguistics, Volume 33, Issue 2, pp.161-199, 2007.

[94] Martha Palmer, Daniel Gildea and Paul Kingsbury. The Proposition Bank: An Annotated Corpus of Semantic Roles. Computational Linguistics, Volume 31, Issue 1, pp.71-106, 2005.

[95] Martha Palmer, Daniel Gildea, Nianwen Xue. Semantic Role Labeling. MORGEN & CLAYPOOL PUBLISHERS, 2010.

[96] Martha Palmer. A Class-Based Verb Lexicon. http://verbs.colorado.edu/~mpalmer/projects/verbnet.html.

[97] Andrew Philpot, Eduard Hovy, Patrick Pantel. The Omega Ontology. Proceedings of the ONTOLEX Workshop at the International Conference on Natural Language Processing（IJCNLP）, 2005.

[98] Carl Pollard, Ivan A. Sag. Head-Driven Phrase Structure Grammar. University of Chicago Press, 1994.

[99] Hoifung Poon, Pedro Domingos. Unsupervised semantic parsing. EMNLP '09 Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Volume 1, pp.1-10, 2009.

[100] Sameer Pradhan, WayneWard, Kadri Hacioglu, James H. Martin, Daniel Jurafsky. Shallow semantic parsing using Support Vector Machines. In Proceedings of NAACL-HLT, 2004.

[101] Sameer Pradhan, Kadri Hacioglu, Valerie Krugler, Wayne Ward, James H. Martin, Daniel Jurafsky. Support Vector Learning for Semantic Argument Classification. Machine Learning, Volume 60, Issue 1-3, pp.11-39, 2005.

[102] Sameer Pradhan, Kadri Hacioglu, Wayne Ward, James H. Martin, Daniel Jurafsky. Semantic role chunking combining complementary syntactic views. CONLL '05 Proceedings of the Ninth Conference on Computational Natural Language Learning, pp.217-220, 2005.

[103] Sameer Pradhan, Wayne Ward, Kadri Hacioglu, James H. Martin, Daniel Jurafsky. Semantic role labeling using different syntactic views. ACL '05 Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pp.581-588, 2005.

[104] Sameer S. Pradhan, Eduard Hovy, Mitch Marcus, Martha Palmer, Lance

Ramshaw, Ralph Weischedel. OntoNotes: A Unified Relational Semantic Representation. ICSC '07 Proceedings of the International Conference on Semantic Computing, pp.517-526, 2007.

[105] Sameer S. Pradhan, Wayne Ward, James H. Martin. Towards Robust Semantic Role Labeling. Computational Linguistics, Volume 34, Number 2, 2008.

[106] Princeton University. WordNet Search 3.1. http://wordnetweb.princeton.edu/perl/webwn.

[107] Vasin Punyakanok, Dan Roth, Wen-tau Yih, Dav Zimak. Semantic Role Labeling via Integer Linear Programming Inference. COLING '04 Proceedings of the 20th international conference on Computational Linguistics, Article No.1346, 2004.

[108] Vasin Punyakanok, Dan Roth, Wen-tau Yih, Dav Zimak, Yuancheng Tu. Semantic role labeling via generalized inference over classifiers. In Proceedings of CoNLL, pp.130-133, 2004.

[109] Vasin Punyakanok, Dan Roth, Wen-tau Yih. The necessity of syntactic parsing for semantic role labeling. IJCAI'05 Proceedings of the 19th international joint conference on Artificial intelligence, pp.1117-1123, 2005.

[110] Vasin Punyakanok, Dan Roth, Wen-tau Yih. The Importance of Syntactic Parsing and Inference in Semantic Role Labeling. Computational Linguistics, Volume 34, Number 2, 2008.

[111] James Pustejovsky, Peter G. Anick. On the semantic interpretation of nominals. COLING '88 Proceedings of the 12th conference on Computational linguistics - Volume 2, pp.518-523, 1988.

[112] Ross Quinlan. Data Mining Tools See5 and C5.0. http://www.rulequest.com/see5-info.html. 2002.

[113] OntoNotes Release 4.0. http://www.bbn.com/NLP/OntoNotes, 2010.

[114] Stephen D. Richardson, William B. Dolan, Lucy Vanderwende. MindNet : acquiring and structuring semantic information from text. ACL '98 Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 2, pp.1098-1102, 1998.

[115] Matthew Richardson, Pedro Domingos. Markov Logic Networks. Machine Learning, Volume 62, Numbers 1-2, pp.107-136, 2006.

[116] Ellen Riloff, Mark Schmelzenbach. An empirical approach to conceptual case frame acquisition. In Proceedings of the Sixth Workshop on Very Large

Corpora, pp.49-56, 1998.

[117] Ronald L. Rivest. Learning Decision Lists. Machine Learning, Volume 2, Issue 3, pp.229-246, 1987.

[118] Rasoul Samad Zadeh Kaljahi. Adapting Self-Training for Semantic Role Labeling. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010.

[119] Anoop Sarkar, Daniel Zeman. Automatic extraction of subcategorization frames for Czech. In Proceedings of the 19th International Conference on Computational Linguistics, pp.691-697, 2000.

[120] James Schmolze, Thomas Lipkis. Classification in the KL-ONE Knowledge Representation System. Proceedings of the Eighth International Joint Conference on Artificial Intelligence, IJCAI. 1983.

[121] Dan Shen, Mirella Lapata. Using Semantic Roles to Improve Question Answering. In Proceedings of the 12th Conference on Empirical Methods in Natural Language Processing, pp.12-21, 2007.

[122] Hiroyuki Shinnou. Detection of Japanese homophone Errors by a Decision List Including a Written Word as a Default Evidence. Proceeding EACL '99 Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics, pp.180-187, 1999.

[123] Norman K. Sondheimer, Ralph M. Weischedel, Robert J. Bobrow. Semantic interpretation using KL-ONE. ACL '84 Proceedings of the 10th International Conference on Computational Linguistics and 22nd annual meeting on Association for Computational Linguistics, pp.101-107, 1984.

[124] Vivek Srikumar, Dan Roth. A Joint Model for Extended Semantic Role Labeling. Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 2011.

[125] Weiwei Sun, Zhifang Sui, Meng Wang, Xin Wang. Chinese semantic role labeling with shallow parsing. EMNLP '09 Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Volume 3, pp.1475-1483, 2009.

[126] Weiwei Sun, Zhifang Sui, Meng Wang. Prediction of Thematic Rank for Structured Semantic Role Labeling. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009.

[127] Weiwei Sun. Semantics-Driven Shallow Parsing for Chinese Semantic Role Labeling. Proceedings of the 48th Annual Meeting of the Association for

Computational Linguistics, 2010.

[128] Weiwei Sun. Improving Chinese Semantic Role Labeling with Rich Syntactic Features. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010.

[129] Mihai Surdeanu, Sanda Harabagiu, John Williams, Paul Aarseth. Using predicate-argument structures for information extraction. In Proceedings of the 41stAnnual Meeting of the Association for Computational Linguistics, pp.8-15, 2003.

[130] Mihai Surdeanu, Lluís Màrquez, Xavier Carreras, Pere R. Comas. Combination Strategies for Semantic Role Labeling. Journal of Artificial Intelligence Research（JAIR）, 29, pp.105-151, 2007.

[131] Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies. In Proceedings of CoNLL-2008, pp.159-177, 2008.

[132] Robert S. Swier, Suzanne Stevenson. Unsupervised Semantic Role Labeling. In Proceedings of the Conference on Empirical Methods in Natural Language Processing（EMNLP）, pp.95-102, 2004.

[133] Ivan Titov, Alexandre Klementiev. A Bayesian Approach to Unsupervised Semantic Role Induction. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 2012.

[134] Erik F. Tjong Kim Sang, Jorn Veenstra. Representing text chunks. EACL '99 Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics, pp.173-179, 1999.

[135] Kristina Toutanova, Aria Haghighi, Christopher D. Manning. Joint learning improves Semantic Role Labeling. ACL '05 Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pp.589-596, 2005.

[136] Kristina Toutanova, Aria Haghighi, Christopher D. Manning. A Global Joint Model for Semantic Role Labeling. Computational Linguistics, Volume 34, 2008.

[137] Tzong-Han Tsai, Chia-Wei Wu, Yu-Chun Lin, Wen-Lian Hsu. Exploiting Full Parsing Information to Label Semantic Roles Using an Ensemble of ME and SVM via Integer Linear Programming. CONLL '05 Proceedings of the Ninth Conference on Computational Natural Language Learning, pp.233-236, 2005.

[138] Richard Tzong-han Tsai, Wen-chi Chou, Yu-chun Lin, Cheng-lung Sung, Wei

Ku, Ying-shan Su, Ting-yi Sung, Wen-lian Hsu. Biosmile: Adapting semantic role labeling for biomedical verbs: An exponential model coupled with automatically generated template features. Workshop: BioNLP Workshop On Linking Natural Language And Biology, 2006.

[139] Richard Tzong-Han Tsai, Wen-Chi Chou, Ying-Shan Su, Yu-Chun Lin, Cheng-Lung Sung, Hong-Jie Dai, Irene Tzu-Hsuan Yeh, Wei Ku, Ting-Yi Sung, Wen-Lian Hsu. BIOSMILE: A semantic role labeling system for biomedical verbs using a maximum-entropy model with automatically generated template features. BMC Bioinformatics 2007, 8:325, 2007.

[140] Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, Yasemin Altun. Support vector machine learning for interdependent and structured output spaces. ICML '04 Proceedings of the 21st international conference on Machine learning, pp.104-111, 2004.

[141] Ushioda, A., Evans, D., Gibson, T., &Waibel, A. The automatic acquisition of frequencies of verb subcategorization frames from tagged corpora. In Boguraev, B., & Pustejovsky, J.（eds.）, SIGLEX ACL Workshop on the Acquisition of Lexical Knowledge from Text, pp.95-106, 1993.

[142] Takehito Utsuro, Yuji Matsumoto . Learning Probabilistic Subcategorization Preference and its Application to Syntactic Disambiguation. Technical Report NAIST-IS-TR97006, Graduate School of Information Science, 1997.

[143] Takehito Utsuro, Yuji Matsumoto. Learning probabilistic subcategorization preference by identifying case dependencies and optimal noun class generalization level. Proceedings of the Fifth Conference on Applied Natural Language Processing, pp.364-371, 1997.

[144] Vladimir Vapnik, A. Lerner. Pattern recognition using generalized portrait method. Automation and Remote Control, Vol. 24, 1963.

[145] Rui Wang, Yi Zhang. Recognizing Textual Relatedness with Predicate-Argument Structures. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2009.

[146] W. A. Woods. Progress in natural language understanding-An application to lunar geology. AFIPS '73 Proceedings of the June 4-8, 1973, national computer conference and exposition, pp.441-450, 1973.

[147] Dekai Wu, Pascale Fung. Can Semantic Role Labeling Improve SMT. In Proceedings of the 13th Annual Conference of the European Association for Machine Translation, pp.218-225, 2009.

[148] Fei Xia, Martha Palmer, Nianwen Xue, Mary Ellen Okurowski, John Kovarik,

Fu-Dong Chiou, Shizhe Huang, Tony Kroch, Mitch Marcus. Developing Guidelines and Ensuring Consistency for Chinese Text Annotation. Proceedings of the second International Conference on Language Resources and Evaluation（LREC 2000）, 2000.

[149] Nianwen Xue, Martha Palmer. Annotating the propositions in the Penn Chinese Treebank. SIGHAN '03 Proceedings of the second SIGHAN workshop on Chinese language processing, Volume 17, pp.47-54, 2003.

[150] Nianwen Xue, Martha Palmer. Calibrating Features for Semantic Role Labeling. Empirical Methods in Natural Language Processing Conference, in conjunction with the 42nd Meeting of the Association for Computational Linguistics, pp.21-26, 2004.

[151] Nianwen Xue. Labeling Chinese Predicates with Semantic Roles. Computational Linguistics, Volume 34, Number 2, 2008.

[152] David Yarowsky. Decision Lists For Lexical Ambiguity Resolution: Application to Accent Restoration in Spanish and French. Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, 1994.

[153] David Yarowsky. Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. ACL '95 Proceedings of the 33rd annual meeting on Association for Computational Linguistics, pp.189-196, 1995.

[154] Szu-ting Yi, Martha Palmer. The integration of syntactic parsing and semantic role labeling. CONLL '05 Proceedings of the Ninth Conference on Computational Natural Language Learning, pp.237-240, 2005.

[155] Szu-ting Yi, Edward Loper, Martha Palmer. Can Semantic Roles Generalize Across Genres? Proceedings of HLT-NAACL 2007, pp.548-555, 2007.

[156] Scott Wen-tau Yih, Kristina Toutanova. Automatic Semantic Role Labeling. HLT-NAACL-06 Tutorial, 2006.

[157] Beñat Zapirain, Eneko Agirre, Lluís Màrquez. Robustness and Generalization of Role Sets: PropBank vs. VerbNet. In Proceedings of ACL-08: HLT, pp.550-558, 2008.

[158] Benat Zapirain, Eneko Agirre, Lluís Màrquez. Generalizing over lexical features: Selectional preferences for semantic role classification. ACLShort'09 Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pp.73-76, 2009.

[159] Benat Zapirain, Eneko Agirre, Lluís Màrquez, Mihai Surdeanu. Improving Semantic Role Classification with Selectional Preferences. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of

the Association for Computational Linguistics, 2010.
[160] Tao Zhuang, Chengqing Zong. A Minimum Error Weighting Combination Strategy for Chinese Semantic Role Labeling. Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), 2010.
[161] チャールズ フィルモア（田中春美, 船城道雄訳）. 格文法の原理. 三省堂, 1975.
[162] 宇津呂武仁, 松本裕治, 長尾眞. 二言語対訳コーパスからの動詞の格フレーム獲得. 情報処理学会論文誌, 34(5), pp.913-924, 1993.
[163] 宇津呂 武仁, 松本 裕治. コーパスからの下位範疇化優先度の学習：隠れ変数を用いた格の依存関係・格要素の汎化レベルの曖昧性の取り扱い. 電子情報通信学会技術研究報告, 96(294), pp.31-38, 1996.
[164] 宇津呂武仁, 宮田高志, 松本裕治. 最大エントロピー法による下位範疇化の確率モデル学習および統語的曖昧性解消による評価. 情報処理学会研究報告, 97(53), pp.69-76, 1997.
[165] 下村 拓也, 竹内 孔一. 名詞の概念体系を利用した規則に基づく意味役割付与モデルの構築. 情報処理学会研究報告. 情報学基礎研究会報告, 2006(94), pp.13-20. 2006.
[166] 河原大輔, 黒橋禎夫. 用言と直前の格要素の組を単位とする格フレームの自動獲得. 情報処理学会研究報告, 2000(107), pp.127-134, 2000.
[167] 河原大輔, 黒橋禎夫, 橋田浩一. 「関係」タグ付きコーパスの作成. 言語処理学会第 8 回年次大会発表論文集, pp.495-498, 2002.
[168] 河原大輔, 黒橋禎夫. 頑健な格解析を実現する格フレーム辞書の自動構築. 言語処理学会第 8 回年次大会, pp.515-518, 2002.
[169] 河原大輔, 黒橋禎夫. 自動構築した格フレーム辞書と先行詞の位置選好順序を用いた省略解析. 自然言語処理, Vol.11, No.3, pp.3-19, 2004.
[170] 河原大輔, 黒橋禎夫. 格フレーム辞書の漸次的自動構築. 自然言語処理, Vol.12, No.2, pp.109-131, 2005.
[171] 河原大輔, 黒橋禎夫. 高性能計算環境を用いた Web からの大規模格フレーム構築. 情報処理学会研究報告 自然言語処理研究会報告, 2006(1), pp.67-73, 2006.
[172] 丸山岳彦, 柏岡秀紀, 熊野正, 田中英輝. 日本語節境界検出プログラムCBAPの開発と評価. 自然言語処理, Vol.11, No.4, pp.39-68, 2004.
[173] 吉川克正, 浅原正幸, 松本裕治. Markov Logic による日本語述語項構造解析. 情報処理学会研究報告, 自然言語処理研究会報告, 199(5), pp.1-7, 2010.
[174] 宮田高志, 宇津呂武仁, 松本裕治. Bayesian Network による下位範疇化の確率モデルおよびその学習. 情報処理学会研究報告 自然言語処理研究会報告, 97(53), pp.77-84, 1997.
[175] 京都大学 大学院情報学研究科 黒橋・河原研究室. 日本語形態素解析システム

JUMAN. http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN.
[176] 京都大学 大学院情報学研究科 黒橋・河原研究室. 日本語構文・格解析システム KNP. http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP.
[177] 橋田浩一, GDA 日本語アノテーションマニュアル 0.74 版, http://i-content.org/gda/tagman.html, 2005.
[178] 橋田浩一. 岩波国語辞典のアノテーション― 照応・共参照・項構造 ―. http://i-content.org/rwcDB/iwanami/doc/tag.html. 2006.
[179] 橋本 順子, 峯恒憲, 雨宮真人. 既存の和語動詞の格フレームを利用したサ変動詞の格フレーム獲得. 情報処理学会研究報告, 自然言語処理研究会報告, 97(69), pp.125-132, 1997.
[180] 橋本力, 河原大輔, 黒橋禎夫, 新里圭司. 構文・照応・評判情報つきブログコーパスの構築. 言語処理学会第 15 回年次大会発表論文集, pp.614-617, 2009.
[181] 桑畑和佳子, 橋本美奈子, 村田賢一. 計算機用日本語辞書の開発. 情報処理学会研究報告 93 巻 42 号, pp.27-34, 1993.
[182] 工藤拓, 松本裕治. 半構造化テキストの分類のためのブースティングアルゴリズム. 情報処理学会論文誌, 45(9), pp.2146-2156, 2004.
[183] 日本語話し言葉コーパスの構築法. http://www.ninjal.ac.jp/products-k/katsudo/seika/corpus/csj_report/. 2006.
[184] 現代日本語書き言葉均衡コーパス設計の基本方針. 大学共同利用機関法人人間文化研究機構国立国語研究所 コーパス整備計画 KOTONOHA http://www.ninjal.ac.jp/kotonoha/ex_2.html, 2011.
[185] 笹野遼平, 河原大輔, 黒橋禎夫. 名詞格フレーム辞書の自動構築とそれを用いた名詞句の関係解析. 自然言語処理, Vol.12, No.3, pp.129-144, 2005.
[186] 山下浩, 田中茂. サポートベクターマシンとその応用. Visual Mining Studio 技術参考資料. http://www.msi.co.jp/vmstudio/materials.html.
[187] 渋木英潔, 荒木健治, 栃内香次. 一文一格の原理と深層格選好に基づいた日本語深層格規則の自動獲得手法. 情報処理学会研究報告 自然言語処理研究会報告, 2003(76), pp.15-22, 2003.
[188] 渋木英潔, 荒木健治, 桃内佳雄, 栃内香次. 深層格の推測手法における自動クラスタリングの利用. FTT2004(第 3 回情報科学技術フォーラム), 情報科学技術レターズ 3, pp.79-80, 2004.
[189] 渋木英潔, 荒木健治, 桃内佳雄, 栃内香次. 深層格選好に基づく深層格推測手法の英文への適用. 言語処理学会第 11 回年次大会発表論文集, pp.1056-1059, 2005.
[190] 渋木英潔, 荒木健治, 桃内佳雄, 栃内香次. 単語概念の深層格選好に基づく深層格推測手法. 電子情報通信学会論文誌, D 情報・システム, J89-D(6),

pp.1413-1428, 2006.
[191] 春野雅彦. 最小汎化とオッカムの原理を用いた動詞格フレーム学習. 情報処理学会研究報告 95 巻 69 号, pp.29-36, 1995.
[192] 小山正太, 乾伸雄, 小谷善行. 「名詞と表層格」パターンに対する深層格対応の推測. 情報処理学会研究報告 自然言語処理研究会報告, 2003(23), pp.153-160, 2003
[193] 小町守, 飯田龍, 乾健太郎, 松本裕治. 共起用例と名詞の出現パターンを用いた動作性名詞の項構造解析. 言語処理学会第 12 回年次大会発表論文集, pp.821-824, 2006
[194] 小町守, 飯田龍, 乾健太郎, 松本裕治. 事態性名詞の項構造解析における共起尺度と構文パターンの有効性の分析. 言語処理学会第 13 回年次大会発表論文集, pp.47-50, 2007.
[195] 小町守, 飯田龍, 乾健太郎, 松本裕治. 名詞句の語彙統語パターンを用いた事態性名詞の項構造解析. 自然言語処理, Vol.17, No.1, pp.141-159, 2010.
[196] 小町守, 飯田龍. BCCWJ に対する述語項構造と照応関係のアノテーション. 『現代日本語書き言葉コーパス』完成記念講演会予稿集, pp.77-82, 2011.
[197] 松林優一郎, 岡崎直観, 辻井潤一. 自動意味役割付与における意味役割の汎化. 自然言語処理, Vol.17, No.4, pp.59-89, 2010.
[198] 上之薗和宏, 榎津秀次, 古宮誠一. 遺伝的プログラミングによる係り受け関係からの格フレーム辞書自動生成システム. 電子情報通信学会技術研究報告, KBSE, 知能ソフトウェア工学, 103(709), pp.43-48, 2004.
[199] 新納浩幸. 決定リストを弱学習器としたアダブーストによる日本語単語分割. 自然言語処理, Vol.8, No.2, pp.3-18, 2001.
[200] 石川徹也, 坂本義行, 佐藤雅之. Mu プロジェクトにおける意味マーカの概念と体系. 情報処理学会研究報告 86 巻 32 号, pp.1-12, 1986.
[201] 石川徹也, 坂本義行. 動詞意味機能に基づく日本語格フレームの生成. 情報処理学会研究報告 89 巻 27 号, pp.1-7, 1989.
[202] 前川喜久雄, 小磯花絵, 籠宮隆之, 古井貞熙, 井佐原均. 共通日本語話し言葉コーパスの設計. 日本音響学会 2000 年春季研究発表会講演論文集, pp.193-194, 2000.
[203] 大石亨, 松本裕治. 格パタン分析を利用した深層格獲得手法について. 情報処理学会論文誌, 36(11), pp.2597-2610, 1995.
[204] 大竹清敬, 根津雅彦, 増山繁, 山本和英. 語順を考慮した格フレームの提案と獲得手法. 電子情報通信学会論文誌, D-II, 情報・システム, II-パターン処理, J83-D-II(3), pp.1060-1063, 2000.
[205] 竹内孔一, 乾健太郎, 竹内奈央, 藤田篤. 意味の包含関係に基づく動詞項構造の

細分類. 言語処理学会第 14 回年次大会発表論文集, pp.1037-1040, 2008.
[206] 竹内孔一, 土山傑, 守屋将人, 森安祐樹. 類似した動作や状況を検索するための意味役割及び動詞語義付与システムの構築. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション. 109(390), pp.1-6, 2010.
[207] 竹内孔一. 動詞項構造シソーラスの構築. 人工知能学会第 25 回全国大会. 2011.
[208] 長尾真. 自然言語処理. 岩波書店, 1996.
[209] 辻井潤一, 山梨正明. 格とその認定基準. 情報処理学会研究報告 52 巻 3 号, pp.1-7, 1985.
[210] 鶴岡慶雅, 近山隆. ベイズ推定による決定リストのルール信頼度推定法. 情報処理学会研究報告, 自然言語処理研究会報告, 2001(69), pp.91-97, 2001.
[211] 渡邉 陽太郎, 浅原 正幸, 松本 裕治. 述語語義と意味役割の結合学習のための構造予測モデル. 人工知能学会論文誌 25(2), pp.252-261, 2010.
[212] 東優, 峯恒憲, 雨宮真人. 既存の概念辞書を用いた動詞語義による文の分類. 電子情報通信学会技術研究報告96巻294号, pp.39-44, 1996.
[213] 藤田篤, 乾健太郎, 松本裕治. 自動生成された言い換え文における不適格な動詞格構造の検出. 情報処理学会論文誌, Vol.45, No.4, pp.1176-1187, 2004.
[214] 内元清貴, 丸山岳彦, 高梨克也, 井佐原均. 『日本語話し言葉コーパス』における係り受け構造付与. 平成 15 年度国立国語研究所公開研究発表会, pp.35-36, 2003.
[215] 白井清昭, 橋本泰一, 西館耕介, 徳永健伸, 田中穂積. 決定リストを用いた形容詞の修飾先の決定. 言語処理学会第 7 回年次大会, pp.253-256, 2001.
[216] 八木豊, 橋本泰一, 美野秀弥, 徳永健伸, 田中穂積. 決定リストにおける規則の適用順序に関する考察. 情報処理学会研究報告, 自然言語処理研究会報告, 2001(112), pp.21-26, 2001.
[217] 飯田龍, 小町守, 乾健太郎, 松本裕治. NAIST テキストコーパス: 述語項構造と共参照関係のアノテーション. 情報処理学会自然言語処理研究会予稿集, NL-177-10, pp.71-78, 2007.
[218] 飯田龍, 小町守, 井之上直也, 乾健太郎, 松本裕治. 述語項構造と照応関係のアノテーション:NAIST テキストコーパス構築の経験から. 自然言語処理, Vol.17, No.2, pp.25-50, 2010.
[219] 肥塚真輔, 岡本紘幸, 斎藤博昭, 小原京子. 日本語フレームネットに基づく意味役割推定. 自然言語処理, 14(1), pp.43-66, 2007.
[220] 平博順, 永田昌明. 構造学習を用いた述語項構造解析. 言語処理学会第 14 回年次大会発表論文集, pp.556-559, 2008.
[221] 平博順, 藤田早苗, 永田昌明. 決定リストを用いた述語項構造解析. 言語処理学会第 15 回年次大会予稿集, pp.368-371, 2009.

[222] 本木実, 嶋津好生, 高橋直人. 階層型ニューラルネットワークによる深層格解析. 情報処理学会論文誌, 41(10), pp.2852-2862, 2000.
[223] 林部祐太. "意味"を計算機で扱う一方法 ―格解析入門―. 第 2 回言語学×自然言語処理合同勉強会, 2010.
[224] 浪岡保男, 浦谷則好, 相沢輝昭. ニュース文における深層格抽出手法. 情報処理学会研究報告 77 巻 3 号, pp.1-8, 1990.
[225] 颯々野学, 宇津呂武仁. 統計的日本語固有表現抽出における固有表現まとめ上げ手法とその評価. 情報処理学会研究報告, 自然言語処理研究会報告, 2000(86), pp.1-8, 2000.
[226] 特定非営利活動法人 言語資源協会(GSK) 言語資源カタログ. http://www.gsk.or.jp/catalog.html.
[227] Ruslan Mitkov. Anaphora Resolution, Studies in Language and Linguistics, Pearson Education, 2002.