

コーパスを用いた言語知識の獲得

Acquisition of Linguistic Knowledge from Natural Language Corpora

宇津呂 武仁* 松本 裕治*
Takehito Utsuro Yuji Matsumoto

* 奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology.

1995年1月9日 受理

Keywords: natural language processing, knowledge acquisition, natural language corpora.

1. はじめに

質が高く対象範囲の広い自然言語処理システムを実現するためには十分な量と質を持った辞書の構築が不可欠である。しかし、人手による辞書の作成には膨大な労力が必要であるうえ、記述の一貫性を保つのが困難であるなど多くの問題がある。特に、今後、さまざまな分野において、分野依存の辞書が必要になってくると考えると、これらの分野依存の辞書を個別に人手で作成することは、現実的に困難であると思われる。これに対して、近年、大量の電子化テキストや機械可読辞書が入手可能になってきており、これらの自然言語資源から言語知識、特に、語彙に関する知識を獲得する研究が盛んになっている*1。

語彙に関する最も基礎的な情報の獲得としては、クラスタリングによる語の意味分類や共起知識の獲得などがあるが、これらについては本号の工藤・井ノ上両氏の解説[工藤 95]で述べられている。これに対して、本稿では、動詞の格フレームや語の階層構造などのより意味的な情報をコーパスおよび辞書から獲得する手法について最近の研究動向を概説する。以下では、単言語コーパスからの語彙的知識獲得(2章)、辞書の定義文からの語彙的知識獲得(3章)、対訳コーパスからの構造的言語知識の獲得(4章)について順に述べる。

2. 単言語コーパスからの語彙的知識の獲得

本章では、単言語コーパスからの動詞・名詞の語彙的知識の抽出を扱った研究を紹介する。

2・1 動詞の語彙的知識の獲得

動詞の語彙的知識のうち、自然言語コーパスからの獲得が比較的良好に研究されているものとして動詞の格フレームに関するものがある。ここでは、まず、動詞の語彙的知識獲得に関する研究を大きく以下のように分類する。

- (1) 動詞が表層的・統語的にどのような引数をとるかを規定した表層格フレームあるいは下位範ちゅう化フレームレベルの知識を獲得するもの[Brent 93, 平岡 94, Manning 93, Poznański 93, Resnik 92, Ushioda 93, 内元 94].
- (2) 表層の手掛りをもとに、動詞の引数となる要素が深層的・意味的にどのような役割を持つかを規定した深層格フレームあるいは Thematic Structure レベルの知識などを獲得するもの[Aone 93, Brent 91, Liu 93, 大石 94, Pugeault 94, Webster 89, Zernik 90].
- (3) 動詞がどのような名詞と共起するかという基準で動詞間の類似度を測り、動詞の意味的クラスタリングを行うもの[Basili 93, Fukumoto 94].

これらのうち、(1)の表層格フレームあるいは下位範ちゅう化フレームの獲得において問題となるのは、**a**コーパスの形態(品詞づけなし、品詞づけあり、構文解析済み)、**b**コーパスに対する解析の度合(パターンマ

*1 自然言語コーパスや機械可読辞書から語彙的知識を獲得する研究の重要性については、Zernik のサーベイ[Zernik 91]に詳細に述べられているので参考にされたい。

ッチング, 構文解析), ㉔表層的・統語的な引数の種類, ㉕格要素の名詞に対する意味的制約の記述の度合, といったことである。これに対して, [Brent 93]では, 品詞づけされていない英文コーパスに対して, 代名詞・助動詞・冠詞・接続詞・句読点だけからなる小規模な辞書だけを用いて簡単な構文解析を行い, 引数となるものが, 名詞句, that 節, 不定詞句, 名詞句+that 節, 名詞句+不定詞句, 名詞句+名詞句の6種類に限定した下位範ちゅう化フレームを獲得している。[Manning 93]では, 統計的品詞づけプログラムおよび統語的曖昧性を含む部分をスキップする構文解析プログラムにより品詞づけされていない英文テキストを解析した結果から, 19個の下位範ちゅう化フレームを獲得することを試みている。[Poznański 93]では, 構文解析ずみの英文コーパスを用いて, 既存のシソーラス(LLOCE: Longman Lexicon of Contemporary English)中の動詞の意味カテゴリーと動詞の表層格フレームを対応づけることを試みている。動詞の表層格フレームとしては, 下位範ちゅう化パターンを表すLLOCEの文法コード(約50個)に格要素の名詞の意味クラス(18個)を付加したものをを用いている。[Ushioda 93]では, 品詞づけされた英文コーパスに対して, 正規表現で記述されたパターンマッチング規則を適用し, [Brent 93]と同じ6種類の下位範ちゅう化フレームの獲得を行っている。また, [Resnik 92]では, 品詞づけされた英文コーパスおよび英語シソーラス(WordNet [Miller 90])を用いて, 動詞の直接目的語がとり得る意味クラスを記述することを試みている。

一方, 日本語に関しては, 構文解析ずみコーパスを入力として, 格要素の名詞の意味的制約を記述するのに名詞のクラスタリング結果のクラスタを用いるもの[平岡 94]や, 構文解析ずみの動詞句の格要素に名詞シソーラス中の意味カテゴリーを付与した用例に対して, 類義語の用例を利用することによって動詞の多義性を類別する手法を述べたもの[内元 94]などがある。これらの日本語を扱った研究では, 格助詞パターンの種類に対する制限はなく, 任意のパターンを扱っている。また, 一般に, これら(1)の表層格フレーム獲得においては, 格の任意性の判定が困難な場合が多い。頻度情報をもとに格の任意性を判定しているもの[Manning 93]もあるが, いまのところ有力な方法は確立されていない。

次に, (2)の深層格フレームあるいは Thematic

Structureの獲得の研究においては, まず, 深層格あるいは深層格パターンとしてどのようなものを想定するかが問題となるが, これについては言語学の分野で比較的良好に用いられているもの(例えば, [Jackendoff 90]など)をそのまま用いることが多い。ここでは, まず, (2)に関する研究を, ㉖例文ごとの深層格割当てを中心に扱ったもの[Liu 93, Pugeault 94, Zernik 90], ㉗深層格フレームなどの観点からあらかじめ動詞のクラスを設定しておき, 動詞ごとのクラス割当てを中心に扱ったもの[Aone 93, Brent 91, 大石 94, Webster 89], の二つに分けて考える。

まず, ㉖では, それぞれの深層格に対してどのような表層的手掛りを設定するかが問題となる。例えば, 英文の構文解析ずみコーパスから動詞の Thematic Structure を獲得する手法を述べた[Liu 93]では, 行為主体を表す意味的役割“agent”について, ①Animateである, ②主語になり得る, ③目的語にならない, ④前置詞句になる場合の前置詞は“by”である, ⑤意図を表す句(“in order to”, “deliberately”など)とともに現れやすい, ⑥命令文で省略される主語である, といったヒューリスティクスを用意している。あらゆる意味的役割に対してこのようなヒューリスティクスを用意しておき, 意味的役割の割当ての曖昧性を解消する。また, [Pugeault 94]では, 仏語動詞の深層格割当てにおいて, 動詞の意味的クラス・格要素の名詞の粗い意味分類・前置詞などを手掛りとして用いている。

一方, ㉗では, 動詞のクラスとしてどのようなものを用意するかが問題となる。例えば, [大石 94]では, 表層格パターン間の言換え可能性によって日本語動詞のクラスを規定し, 各クラスに対して深層格フレームを用意している。ただし, 同一の格助詞が複数の意味で使われているため, 表層格パターンと深層格フレームとの対応に複数の可能性が生じてくるので, [森山 88]の連語論的分析に基づき, 格助詞の言換え可能性*2を利用して表層格パターンと深層格構造との対応の可能性を絞り込んでいる。例えば, 「ガ, ニ」および「ガ, ト」という二つの言換え可能な格助詞パターンを持つ動詞の場合は, この言換え可能性によって, 相互動作を表す動詞でありニ格とト格の深層格が“object”であると決定される。「ぶつかる」, 「関係する」, 「関連する」, 「結びつく」などがこのクラスに属する。また, [Webster 89]では, 深層格のうちの必須格および任意格の組によって動詞のクラスを規定し, 深層格割当てずみの用例を入力として動詞のクラスを決定している。

*2 英語において, 同様の言換え可能性(alternation)に基づき動詞を分類したものとして[Levin 93]がある。

2・2 語彙的意味論に基づくコーパスからの語彙知識獲得

語彙知識獲得においては、語の意味表現としてどのような形式を採用するかが重大な問題である。動詞については表層格や深層格のような一般的に認められる語彙知識が仮定できるが、名詞については一般的な同意はない。上位下位概念は名詞の基本的な意味関係と考えられるが、語の意味的な多義性や比喩的また換喩的用法などがこのような知識の抽出を困難にしている。

[Pustejovsky 91, Pustejovsky 93b]では、名詞の多義性や換喩的用法を個別の現象として分類するのではなく、共通の意味表現から生成的に派生させるという方法で名詞の意味表現を捉え、これを Generative Lexicon という名詞の意味表現形式として提案している。Generative Lexicon の基本的な表現は Qualia Structure と呼ばれ、次の4項目よりなる。

- constitutive role (構成役割, 機能的な意味)
- formal role (外見役割, 外面的な意味)
- telic role (目的役割, 何に用いるか)
- agentive role (主体役割, どのようにして生じるか)

例えば, “book” を表現する Qualia Structure は次のように記述される。

```

book(x, y)
CONST=information(y)
FORMAL=bound-pages(x) ∨ disk(x)
TELIC=read(T, w, y)
AGENTIVE=artifact(x) ∧ write(T, z, y)

```

“book” は、情報を表す y と外見を表す x の2項関係であり、 y は読まれることを目的とし、書かれることによって生まれる。 x は人工物である。 T は、「遷移(transition)」を表すイベントタイプ、 w と z は特に明示されていない動作主である。この表現により、“I read a book”, “The book is heavy”, “He finished a book”^{*3} の文に現れる “book” がどのような意味を指すかが適切に説明できる。

[Anick 90, Pustejovsky 92, Pustejovsky 93a]では、Generative Lexicon のような意味表現を仮定することにより、コーパスに現れるさまざまな表現(複合名詞、前置詞句を伴う名詞句、動詞との共起関係など)を観察することによって、名詞の Qualia Structure を徐々に獲得する考え方について述べられている。

動詞についても事前に設定された知識に基づいて、

* 3 彼が本を読む人か書く人かによって telic, agentive のいずれの読みも可能である。

* 4 head word, kernel word, key word などさまざまな呼び方をされている。

より詳細な意味情報を獲得することが考えられる。[Zernik 89] は、意味的に類似している動詞が同じようなバリエーションの構文パターン(これを category と呼んでいる)を持つことに注目し、未知の動詞がどの category に属するかを、(構文解析済み)コーパスに出現する構文パターンから推定する方法を示している。[大石 94] も格構造パターンのバリエーションや格助詞の置換えなどの可能性を事前に考慮しておくことによって、動詞の深層格の決定が行えることを示したといえる。

2・3 未知の語彙を動的に獲得するモデル

語彙的意味論に基づくコーパスからの語彙知識獲得と関連した研究として、テキストを解析中に未知の語彙が出現したときに、その語彙に関する知識を動的に獲得するモデルの研究がある [Jacobs 87, Martin 87, Wilensky 90, Zernik 87, Zernik 89]。例えば, [Wilensky 90] では、複数の用法を持つ語のある用法が未知のときに、これを他の類似した語の複数の用法から類推して獲得する手法が述べられている。また、システムが、機能的意味を持った名詞 “breakfast” について、冠詞がつく一般的な用法 “the breakfast” という用法は知っているが、冠詞がつかない “at breakfast” という用法を知らないという状況で “at breakfast” という表現に出合ったとする。このとき、機能的意味を持つ他の名詞 “class” の既知用法 “the class” や “at class” から類推することにより、システムは、“at breakfast” という用法における “breakfast” が「朝食を食べている間」という意味であることを学習する。その他、同様のモデルによって、比喩表現 [Martin 87]、慣用句 [Zernik 87] を扱ったものなどがある。

3. 辞書の定義文からの語彙的知識の獲得

語の知識を抽出する資源としての辞書の有用性は早くから認識され、機械可読辞書 (Machine Readable Dictionary: MRD) から語に関する知識を自動的に抽出する試みは種々行われてきた。辞書における語の定義は、見出し語の意味的な上位語 (genus term) およびそれを他の同じレベルの語から識別するための差異 (differentia) からなると考えられるので、その上位語^{*4} を取り出すことが重要である。

初期の研究として, [Amsler 81] では、機械可読な Merriam-Webster Pocket Dictionary の動詞と名詞の定義文から手作業で曖昧性を除去して上位語を抽出し、そこから語の階層構造を構築することの可能性に

ついて調査している。彼は、後の多くの研究の基本になる次のような問題点や方法論を指摘している。

1. 辞書では語の定義が自然言語で行われ、上位語の意味的な曖昧性が解消されていないため、もつれた (tangled) 階層構造しか得られないこと。
2. 階層構造の上位に位置する語 (cause, thing, class など) がループを形成する傾向があること。
3. “a type of…” の表現のように “of” の直前に現れるある種の語は、上位語ではなく、その直後の語が上位語になること。また、“of” を伴う *type*, *group*, *set* のような語を分類しておくことによって語の間の他の意味関係 (part-of など) が抽出できること。

[Chodorow 85] は、Webster 7th New Collegiate Dictionary を対象に、パターンマッチ手法によって定義文から上位語を取り出すためのヒューリスティクスを与えた。動詞については、“to” に続く単動詞を取ることによってほぼ 100% の精度で上位語が抽出できる。また、名詞については、前置詞や関係詞の直前の名詞句の主辞を選択する。ただし “of” を伴うある種のパターンについてはその直後の名詞句の主辞を選ぶ。これにより約 98% の精度で上位語を抽出可能であると報告している。LDOCE (Longman Dictionary of Contemporary English) を対象にして同様の研究が [Nakamura 88] によって、また、新明解国語辞典を対象にした研究が [Tsurumaru 86, 鶴丸 91] によって行われている。[Chodorow 85] では、何らかの意味情報 (語に与えられた荒い意味マーカや “er”, “man” などの語尾形態から得られる情報) を仮定して、語が多義性を持つかどうか判断する方法が示唆されているが、その他の研究では語の多義性を考慮していないため、得られる階層構造は性質の継承を保証するような真の IS-A 階層とはいえない。

[Guthrie 90] は、機械可読版の LDOCE が持っている box code (意味分類) と subject code (分野情報) を利用することによって、ある程度自動的に上位語の意味的多義性の解消を行うことができることを示している。

辞書の定義文から、階層構造だけでなく動詞の構文情報や格情報を取り出す研究もある。[Tomiura 92] では、動詞の格パターンと定義文が記述された辞書 (IPAL 辞書) から動詞の格関係を含めた意味的な階層関係を抽出する方法を提案している。動詞にはさまざまな用法の変化が見られる。特に動作動詞については、“Kim docked the boat.” と “The boat docked.” に見られるような表層格構造の変化 (causative-in-

choative alternation, 原因起動変化) などのさまざまな表層構造を持ち得る。[Sanfilippo 92, Sanfilippo 94] は、これらの可能な変化パターンの関係を事前に作成しておき、LDOCE および LLOCE に記述されている構文コードや意味コードを参照して、構文および意味情報を持った動詞の素性構造を半自動的に得る方法を提案している。

辞書からの語彙知識抽出は、現状ではまだ自然言語処理システムの性能を向上し得るだけの結果を与えていない。それにはいくつかの理由がある。[Ide 93] は、MRD からの語彙知識抽出の問題点を、「MRD は有用な情報を含むか」と「MRD からの情報抽出は簡単か」の 2 点について簡潔に述べている。前者については、現在の多くの MRD は人間用に作成された辞書の電子化版であり、上位語の選択に統一性がない。また、上位語として自然言語を用いざるを得ないため上位語のループを生じてしまう。さらに、上位概念以外の意味的な記述はきわめて貧弱などの問題点がある。後者については、[Chodorow 85] の研究などで簡単なパターンマッチ法によってかなりの精度で上位語が取り出されているように見えるが、実際に取り出された上位語は、階層構造を記述するための適切な上位語でない場合が多いし、上位語は往々にして意味的な多義性を含む。また、上位語以外の情報の記述には多様な言語表現が用いられており、言語解析が必要である。ただし、[Ide 93] は、辞書からの語彙知識抽出の研究がまったくむだであったとは結論していない。例えば、複数の辞書から、上位語を抽出することによって精度向上が得られたこと (五つの MRD から取り出された結果を合わせることによって、問題のある箇所が 55~70% から 5% に削減された) を述べている。複数の資源からの知識抽出を考えることは重要で、MRD からの知識抽出はコーパスからの知識抽出とは相補的な関係にあると考えるのが自然であろう。

MRD から上位語以外の意味情報を抽出する研究も行われているが、まだ具体的な成果は上がっていない。[Montemagni 92] は、上位語以外の意味情報を MRD から抽出するにはパターンマッチ法ではなく定義文の構文解析が必要であること、また、定義文の記述は多様であり、定義文用の文法を記述するよりは広い範囲を対象とする一般的な文法が必要であると述べている。

4. 対訳コーパスからの構造的言語知識の獲得

単言語コーパスと比べると、対訳コーパスから言語

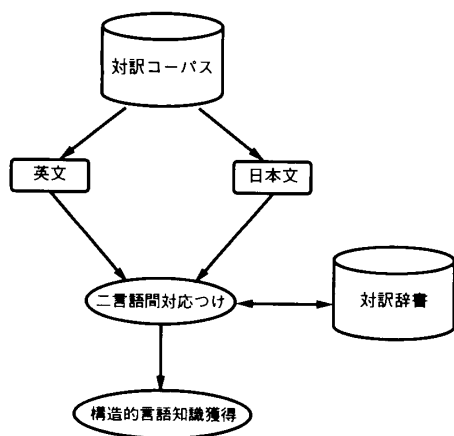


図1 対訳コーパスからの構造的言語知識獲得

知識を獲得することの利点・必要性としては、①二言語の情報を利用することにより、単言語だけでは解消不可能な統語的・語彙的曖昧性を絞り込むことができる、②機械翻訳での利用を目的としてコーパスから言語知識を獲得する場合は、対訳コーパスの利用が不可欠である、といった点があげられる。対訳コーパスからの構造的言語知識獲得の過程は、図1に示すように^{*5}、①対訳文の内部構造の対応づけ、②対訳文の内部構造を対応づけた結果からの構造的言語知識獲得、の二つに分けることができる^{*6}。

①の対訳文の内部構造の対応づけを扱った研究としては、[Grishman 94, Kaji 92, 北村 94, Matsumoto 93, 宇津呂 92]などがある。これらの手法の基本的考え方は、対訳辞書中の訳語の対応を利用して対訳文の統語構造中の部分どうしを対応づけるという点にある。ただし、これらの手法の間には、統語構造の表現形式として何をを用いるかなどに関する違いがあり、おおまかに以下のような分類をすることが可能である。

- (1) 一単語対一単語の対応だけを許して、素性構造で表現された表層格構造を二言語間で対応づける [宇津呂 92]。
- (2) 依存構造で表現された統語構造の部分どうしを二言語間で対応づける [Grishman 94, 北村 94, Matsumoto 93]。
- (3) 句構造木で表現された統語構造の句どうしを対応づける [Kaji 92]。
- (4) 依存構造を対応づける点は(2)と類似している

*5 ここでは、主として日本語と英語の対訳コーパスを対象として論文を紹介するので、図1中の対訳コーパスも、便宜上日英のものになっている。

*6 一般に対訳コーパスから言語的知識を獲得する際には、まず二言語間で文対応をとる必要があるが、ここでは、その過程は省略してある。対訳テキストの文対応については、本特集の工藤・井ノ上両氏の解説[工藤 95]を参照されたい。

英語: He is second to none in eloquence.
日本語: 弁舌では彼の右に出る者はいない。

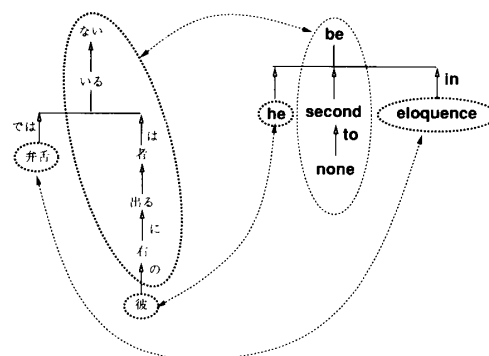


図2 対訳文の統語構造の対応づけ

が、人間による翻訳例と機械翻訳システムによる翻訳例の構文的不整合性(mismatch)から依存構造の対応部分を見つけ出し、正しい翻訳パターンを獲得する [Watanabe 93]。

ここでは、例として、[北村 94, Matsumoto 93]の依存構造照合の手法を紹介する。図2に、“he/彼”、“eloquence/弁舌”という訳語対応のもとで対訳文の統語構造中の部分どうしの対応がつけられた結果の例を示す。この方法では、①有用な訳語対応の情報が利用できる部分についてはできるだけ細かな対応を優先する、②有用な訳語対応の情報が利用できない部分についてはできるだけ粗い対応を優先する、という基準で対応結果の評価値を求め、評価値最大の対応結果を一意に求めている。これにより、二言語間で最も類似した統語構造が選択されるので、単言語での統語的曖昧性の解消も同時に行うことができる。和英辞典中の対訳例に対して行った実験では、64%の割合で正しい対応結果が一意に得られ、また、一意かどうかに限らず正しい対応結果を含む結果が得られる割合は89%であった。その他、[Grishman 94]では、1回の統語構造対応づけの結果から訳語の対応を抽出し、抽出された訳語対応をフィードバックさせてもう一度統語構造対応づけを行うという反復を繰り返すことにより、統語構造対応づけの精度を徐々に向上させていく手法が述べられている。

次に、②の構造的言語知識獲得を扱った研究を、獲得対象が何であるかという観点でおおまかに分類すると以下ようになる。

- (1) 翻訳規則を獲得するもの [Almuallim 94, Kaji 92, 北村 94, Tanaka 94, Watanabe 93]。
- (2) 対訳辞書の訳語対応に付加的情報を追加していくもの [Klavans 90]。

(2) 単言語の表層格フレームを獲得するもの[宇津呂 93].

(4) 訳語の組によって記述された対訳シソーラスを獲得するもの[浦本 93].

(1)の翻訳規則を獲得するもののうちで,[Kaji 92]では,句構造木を二言語間で対応づけた結果の一部を変数化することにより翻訳パターンを獲得している.また,[北村 94]では,日英の対訳例の依存構造を二言語間で対応づけた結果を蓄積し,日本語格要素のシソーラス中の位置によって日本語動詞の英訳語を選択する規則を獲得している.一方,[Almuallim 94, Tanaka 94]では,同様の動詞の訳語選択規則を,格要素の名詞のシソーラス中での位置を用いた決定木の形式で学習する手法を述べている.その他,(2)の[Klavans 90]では,英仏対訳辞書から得られる動詞の翻訳規則に対して,対訳コーパスから得られる詳細な語彙情報・頻度情報などを追加する手法が述べられている.

これらの研究は,いずれも機械翻訳に直接用いることを想定した研究であったが,それ以外にも,(3)および(4)のように単言語では困難であった語彙的多義性の解消に対訳の情報を利用する研究がある.これらの研究では,二言語間で語彙の意味的体系が異なるために,語彙的多義性の分布が二言語間で異なることを利用して,語彙的多義性の解消を行っている.例えば,[宇津呂 93]では,日英間で動詞の表層格構造を対応づけた結果から,日本語の動詞の表層格フレームを獲得する手法を述べている.その際には,英語側の情報を利用することにより,半自動的に日本語動詞の多義性を解消している.例えば,「買う」という動詞の場合は,“buy”, “incur”, “appreciate”といった英語の動詞の違いにより,それぞれ「物を購入する」,「恨み・反感などを招く」,「努力などを評価する」という意味の格フレームが獲得される.格要素の名詞の意味的制約の記述には,名詞シソーラス(分類語彙表[国研 93])中の意味カテゴリーを用いている.一方,[浦本 93]では,辞書

の定義文からのシソーラス構築において,対訳になっている定義文を利用することにより上位語の語義の曖昧性を解消しながら,訳語の組によって記述された対訳シソーラスを構築する手法を述べている.

5. おわりに

コーパスおよび辞書などの言語資源からの語彙知識獲得に関する研究のサーベイを行った.最後に,この分野の研究に関して注意すべき点をいくつかあげておきたい.

構文解析などが行われた括弧つきコーパスや品詞のタグつきコーパスは生コーパスに比べて入手が困難であるが,明らかにより有用である.解析済みコーパスを自動作成する統計的ツールの整備も重要である.

英語の辞書には,LDOCEのように意味カテゴリーが付与されたり限られた基本単語だけを用いて定義を記述するようなものや,COBUILD(Collins COBUILD English Language Dictionary)のように定義文が見出し語を使った完全な文で記述されているものなど,記述法が統一されたものがある.日本語の辞書では,このように統一された思想で書かれた辞書はないし,電子化された辞書の数も少ない.

現在のところ,自動獲得の対象は,動詞の格フレームや語の階層関係が中心である.それ以外の意味情報については,意味表現の枠組み自体を詳細に検討する必要がある.Generative Lexiconのように派生的に生成できる意味を別の項目と考えない捉え方は重要である.しかし,なお多義性や同形異義語の識別,比喩や熟語などの特殊な用法と通常の用法の識別などの問題は未解決である.

また,今後は,コーパスや辞書だけでなく,複数の資源を有機的に利用して語彙知識を抽出する研究が行われるだろう.

◇ 参 考 文 献 ◇

- [Almuallim 94] Almuallim, H., Akiba, Y., Yamazaki, T. and Kaneda, S.: Induction of Japanese-English Translation Rules from Ambiguous Examples and a Large Semantic Hierarchy, 人工知能学会誌, Vol. 9, No. 5, pp. 730-740(1994).
 [Amsler 81] Amsler, R.: A Taxonomy for English Nouns and Verbs, *Proc. 19th Annual Meeting of ACL*, p. 133-138(1981).
 [Anick 90] Anick, P. and Pustejovsky, J.: An Application

- of Lexical Semantics to Knowledge Acquisition from Corpora, *Proc. 13th COLING*, Vol. 2, pp. 7-12(1990).
 [Aone 93] Aone, C. and Mckee, D.: Acquiring Predicate-Argument Mapping Information from Multilingual Texts, *Proc. Workshop: Acquisition of Lexical Knowledge from Text*, pp. 107-116(1993).
 [Basili 93] Basili, R., Pazienza, M. T. and Velardi, P.: Hierarchical Clustering of Verbs, *Proc. Workshop: Acquisition of Lexical Knowledge from Text*, pp. 70-81(1993).

- [Brent 91] Brent, M. R.: Automatic Semantic Classification of Verbs from Their Syntactic Context: An Implemented Classifier for Stativity, *Proc. 5th ACL European Chapter*, pp. 222-226(1991).
- [Brent 93] Brent, M. R.: From Grammar to Lexicon: Unsupervised Learning of Lexical Syntax, *Computational Linguistics*, Vol. 19, No. 2, pp. 243-262(1993).
- [Chodorow 85] Chodorow, M., Byrd, R. and Heidorn, G.: Extracting Semantic Hierarchies from a Large On-Line Dictionary, *Proc. 23th Annual Meeting of ACL*, pp. 299-304(1985).
- [Fukumoto 94] Fukumoto, F. and Tsujii, J.: Automatic Recognition of Verbal Polysemy, *Proc. 15th COLING*, pp. 762-768(1994).
- [Grishman 94] Grishman, R.: Iterative Alignment of Syntactic Structures for a Bilingual Corpus, *Proc. 2nd Annual Workshop on Very Large Corpora (WVLC2)*, pp. 57-68(1994).
- [Guthrie 90] Guthrie, L., Slator, B. M., Wilks, Y. and Bruce, R.: Is there content in empty heads?, *Proc. 13th COLING*, Vol. 3, pp. 138-143(1990).
- [平岡 94] 平岡冠二, 松本裕治: コーパスからの動詞の格フレーム獲得と名詞のクラスタリング, *情処学研報*, Vol. 94, No. 78(94-NL-104) pp. 79-86(1994).
- [Ide 93] Ide, N. and Véronis, J.: Extracting Knowledge Bases from Machine-Readable Dictionaries: Have We Wasted Our Time?, *Proc. Int. Conf. on Knowledge Building and Knowledge Sharing*, pp. 257-266(1993).
- [Jackendoff 90] Jackendoff, R. S.: *Semantic Structures*, MIT Press(1990).
- [Jacobs 87] Jacobs, P. and Zernik, U.: Acquiring Lexical Knowledge from Text: A Case Study, *Proc. 5th AAAI*, pp. 739-744(1987).
- [Kaji 92] Kaji, H., Kida, Y. and Morimoto, Y.: Learning Translation Templates from Bilingual Text, *Proc. 14th COLING*, pp. 672-678(1992).
- [北村 94] 北村美恵子, 松本裕治: 二言語対訳コーパスからの翻訳知識の自動獲得, *信学技報*, Vol. 94, No. 5(NLC94-2), pp. 9-16(1994).
- [Klavans 90] Klavans, J. and Tzoukermann, E.: The BICORD System: Combining Lexical Information from Bilingual Corpora and Machine Readable Dictionaries, *Proc. 13th COLING*, Vol. 3, pp. 174-179(1990).
- [国研 93] 国立国語研究所: 分類語彙表, 秀英出版(1964, 1993).
- [工藤 95] 工藤育男, 井ノ上直己: コーパスに基づく共起知識の獲得とその応用, *人工知能学会誌*, Vol. 10, No. 2, pp. 205-212(1995).
- [Levin 93] Levin, B.: *English Verb Classes and Alternations*, The University of Chicago Press(1993).
- [Liu 93] Liu, R.-L. and Soo, V.-W.: An Empirical Study on Thematic Knowledge Acquisition based on Syntactic Clues and Heuristics, *Proc. 31th Annual Meeting of ACL*, pp. 243-250(1993).
- [Manning 93] Manning, C. D.: Automatic Acquisition of a Large Subcategorization Dictionary from Corpora, *Proc. 31th Annual Meeting of ACL*, pp. 235-242(1993).
- [Martin 87] Martin, J. H.: Understanding New Metaphors, *Proc. 10th IJCAI*, pp. 137-139(1987).
- [Matsumoto 93] Matsumoto, Y., Ishimoto, H. and Utsuro, T.: Structural Matching of Bilingual Texts, *Proc. 31th Annual Meeting of ACL*, pp. 23-30(1993).
- [Miller 90] Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K.: Five Papers on WordNet, CSL Report 43, Cognitive Science Laboratory, Princeton University (1990).
- [Montemagni 92] Montemagni, S. and Vanderwende, L.: Structural Patterns vs. String Patterns for Extracting Semantic Information from Dictionaries, *Proc. 14th COLING*, pp. 546-552(1992).
- [森山 88] 森山卓郎: 日本語動詞述語文の研究, 明治書院(1988).
- [Nakamura 88] Nakamura, J. and Nagao, M.: Extraction of Semantic Information from an Online English Dictionary and its Evaluation, *Proc. 12th COLING*, pp. 459-464(1988).
- [大石 94] 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, *情処学研報*, Vol. 94, No. 78(94-NL-104) pp. 71-78(1994).
- [Poznański 93] Poznański, V. and Sanfilippo, A.: Detecting Dependencies between Semantic Verb Subclasses and Subcategorization Frames in Text Corpora, *Proc. Workshop: Acquisition of Lexical Knowledge from Text*, pp. 82-94(1993).
- [Pugeault 94] Pugeault, F., Saint-Dizier, P. and Monteil, M.-G.: Knowledge Extraction from Texts: A Method for Extracting Predicate-argument Structures from Texts, *Proc. 15th COLING*, pp. 1039-1043(1994).
- [Pustejovsky 91] Pustejovsky, J.: The Generative Lexicon, *Computational Linguistics*, Vol. 17, No. 4, pp. 409-441(1991).
- [Pustejovsky 92] Pustejovsky, J.: The Acquisition of Lexical Semantic Knowledge from Large Corpora, *Proc. 5th DARPA Speech and Natural Language Workshop*, pp. 243-248(1992).
- [Pustejovsky 93a] Pustejovsky, J., Bergler, S. and Anick, P.: Lexical Semantic Techniques for Corpus Analysis, *Computational Linguistics*, Vol. 19, No. 2, pp. 331-358(1993).
- [Pustejovsky 93b] Pustejovsky, J. and Boguraev, B.: Lexical Knowledge Representation and Natural Language Processing, *Artif. Intell.*, Vol. 63, No. 1-2, pp. 193-223(1993).
- [Resnik 92] Resnik, P.: A Class-based Approach to Lexical Discovery, *Proc. 30th Annual Meeting of ACL*, pp. 327-329(1992) (Student session).
- [Sanfilippo 92] Sanfilippo, A. and Poznański, V.: The Acquisition of Lexical Knowledge from Combined Machine-Readable Dictionary Sources, *Proc. 3rd ANLP*, pp. 80-87(1992).
- [Sanfilippo 94] Sanfilippo, A.: Word Knowledge Acquisition, Lexicon Construction and Dictionary Compilation, *Proc. 15th COLING*, pp. 273-277(1994).
- [Tanaka 94] Tanaka, H.: Verbal Case Frame Acquisition from a Bilingual Corpus: Gradual Knowledge Acquisition, *Proc. 15th COLING*, pp. 727-731(1994).
- [Tomura 92] Tomura, Y., Nakamura, T., Hitaka, T. and Toshida, S.: Logical Form of Hierarchical Relation on Verbs and Extracting it from Definition Sentences in a Japanese Dictionary, *Proc. 14th COLING*, pp. 574-580(1992).
- [Tsurumaru 86] Tsurumaru, H., Hitaka, T. and Yoshida, S.: Automatic Thesaurus Construction from an Ordinary Japanese Language Dictionary, *Proc. 11th COLING*, pp. 445-447(1986).
- [鶴丸 91] 鶴丸弘昭, 竹下克典, 伊丹克企, 柳川俊英, 吉田 将: 国語辞典を用いたシソーラスの作成について, *情処学研報*, Vol. 91, No. 37(91-NL-83) pp. 121-128(1991).
- [内元 94] 内元清貴, 宇津呂武仁, 長尾 真: 動詞の語彙的知識獲得における類義語の用例を用いた多義性の類別, *情処学研報*, Vol. 94, No. 47(94-NL-101), pp. 105-112(1994).
- [浦本 93] 浦本直彦: 対訳コーパスからのパラレルシソーラ

- スの抽出, 人工知能学会第7回全大, pp. 463-466(1993).
- [Ushioda 93] Ushioda, A., Evans, D. A., Gibson, T. and Waibel, A.: The Automatic Acquisition of Frequencies of Verb Subcategorization Frames from Tagged Corpora, *Proc. Workshop: Acquisition of Lexical Knowledge from Text*, pp. 95-106(1993).
- [宇津呂 92] 宇津呂武仁, 松本裕治, 長尾 眞: 日英対訳文間の素性構造照合による統語的曖昧性の解消, *情処学論*, Vol. 33, No. 12, pp. 1555-1564(1992).
- [宇津呂 93] 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, *情処学論*, Vol. 34, No. 5, pp. 913-924(1993).
- [Watanabe 93] Watanabe, H.: A Method for Extracting Translation Patterns from Translation Examples, *Proc. 5th TMI*, p. 292-301(1993).
- [Webster 89] Webster, M. and Marcus, M.: Automatic Acquisition of the Lexical Semantics of Verbs from Sentences Frames, *Proc. 27th Annual Meeting of ACL*, pp. 177-184(1989).
- [Wilensky 90] Wilensky, R.: Extending the Lexicon by Exploiting Subregularities, *Proc. 3rd DARPA Speech and Natural Language Workshop*, pp. 365-370(1990).
- [Zernik 87] Zernik, U. and Dyer, M. G.: The Self-Extending Phrasal Lexicon, *Computational Linguistics*, Vol. 13, No. 3-4, pp. 308-327(1987).
- [Zernik 89] Zernik, U.: Lexicon Acquisition: Learning from Corpus by Capitalizing on Lexical Categories, *Proc. 11th IJCAI*, pp. 1556-1562(1989).
- [Zernik 90] Zernik, U. and Jacobs, P.: Tagging for Learning: Collecting Thematic Relations from Corpus, *Proc. 13th COLING*, Vol. 1, pp. 34-39(1990).
- [Zernik 91] Zernik, U.: Introduction, Zernik, U. (ed.), *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, pp. 1-26, Lawrence Erlbaum Associates(1991).

 著 者 紹 介



宇津呂 武仁(正会員)

1989年京都大学工学部電気工学第二学科卒業。1994年同大学院博士課程修了。工学博士。同年、奈良先端科学技術大学院大学助手。現在に至る。自然言語処理の研究に従事。情報処理学会、日本ソフトウェア科学会、言語処理学会、ACL各会員。



松本 裕治(正会員)

1977年京都大学工学部情報工学科卒業。1979年同大学院工学研究科修士課程修了。同年、電子技術総合研究所入所。1984～85年英国インペリアルカレッジ客員研究員。1985～87年(財)新世代コンピュータ技術開発機構に出向。京都大学大型計算機センター助教授。京都大学工学部助教授を経て、1993年奈良先端科学技術大学院大学情報科学研究科教授。現在に至る。工学博士。自然言語処理、理論プログラミングに興味を持つ。情報処理学会、日本ソフトウェア科学会、言語処理学会、AAAI, ACL, ACM各会員。