

言葉の意味を計算機で扱う

黒橋 禎夫

京都大学大学院 情報学研究科

kuro@i.kyoto-u.ac.jp

1 意味とは

辞書で「意味」という語を引くと、ほぼどの辞書でも次のような3つの語義が与えられている。

1. 言葉・行動などが直接的に表す内容。
2. 言葉・行動などが間接的に表す(そこに込められた, そこから汲み取れる, そのうらにある)内容。
3. 価値, 重要性。

ここではこの1と2の語義を対象として議論をすすめる¹。

「言葉の意味を計算機で扱う」とは, この2つの語義でいうところの意味を計算機によって計算する(求める)ことである。もちろんこの2つの語義は明確に区別できるものではないが, その違いを意識しておくことは大事である。ここでは, 語義1に関する処理を直接的意味処理, 語義2に関する処理を間接的意味処理とよぶことにする。

より具体的には, 直接的意味処理, 間接的意味処理にはそれぞれ次のような処理が考えられる。

直接的意味処理 言葉が直接的に表す内容の計算。

1. 語の意味の明確化
例) bank 銀行か土手か
2. 語の関係の明確化
例) 私の手帳 私の持っている手帳

間接的意味処理 言葉が間接的に表す内容の計算。

1. 語からの連想
例) 本 読み書き
2. 言い換え関係の推測
例) 太郎から本を借りた 太郎が本を貸した
3. 因果関係の推測
例) 皿が落ちる 皿がわるる

¹3番目の語義について考えてみることも大きな意味があるが, ここではとりあえず対象外としておく(3番目の語義はこの下線の「意味」の用法)。

4. 話者の意図の推測

例) それは塩ですか 塩をとってください

人間のテキストの理解やコミュニケーションがこのような「意味」に大きく依存していることは間違いなく, これらが今後の言語情報処理において中心的課題になっていくと考えられる。

このような意味処理の特徴は, 統語処理の問題, たとえば形態素解析, 構文解析などと比較して, 次の2点であると考えられる。

- 問題設定(問題の正解の設定など)が明確でない。
- どのような知識をどうやって用意するかが最大の問題である。

以下ではこの2点を意識しながら議論を進める。

2 従来の研究

まず, これまでの意味処理の研究を概観する。

2.1 直接的意味処理について

語の意味の明確化の問題はこれまで盛んに研究されてきた(最近ではこの問題を対象としたコンテストもある[1])。これは, まず各語の複数の意味(語の多義性)を何らかの基準で定義しておき, 与えられた文・文章において各語の意味を決定するという問題である。語の多義性の定義によって次のように大別することができる。1の場合には誰の目にも正解は明らかであるが, 2の場合には人間でも判断がゆれる, あるいは語の多義性の設定が細かすぎるといった場合がある。

1. 正解がある程度明確な場合

例) かな漢字変換: せんたく 選択, 洗濯

フランス語等の発音記号の復元:

cote côte(海岸), côté(側)

2. 正解があまり明確でない場合

例) 辞書の語義: interest interest₁, interest₂
シソーラスの意味クラス: WordNet の synset
翻訳における訳語選択: wear 着る, はく

これらの意味の明確化は, いずれかの意味と強く共起するような手がかり (表現) によって行われる. たとえば「洗剤でせんたく」とあれば「洗濯」と推測できる. 問題は, このような手がかりをどのように集めるかという点にある.

1 の場合には, 手がかりの学習データを比較的容易に大量に集めることができる. 例えば, かな漢字変換では, 日本語のテキストの漢字部分を読み直せば, かな漢字対応とその文脈という学習データを得ることができる. 同様に, フランス語等の発音記号の問題でも, 発音記号のついた正規のテキストが学習データとなる [2].

一方, 2 の場合は学習データの収集はそれほど簡単ではない. 辞書の語義の区別を問題とする場合には, 各語義の定義文や用例中の語が手がかりとなるが, それだけでは十分な手がかりとはいえない [3].

シソーラスを意味クラスを定義したものと考え, 語がシソーラス中の複数の意味クラスに登録されている場合に, ある文脈でのその語の出現がどの意味クラスのものであるかを明らかにするという問題設定も考えられる. この場合, コーパス中の各語をシソーラスの意味クラスに汎化することにより, 意味クラスとしての共起情報を学習することができる [4]. ある語について見ると意味クラスへの汎化は一意には定まらないが (その解決がこの問題設定であるが), 中には対応する意味クラスが一つしかない (多義でない) 語もあり, 全体として見れば意味クラスの共起をある程度知ることができるという点がポイントである.

訳語選択の場合, 例えば名詞と動詞のペア (動詞句) の訳し分けでは, 可能な組み合わせをすべて作ってみて, 目的言語のコーパスにおいてそのような組み合わせが実際に出現しているかどうかを調べるという方法も提案されている [5].

他の方法としては, とにかく正解コーパスを作ってしまうということも考えられる. たとえば, コーパス中の各語に WordNet の synset 番号を付与した SemCor コーパスは意味処理の学習・テストコーパスとしてよく利用されている [6]. また, Yarowsky は, 1 つのテキスト中では語の意味は一定であるという経験則を利用し, まず人間が各意味の典型的な使用例をコーパス中で指定し, そこから徐々に手がかりを学習していく

という方法を提案している [7].

語の関係の明確化の問題については, 用言と体言の間の Agent, Object, Experiencer などの関係を議論した Fillmore の格文法が代表的なものである [8]. 日本語では, 語順の入れ替り, 省略などの問題に対応するために特に機械翻訳の研究で格文法の考え方が早くから取り入れられてきた. しかし, どのような関係を設定するかについてはコンセンサスの得られるようなものは存在しないといってよい.

NTT は日本語語彙大系という, 木構造にまとめられた約 3,000 の意味素とそれを用いて記述した用言の構文パターンの辞書を作成している [9]. これは日英翻訳において体言との関係に基づいて用言の訳語を選択することを主目的としたものであるが, 大規模な意味素システムによる処理を実現したものとして注目されている. このような大規模な人工システムの保守・拡張が今後どの程度可能であるのか, また, 現在のコーパス等に恵まれた時代にこれに近い精度のものが自動的に学習可能であるかどうか, などの点が興味深い.

語の関係の明確化の問題でもう一つよく議論されてきたものとして, 日本語の「A の B」の形の名詞句における A と B の意味関係の問題がある. これに所有, 対象, 目的など, 関係のラベルを付与するという問題設定で多数の研究が行われてきたが, 格文法の場合と同様に, コンセンサスの得られるような関係のセットは存在しないといってよい.

一方, やはり日英翻訳において名詞句「A の B」をどのように訳すか (B of A, B in A, A' B, AB など) という問題がある. これについて隅田らは大量の翻訳例 (日本語の名詞句とその翻訳のペア) を収集し, 新たな名詞句の翻訳は, それと似た用例を見つけてその翻訳に従うという方法を提案した [10]. この方法は用例に基づく翻訳の実現例としても注目をあつめた.

2.2 間接的意味処理について

間接的意味処理では, 直接的意味処理の場合とは異なり, ある処理をある程度実現したとしても, それが自然言語処理アプリケーション等の改善に直接つながることは現時点では少ない. そのため, 処理の実現というよりは, ある理論または方法論によってどのような知識がどの程度集積可能であるかという点に重点がおかれて研究が進められている.

このような問題で最もよく参照されるのは Cyc プロジェクトであろう [11]. 世界の常識を, 基本的に人

手で、形式言語の上で書き下すことを目的に 1984 年に始まったが、「できた」という報告はない。

これに対して、MindNet プロジェクトでは常識は既存の(人間用の)辞書の中にはほぼ与えられていると考え、英語の複数の辞書を構文解析して辞書中の同一語をリンク付けした知識ベースを構築している [12]。現在のところ、PP-Attachment の問題への応用などの報告がある。

Fillmore らによる FrameNet プロジェクトは、コーパス中の用例を重視する立場で、数十の用言カテゴリに着目し、それと関係する項(例えば「貸し借り」に対して貸手、借手など)を分類して、コーパス中でタグ付けしていくという作業を行っている [13]。関係タグの設計は議論を重ねれば重ねるほど詳細化しているようで、格文法と同様に関係を人工的に分類することに難しさがあると思われるが、今後の動向が注目される。

一方、Pustejovsky らは、語/文のダイナミックな解釈をモデル化するものとして Generative Lexicon という理論を提案している [14]。これは、例えば「本」という名詞の意味構造の中に「読み書き」のようなことが書かれており、これが enjoy a book の解釈(enjoy reading a book)を助けていると考える。このような議論を形式言語の上で緻密に行っているが、一方で WWW テキストからの語の意味構造の学習についても研究しており、非常に興味深い。

3 我々の研究：言葉で言葉を扱う

以下では我々の最近の意味処理の研究を紹介したい。我々の意味処理に対する立場は「知識の取り扱いを、人工的言語(形式言語)ではなく自然言語そのもので行う」というものである。すなわち、

- 人工的カテゴリを導入しない²。すなわち、格解析における Agent, Object, Experiencer など、名詞句「A の B」の解析における所有、対象、目的などといったラベルを用いない。もちろん計算機が自動的にクラスタリング、カテゴリ化を行うことはよいが、人間が導入することは行わない。

²ただし、語間の類似度や意味クラスが必要となる場合、現時点では人手で作られたシソーラス(分類語彙表、WordNet など)を利用しており、これは人工的カテゴリの導入ということになる。これは、1) 一度にすべてを自動学習的に行うことは難しい、2) 語の上位下位/同義類議の関係は人間が定義しても(詳細は別として大まかには)比較的安定である、などの理由による。しかし、将来的にはシソーラスもコーパス等から自動学習すべきであると考えている。

- 人工的言語(形式言語)ではなく、自然言語を知識の記述形式と考える = 自然言語で記述された既存の一般のテキストを知識源として利用する。

- 知識の操作も言い換え(paraphrase)などの自然言語上の操作によって行う。

知識の取り扱いを形式言語で行うことは、曖昧性がない点、(世界が無矛盾であれば)形式的操作によって推論が行える点などが利点であろう。しかし、ある小さなドメインに限るなら別として、この世界全体を対象として満足できる形式言語を設計し、それで知識を書き下していくことが本当にできるのかという問題があり、実際上それらはほとんど不可能であると思われる。

一方、自然言語で知識を取り扱うならば、もちろん設計は必要なく、この世界を記述することも可能である(すでに人間は世界について様々なことをテキストとして記述してきた)。自然言語処理の技術が未熟であった時代には、自然言語で何かが書いてあっても計算機にとっては意味がないという状況だったが、現在では構文解析レベルで歯が立たないということはない。そして意味解析レベルの問題はこれからその知識を使って解決していく問題である。推論の問題についても、形式言語が仮定していたのは無矛盾な世界であるが、実際の世界はそうではないので形式言語に advantage があるとは考えられない。

このようなことから、我々は上記のように「知識の取り扱いを自然言語で行う」という立場をとっている。このアプローチの最大の問題は、やはり「自然言語で書かれた知識を本当に十分に使いこなせるか」という点にあるだろう。これを実現するために我々が行っているいくつかの研究を紹介したい。

3.1 コーパスからの格フレームの獲得

日本語の場合には、格要素の語順の入れ替え、省略、提題助詞/連体修飾などに関連して格助詞が明示されない場合などがあるため、構文解析を行うだけでは文の解析としてまったく不十分であり、格解析を行ってはじめて文の解析が一段落するといつてよい。

そこで、格フレーム辞書をどのように構築するかということが問題となる。さきほど述べた NTT の日本語語彙体系はこれを人手で構築したものであるが、現実のデータを過不足なく説明できるかどうか、さらに作成・保守のコストなどを考えるとテキストからの自

動学習の方法を検討すべきであることは明らかである。

一つの方法は、構文情報の付与されたコーパスを用いる方法である [15]。しかし、格フレームは名詞と動詞の組み合わせを記述したものであるから、その学習には膨大な学習データが必要となり、いわゆるタグ付きコーパスの量はそれに対して十分とはいえない。

そこで我々は、これまでに開発してきた構文解析システム KNP を用いて大量のコーパスを構文解析し、その結果から格フレームを学習するというを行った [16]。このアプローチの場合には、構文的曖昧性 (KNP の解析エラー) と、意味的曖昧性 (用言の意味がひとつでなく、それによって格フレームのパターンもかわりえる) をどのように扱うかが問題となる。前者については、KNP の構文解析において信頼度の低い部分は用いない、ある用言にとって頻度が低い格要素は削除する、などの対処を行った。また後者については、格フレームの変化は用言の直前の格要素に依存すると考えられるので、直前の格要素と用言のペアについて、たとえば「友達になる」「中心になる」などについてそれぞれ格フレームを学習し、必要であればあとからクラスタリングを行うという手法をとった。

このような方法により、毎日新聞 7 年分の 360 万文から約 50,000 用言について格フレームを学習した。学習した格フレームの例を以下に示す。

友達になる

ガ: ヘビ, カッパ, 動物

ト: 女の子, 留学生, 人, 子供, 少年, 辞書, 生徒, 生物, サメ, バッハ, カエル, 自然

中心になる

ガ: 問題, 議員, 会, 日本, 氏, 人, メンバー, こと, 委員会, 者, 協力, 弁護士, 部, 活動

デ: 選, 会談, 協議, 面, 聴取, 問題, 中, 内, 作成, 日本, 会, 観, 部門, 対策, 台湾, 北部

政治家になる

ガ: 女性, 女房, 軍人, マーフィー, 氏, 者, 人

この辞書を用いて、KNP を拡張し、実用規模の格解析システムを構築した。これにより、たとえば「どのような政党でも政権を担える」「ドイツ語も話す先生」で「政党」がガ格、「ドイツ語」がヲ格、「先生」がガ格であることを正しく求められるようになった。

この格フレーム辞書は、文の格解析だけでなく省略、照応の解析にも不可欠であり、言語処理の多くの場面で利用される非常に重要なツールになると考えている。

3.2 名詞句「A の B」の解釈

「A の B」の形の名詞句の意味解釈は日本語の意味処理において最も困難な課題であった。従来の方法では、まず意味関係のセットを設定し、それによって関係を説明するというを行っていた。例えば、次のようなものである。

私の車	所有
机のあし	全体部分
グレーの制服	様態
ラグビーのコーチ	対象
野球の選手	範疇
風邪のウイルス	結果
旅行の準備	目的
トランプの手品	道具
...	

しかし、どのように関係セットを設計しても、複数に属する、あるいはどれにも属さないと思われるような名詞句の例が際限なく現れ、納得できる関係セットを得ることは非常に難しい状況であった。

これに対して、我々は、これらの関係は後者の名詞 (B) のもつ意味構造のスロットを前者の名詞 (A) が埋めるという関係であり、さらに名詞の意味構造は多くの場合国語辞書に記述されているということを示した [17] (実験では例解小学国語辞典 [18] を用いた)。すなわち、名詞句「A の B」の解釈は B の語義文中の適切な名詞に A を対応付けることであるとした (対応付けはシソーラスで計算される語間の類似度を基準とする)。たとえば、「ラグビーのコーチ」の解釈は「ラグビー」を「コーチ」の定義文中の「スポーツ」に対応付けることであるとする。「風邪のウイルス」でも同様である。

「ラグビーのコーチ」

コーチ スポーツ で、そのやり方などを教えること。また、その人。

「風邪のウイルス」

ウイルス インフルエンザなどの 病気 をおこす、ひじょうに小さな生物。

この解析結果から B の定義文中の対応部分を A で置き換えることにより名詞句の言い換えを得ることが

できる。たとえば上の例では「ラグビーのコーチ」が「ラグビーで、そのやり方などを教えること。また、その人。」と言い換えられる。これらの結果は、人々が共通して納得できるものであると述べている。また、どうしても関係セットを導入したい場合や、この結果を他言語への翻訳の規則に対応させたい場合は、この言い換えられた（詳細化された）表現をもとにクラストリングを行ったり、用例間のマッチングを行うことが可能である。

3.3 国語辞典を用いた常識問題の推論

MindNet の研究でも指摘されているように、辞典には非常に豊富な情報が与えられており、これを十分に使いこなすことができれば、いわゆる「常識」をかなりの程度理解できる言語処理システムを実現することができるだろう。

一つの方法は、辞書中の同一語をリンクでつなぎ、その間の活性伝搬のような処理によって「連想」のような処理を実現することである。しかし、そのような方法では、そこからさらに次のステップに進むための知見を得ることは難しい。そこで、もっと明示的に、辞書内の知識を推論のステップのように積み重ねていくことにより、どの程度常識に関する問題に答えられるかということを考える [19]。

常識に関する問題として、たとえば

時計で？がわかる。

堤防が崩れると？があふれる。

というような問題を考えると、辞書の記述をもとにしてこれらに回答するためには次のような知識の利用が必要となる。

時計 時刻を知るための器械

↓ (辞書定義文の変換)

時計で時刻を知る

↓ (Xを知る = Xがわかる)

時計で 時刻 がわかる

堤防 川や海があふれないように築いたもの

↓ (辞書定義文の変換)

川や海があふれないように堤防を築く

↓ (Xのために Y → \bar{Y} だと \bar{X} ³)

堤防を築く と川や海があふれる

³ \bar{X} は X の否定。

↓ (築く ↔ 崩れる)

堤防が崩れると 川や海があふれる

現在、例解小学国語辞典の定義文と用例、同義語、対義語などの情報によって上記のように常識の問題に答えるシステムを作成し、種々の検討を行っている [20]。

3.4 辞書の記述の構造化

前節までの研究では、辞書の定義文・用例を構文・格解析したものを知識源としており、ある意味で辞書の記述をすべて平等に扱っていた。しかし、人間の知識の把握は、もっと構造化された、メリハリのついたものであると考えられる。

そこで、辞書の定義文の記述を共通特徴、共通変数、個別特徴という3つの部分に自動的に構造化することを考えた [21]。たとえば、次のようないくつかの生き物の定義文を考える。

鮎 きれいな 川に すむ、姿の美しい 魚。

ゴリラ アフリカの 森林に すむ、体が大きなサルのような 動物。

ほととぎす 夏鳥として渡ってきて、森林に すむ 鳥。

これらの定義文をみると「すむ」という表現がすべての定義に現れており、これがこれらの生き物の共通の特徴を示していると解釈できる。また、「どこに」すむかと、その生物の種別が各定義に現れており、その具体的値（「川」「森林」など）は同じではないが、そういう共通のスロット（変数）があるということがわかる。さらに、残った個別の部分、例えば「ほととぎす」の「夏鳥として渡ってきて」という表現は「ほととぎす」に固有の特徴である。このように、定義文の記述の共通性をみれば、その定義された語（概念）間の特徴の区別（構造化）が可能である。

問題は、このような記述の共通性をどのような尺度によって取り出すかという点にある。辞書の定義文全体を対象とすると、様々な頻度で、様々な大きさの記述が共通しており、さらに上の「川」「森林」のようにある意味クラスの語というような共通性まで考慮すると、何をどのような尺度で共通と考えるかが大きな問題となる。

この尺度として、我々はデータ圧縮などでよく利用される情報理論の尺度である MDL (Minimum Description Length) 原理を用いることを考えた。すなわ

ち, MDL という基準によって定義文集合 (= 構文木集合) の「良い」圧縮を行い, シソーラス上の意味クラスに汎化されて圧縮された部分を共通変数, 普通に圧縮された部分を共通特徴, 残りの部分を個別特徴と解釈することとした.

このような方法により, 例解小学国語辞典の定義文集合の構造化を行ったところ, 文節単位で約 11% が共通特徴, 約 58% が共通変数となる構造化が行われた. 上述の例はその具体例である. 現在この構造化の結果を検討中であるが, その結果は多くの場合, 人間の直感にあうものとなっている.

このように, 自然言語テキストをその内部の性質 (表現の繰り返しなど) によって自動的に構造化できれば, 自然言語テキストを知識源とするアプローチの可能性が大きく広がるものと考えられる.

3.5 対話的ヘルプシステム, 自動リファレンスサービスシステム

これまでに述べてきたような研究が進展し, 自然言語で書かれた内容を計算機がそのまま知識源として利用できることになれば, 自然言語処理アプリケーションには (SF の世界のような) 大きな可能性が生まれる. 例えば, これまで単にキーワードのマッチング対象でしかなかった検索対象テキストが, そこでは知識源となるのであるから, 世界中に広がる WWW テキストに対して, それが知識として消化・理解された上での検索ができるようになる.

そうすればその次の段階として, 知識源 (= テキスト) を背景として, 人間の質問に対話的に答えるようなシステムが可能になると考えられる. このような考えから, 自然言語テキストを知識源としてどのような対話ができるかという点に主眼を置いて, 京都大学総合情報メディアセンターの対話的ヘルプシステムの研究を行っている [22]. このシステムは, メディアセンターの計算機環境, アプリケーションの使用法などについてキーボード入力による対話を行うことができるシステムである.

現在のところ, まだ, 自由に書いた, すでに存在するテキストを十分に使いこなす技術があるわけではないので, 「~する場合は~する」「~するためには~する」というような if-then 形式に制限して知識を記述している.

システムは, ユーザの質問と上記のような知識とのマッチングを行い, 質問に該当する知識を見つけ, そ

れをユーザに提示する. さらに, 質問を知識に照らして解釈することにより,

1. 相手の発話の中に自分の知らない言葉があれば聞き返す.
2. 相手の発話を文脈にそって解釈する.
3. 相手の発話が曖昧な場合 (自分にとって複数解釈可能な場合) 聞き返す.

というような, 対話における最も基本的なやりとりを実現している.

このシステムは 99 年 7 月からメディアセンターのホームページを通して公開しており, これまでにすでに 2000 件の質問を受け付け, 約 40% に適切な回答を行っている. センター利用者の質問は多岐にわたり, TA (ティーチングアシスタント) でも質問に答えられない場合がある. システムは適合する質問がない場合「(知識がないので) その質問にはこたえられません」と発話するが, これも失敗と数えた上での約 40% の回答成功率は決して低くないと考えている.

以下に実際の対話例を示す.

<Thu Aug 19 21:20:52 1999>

U1: mule でメールを読むにはどうすればいいですか.

S1: “mh-e” “mew” からメールソフトを選択してください.

U2: mh-e を使います.

S2: まだ mh-e を起動していない場合は, mule 上で “M-x mh-rmail” として下さい. 新しくメールを取り込みたい場合は “i” を押してください.

<Thu Sep 16 22:32:20 1999>

U1: UNIX で, Mew を使うにはどうすればいいのかわか.

S1: Mule を起動してから mule 上で “M-x mew” としてください. 具体的な使い方は “http://www.mew.org/” を参照してください.

U2: メールを消去するにはどうすればいいのかわか.

S2: Summary モードで “d” を押して ‘D’ マークを付けた後に, 実行の “x” を押してください.

最初の例では, 知識ベースにメールを読む方法が二つあるので, それを聞き返している. 次の例では, ユー

ザの2番目の質問を「Mewを使っている」という文脈で正しく解釈している。

さらにこのシステムの枠組みをそのまま用いて、京都大学附属図書館の自動リファレンスサービスシステムを構築している[23]。ここでは、リファレンスサービス用の知識と図書館分類のデータベースを用意することにより、「江戸時代の農民一揆に関する本を探したい」というような質問に対して「歴史・地理の本は図書館2階のGの書架にあります」と回答することを実現している。このシステムも近々サービスを開始する予定である。

3.6 おわりに

本稿では「言葉の意味を計算機で扱う」ということについて簡単に問題の整理を行った後、まず従来の研究を概観し、次に我々の最近の意味に関連する研究を紹介した。

言語情報処理をとりまく環境はこの10年間で激変した。超高速の計算機、ギガ単位の記憶容量、世界中に分散した巨大なコーパスや辞書が利用可能となった。これにともない、いよいよ本当の言語情報処理、すなわち意味を取り扱う言語情報処理ができる時代になった。

我々は、意味の取り扱いに breakthrough をもたらす最大のポイントは「知識を言語で表し、それを十分に扱う技術を成熟させること」であると考えている。我々の一連の研究はこの考え方を追求したものであり、本稿ではその有望性のある程度示せたのではないかと考えている。

もちろん、意味に対するアプローチはこのようなものだけではない。様々なアプローチの間の建設的な議論が「言葉の意味を計算機で扱う」ことを可能にする信じている。

謝辞

日頃より刺激的な議論をしていただいている京都大学大学院情報学研究科の松山隆司先生、河原達也先生、杉本晃宏先生に感謝いたします。また、本稿に様々なコメントをいただきました北陸先端科学技術大学院大学の佐藤理史先生、通信総合研究所の内元清貴氏に感謝いたします。

参考文献

- [1] A. Kilgarriff. SENSEVAL: An Exercise in Evaluating Word Sense Disambiguation Programs. In *Proc. of LREC*, Granada, pp.581-588, 1998.
- [2] D. Yarowsky. Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French. In *Proc. of the 32nd Annual Meeting of ACL*, pp.88-95, 1994.
- [3] M. Lesk. Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. In *Proc. of ACM SIGDOC Conference*, 1986.
- [4] D. Yarowsky. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. In *Proc. of COLING'92*, 1992.
- [5] 野美山浩. 目的言語の知識を用いた訳語選択とその学習性. 情報処理学会 自然言語処理研究会 86-2, 1991.
- [6] G. Miller, C. Leacock, and R. Teng. A semantic concordance. In *Proc. of the ARPA Human Language Technology Workshop*, pp.303-308, 1993.
- [7] D. Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods. In *Proc. of the 33rd Annual Meeting of ACL*, pp.189-196, 1995.
- [8] C. J. Fillmore (田中春美, 船城道雄訳). 格文法の原理. 三省堂, 1975.
- [9] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林編. 日本語語彙大系. 岩波書店, 1997.
- [10] E. Sumita, H. Iida and H. Kohyama. Translating with Examples: A New Approach to Machine Translation. *Proc. of 3rd TMI*, 1990.
- [11] D. B. Lenat and R. V. Guha. *Building Large Knowledge-Based Systems; Representation and Inference in the Cyc Project*. Addison Wesley Publishing Company, Inc., 1989.
- [12] S. D. Richardson, W. B. Dolan and L. Vanderwende. MindNet: acquiring and structuring semantic information from text. In *Proc. of COLING-ACL'98*, 1998.

- [13] C. F. Baker, C. J. Fillmore, and J. B. Lowe. The Berkeley FrameNet Project. In *Proc. of COLING-ACL'98*, pp.86–90, 1998.
- [14] J. Pustejovsky. *The Generative Lexicon*. The MIT Press, 1995.
- [15] 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 自然言語処理研究会 119-11, pp.69–76, 1997.
- [16] 河原大輔, 鍛冶伸裕, 黒橋禎夫. 大規模コーパスからの格フレーム辞書構築とそれを用いた格解析. 言語処理学会 第6回年次大会, 2000.
- [17] S. Kurohashi and Y. Sakai. Semantic Analysis of Japanese Noun Phrases : A New Approach to Dictionary-Based Understanding. In *Proc. of the 37th Annual Meeting of ACL*, pp.481–488, 1999.
- [18] 田近洵一 編. 例解小学国語辞典. 三省堂, 1997.
- [19] B. Katz. Using English for Indexing and Retrieving, *Artificial Intelligence at MIT*. Vol.1, MIT Press, pp.134–165, 1990.
- [20] 黒橋禎夫, 酒井康行. 辞書とコーパスからの世界知識の自動抽出. 「知識発見のための自然言語処理」シンポジウム, 1999.
- [21] M. Tsuchiya and S. Kurohashi. Automatic Discovery of Definition Patterns Based on the MDL Principle (poster). In *Proc. of the Second International Conference, Discovery Science '99 (Lecture Notes in Artificial Intelligence 1721)*, pp.364–365, 1999.
- [22] 日笠亘, 藤井綱貴, 黒橋禎夫. 入力質問と知識表現の柔軟なマッチングによる対話的ヘルプシステムの構築. 情報処理学会研究会 自然言語処理研究会 134-14, 1999.
- [23] 平田大志, 日笠亘, 藤井綱貴, 黒橋禎夫. 図書館の自動リファレンス・サービス・システムの構築. 言語処理学会 第6回年次大会, 2000.