

知的電子化英和辞書

-前置詞の曖昧性解消システムの実装-

永井野 亮 佐川 雄二 杉江 昇
名城大学大学院 理工学研究科

[c3002020@ccmailg](mailto:c3002020@ccmailg.meijo-u.ac.jp), [sagawa@ccmfs](mailto:sagawa@ccmfs.meijo-u.ac.jp), [sugie@ccmfs](mailto:sugie@ccmfs.meijo-u.ac.jp) } .meijo-u.ac.jp

あらまし ここ数年、計算機の普及に伴い、様々な英和辞書が電子化されている。電子化されることにより、単語の検索時間を短縮し、英文を読む際の手間を省く効果が現れている。そこで我々は、英語学習環境下で使用する英和辞書として、電子化英和辞書に、英語学習時に役立つ知的な機能を付加したシステムを作成することにより英語学習の効率が上がる事ができると考えた。その機能の一つとして、英語学習初心者が辞書を引く際に難関となる「多義性を持つ英単語からの文に合う意味の選択」を支援する機能を提供することを目的とする。本稿では、多くの多義性を持ち、初心者にとって分かりづらい語として前置詞を取り上げ、その多義曖昧性を解消するための機能について、その考え方からシステムの実装について報告する。

The Intelligence Dictionary for English to Japanese

- Implementation of a system to disambiguate prepositions -

Ryo Nagaino Yuji Sagawa Noboru Sugie
Graduate School of Science & Technology, Meijo University

Abstract In recent years, a English-Japanese dictionary is implemented as software or electronic dictionary, because the performance of a computer goes up. Therefore an electronic dictionary reduces efforts to look up word entry. So we thought that we add some intelligent functions to an English-Japanese dictionary. Then this system helps some people with learning English. One of those functions is to disambiguate a word. The purpose of this paper is how to disambiguate a word and to implement that function.

1. はじめに

近年の企業や個人活動のグローバル化により、国内における海外への関心が高まっている。さらに、PCの普及やネットワークの

整備により Internet を通じて海外サイトをみることや英語で書かれた論文にアクセスすることが簡単になってきている。このような背景から英語の学習への関心が強まってき

ている。英語を学習する方法は様々であるが、どの様な学習法でも英和辞書を使用するのは共通していると考えられる。最近では、英和辞書も電子化され単語を引く時間が短縮されて便利になってきている。

しかし、「多義な英単語に対し読んでいる文に合う訳語を選択する時に間違えるか迷ってしまう」という初心者がよく遭遇するトラブルは未解決のままである。なぜなら、辞書を引いている者の知識と経験が必要になるからである。英語の学習初心者や苦手としている者には、知識や経験は少ないと言える為、多義性を持つ英単語を調べる際に、何らかの支援が行えればより使いやすい辞書となることが期待できる。

我々は、多義曖昧性を持つ英単語を文脈にあう可能性のある意味に絞る事を支援する機能を、電子化英和辞書に付加することを研究の目的としている。

本稿では前置詞についての多義曖昧性を解消するシステムについて述べる。

2. 知的電子化英和辞書

知的電子化英和辞書とは、電子化英和辞書に上記で記述した学習を支援する機能を「知的な機能」とし、自然言語処理技術を用いて実現しようというものである。

図1はシステムの大まかな流れである。

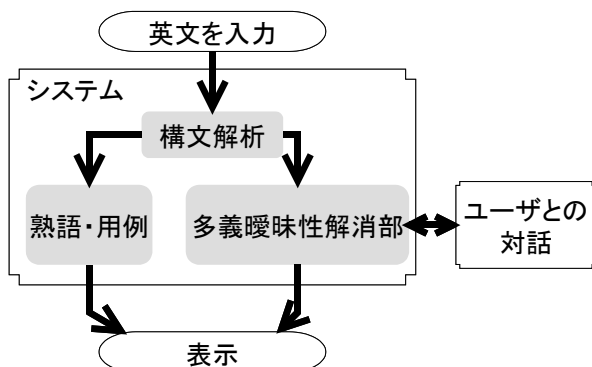


図1. システムの大まかな流れ

まず調べたい英単語が現れたとき、その英単語を含む英文を入力してもらう。次に、その英文はシステム内部で構文解析される。その結果、入力された英文がシステムに登録されている熟語・用例と同じ構造と判定された場合は、それを表示することになる。そうでなかった場合は、多義曖昧性解消部において、英文にあった意味に絞られ、表示されることになる。また、システム単独で多義曖昧性が解消されない場合は、ユーザとの対話によって曖昧性を解消し表示する。

本システムの主要な部分は図1における「多義曖昧性解消部」であり、これが従来辞書にはない「知的」な部分であり、本システムの特長となる。

多義曖昧性解消は、様々な手法が提案されているが、本システムでは特定のアプリケーションを前提としているため、以下の条件が他の一般的な手法と異なってくる。

- ・「意味」を特定するのではなく、辞書に記載されている「訳語」の中からふさわしいものを選択できればよい。
- ・解は必ずしも一つに限定できなくてもよい。
- ・なぜ、その訳語が選択されたかを説明できることが望ましい。

本システムは、英語学習初心者などを対象にしており、システムを使用する環境が「学習」というものであるため、翻訳システムのように意味の整合性を求めているのではない。

また、提示された訳語が一つだけであるとその訳語をそのまま自分の訳に当てはめるだけで学習にはならないと考えられる。つまり、どの訳語を用いるのかはユーザ自身がある程度は考えなければならない。

提示した訳語が選択された理由を説明で

きるようにすることで、学習そのものに役立ち、なおかつ、訳語選択のヒントにもなると考えられる。このような理由から上記した条件が現れてくる。

2.1 前置詞の多義曖昧性解消法^[1]

本システムでは多義曖昧性が多く、英語学習初心者が意味を取り違え易い前置詞をまず対象とした。

前節の最後で示したように、システムが訳語を選択した時に、その選択理由を説明する必要がある為、多義曖昧性解消は統計的な方法は利用できない。そこで、本システムではルールベースの多義曖昧性解消を行わなければならないと考える。

また、前置詞に関する多義曖昧性解消のルールを作成するに当たっては、英文の中で使用される前置詞が、以下で示す情報によって決定されるという特徴を利用した。

- ・ 文の構文構造
- ・ 文中の動詞
- ・ 前置詞の前後にある名詞

例えば、文の構造が受動態で文中に前置詞'by'が存在した場合は、「～によって（受け身の動作主）」という訳語（意味）になる可能性がある。また、"I went to the library"という英文中の to は'went'の様な「物理的な移動」を表す動詞があり、'library'の様な「物理的な場所」を表す名詞が来る場合は、「～へ、に、まで（到着点・行く先）」という訳語（意味）になる。

このように、3つの情報を利用することによりルールは作成され、文から得られる情報と照らし合わせることにより、可能性のある訳語（意味）だけを残していくことができ多義曖昧性は解消できる。

ただし、動詞に関しては上記のように「物

理的な移動」という様に、Schank が概念依存理論[2]において使用している基本動詞を若干拡張して使用している。本来の基本動詞は 11 個であるが「状態」「願望」「所有」という基本動詞を新たに追加し 14 個となっている。

また、名詞に関しても「物理的な場所」というように、単語の直接的な意味を使用するのではなく上位概念である意味素性を利用する。これは、シソーラスに代表される概念の is_a 階層をつかうことで判定できる。

例えば、"too many by one"や"win by a boat's length"のように'by'の後にくる名詞が「数量」の意味である場合の'by'は、【**程度・差異**】と限定できる。このチェックのためには'one'や'length'が「数量」の意味であることを判定できる事が必要である。そこで、

「by の後の名詞が上位概念として<measure>が存在するならば、訳語のカテゴリは【**程度・差異**】である」

というルールを用意しておけば、one については one -> digit -> integer -> number -> definite quantity -> measure, length については length -> dimension -> measure と階層をたどることによりどちらもこのルールにマッチし、適切な訳を選択することができる。(is_a 階層は WordNet による [3])

2.2 前置詞'by'の多義曖昧性解消

今回実装するシステムでは、前置詞'by'の多義曖昧性解消についてのルールを作成し、実装している。その際、小学館「ランダムハウス英和大辞典（第2版）」[4]を参考にルールを作成した。ここに記述されてある前置詞'by'の意味は 23 個ある。また、ルールは 26 個作成した。その一部分を表 1 に示す。

次に、このルールがどの様に使用されていくのか次の2つの例文と表 1 を用いて説明

表1. 前置詞'by'のルール

	英文の構造	byの前にある名詞の意味素性	byの後にある名詞の意味素性	byの前後の名詞の意味関係	英文の動詞の意味	byの訳語
Rule 1	受動態					~によって
Rule 2			数量			~だけ
Rule 3		人or物	人	親子or作品と著者		~から生まれた
Rule 4		人or物	"人or物"の一部分	後が前の一部分	獲得	~のところを

する.

a. The phonograph was invented by Thomas Edison.

b. Eve had two sons by Adam.

まず, この文から上記した3つの情報を取り出すと表2のようになる.

表2. 例文からの情報

	構造	前の名詞の意味素性	後の名詞の意味素性	動詞の意味
a	受動態		人	MBUILD
b	能動態	人	人	POSSESS

ここで, MBUILD とは「思考によって新たな情報を生成する」という意味であり, POSSESS とは「所有」という意味である.

表2の情報を表1のルール全てと比較すると, aはRule 1に当てはまり, bはRule 3に当てはまる. ユーザにはこの当てはまったルールに対応する訳語が提示されることになる.

この2つの例文では各例文に対し1つのルールという対応関係であったが, 他のルールとも当てはまる様ならばそのルールも適用され, ユーザには当てはまる可能性のある訳語としてルールに対応する訳語が全て提示されることになる.

2.3 評価

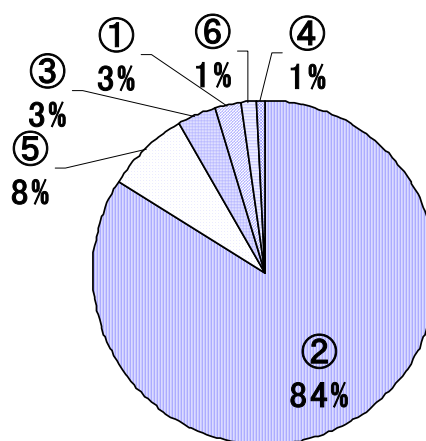
前置詞'by'の多義曖昧性解消をするルールを人手で評価した. ただし, 意味素性に関しては独自に定めたものを使用している.

評価には次に示す英文資料中に含まれる前置詞'by'を含む例文を用いている.

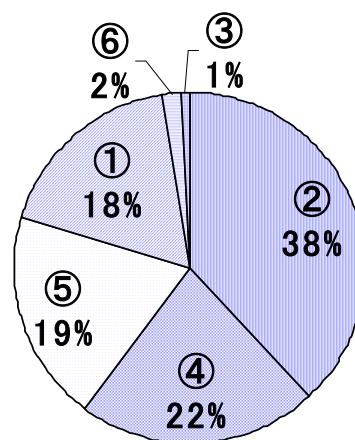
(1)D.O.Hebb "Organization of Behavior"

Chapter 4.

(2)P.H.Winston "Artificial Intelligence Third



(a) 技術論文での評価



(b) 文学作品での評価

- ① 熟語・用例
- ② 構文解析結果を利用したルール
- ③ 'by'の前後にある名詞を利用したルール
- ④ 文中の動詞を利用したルール
- ⑤ 'by'の後ろにある名詞だけを利用したルール
- ⑥ ユーザとの対話により解消

図2. 前置詞'by'の多義曖昧性解消をするルールの評価

Edition” Chapter 10.

(3)AAAI’99 の一部の論文

(4)R.J.Waller “The Bridges of Madison County”

(1)～(3)の資料は技術論文であり，(4)は文学作品である。

評価結果を図2に示す。これは，まず文情報とルールを比較していき，ルールと合うものを全て取り出す。そして，その中に正しい意味を示すルールが選択されている場合に正しく曖昧性が解消されたとして，グラフの対象としてそのルールを入れた結果である。

結果として，システムでの多義曖昧性の解消は95%以上となっている。また，技術論文では，受動態の形が多いため，構文情報による多義曖昧性解消が多くなっている。文学作品では，ルールは全般的に使用されていた。

3. システムの構成

システムを作成する環境は，Windows2000上の VisualC++を使用している。また，構文解析器として SIL International の PC-PATR [5]を利用した。文法は，本システムのために独自で開発した。名詞の意味素性のデータベースとしてプリンストン大学の WordNet を利用している。

従って，本システムの構成は図3に示す様になる。また，図4にシステムのインタフェースを示す。

図4のインタフェースを使用しながら，システムの流れを説明していくと，まず，調べる対象を含む英文もしくは現在読んでいる文章そのものを(b)のフィールドに入力する。ここではコピー&ペーストやテキストファイル指定で読み込めるようになっている。次に，その英文の文脈情報を(a)で指定する。文脈情報を指定することで，専門的に使用される訳語を提示することができる。また，文脈

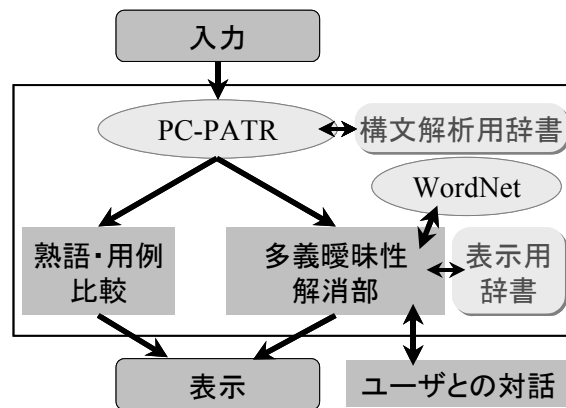


図3. システムの構成

情報として利用するのは，料理・コンピュータなどの分野を表すものと，新聞記事・論文などの文章のタイプを選択してもらうことになる。

その後，入力した英文を読んで行き，調べたい英単語が現れた時にその単語を反転表示させる。そして，(c)の「調べる」ボタンを押すことで，システムは英文にあった意味に絞り，その訳語を(e)に表示する。そしてユーザは絞られた訳語から英文に適した訳語を選択する。また，(d)は選択された英単語や英文を確認するためのものである。

本システムは，曖昧性解消した後のフォローとして，ユーザが選択した訳語に対し，「訳語が候補に挙げた理由」や「対象となる訳語を使用する例文」，「語法」などを表示する機能を作成することにより，より学習の効率が上がると考え，それを表示するための機能を(f)に付加する。

また，対象の単語が熟語の一部として使用されていると判定された場合に，例えば”know ~ by heart”のような場合，本システムでは，表示方法は”know the route by heart”と使われていれば，「the route を暗記している」のように文に合わせて行う。その機能が(g)の部分である。

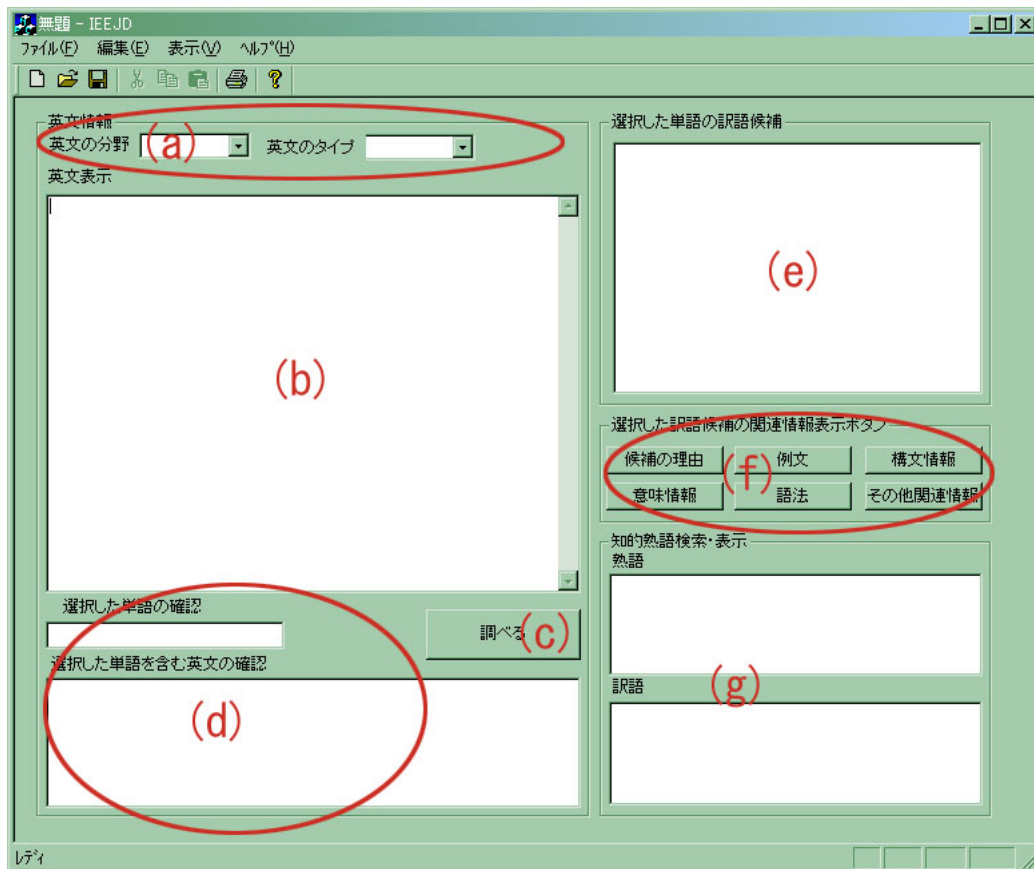


図4. システムのインタフェース

4. おわりに

本稿では、知的電子化英和辞書の主要部分である多義曖昧性の解消の内、前置詞の多義曖昧性解消部分について、その手法とシステムへの実装について述べた。本システムは、開発途中のものであり、前置詞'by'の多義曖昧性解消部を作成しているのだが、他の前置詞の多義曖昧性解消についても行っていくことを考えている。また、本システムが学習支援を目的にしたものであるために、より学習しやすいシステムを作成していくことを目指していく必要もあると考えている。

また、人手で行ったルールの評価においても、独自の意味素性を利用しているため、WordNet との整合性をとる必要もある。

参考文献

- [1] Ryo Nagaino, et al. : "English-Japanese dictionary system with intelligence", Pacific Association for Computational Linguistics 2001, pp.182-188, (2001).
- [2] 田中穂積 : "自然言語処理 ー基礎と応用ー", 電子情報通信学会, pp.51-55 (1999)
- [3] Christiane Fellbaum: " WordNet: an electronic lexical database" , MIT Press (1999).
- [4] 小学館 ランダムハウス英和大辞典 第2版 編集委員会 : "ランダムハウス英和大辞典", 小学館, p.384 (1994) .
- [5] Summer Institute of Linguistics: " PC-PATR -A syntactic parser- " , <http://www.sil.org/pcpatr/>