

# 単語の意味論的進化情報を付加した 英語シソーラスデータベースシステムの構築

○山崎 達也  
藤代 一成  
池辺 八州彦

Developing an English Thesaurus Database System with Evolving Semantics of Words ○Tatsuya Yamazaki  
Issei Fujishiro  
Yasuhiko Ikebe

## Abstract

*In traditional English thesauri, it is difficult for a user to choose from a synonym list, adequate words which can be used in his/her context. To resolve this difficulty, this paper designs and implements an English thesaurus database system with evolving semantics of words. The salient feature of the thesaurus database system is that it keeps track of word evolution from the Proto-Indo-European to modern English.*

## 1 英語シソーラス

英語シソーラスは、英語句を意味によって分類配列し、各語句について同義語、反意語などを記述した一種の類語辞典であり、主として、文章の作成において、表現したい意味内容を的確に表す英語句を選ぶ目的に使われる。本研究では、数多くあるシソーラスの中でも定評のある Roget's シリーズの一つである、Roget's II The New Thesaurus (以下 Roget's II と略す) [4] をとりあげる。通常英語シソーラスは、各語句の意味記述をもたないために英英辞典等と併用されるが、Roget's II は、各語句の間の同義、反意などの関係を示しながら意味記述も併せて示しているため、従来のものより利便性が高い。

## 2 Roget's II の問題点

しかしながら、Roget's II でも、シソーラス本来の目的の達成度の観点から見ると、まだ多くの問題点を残している。その問題点は、構造上と内容上の問題点の二つのカテゴリーに分けることができる。構造上の問題点については既に文献 [2] で論じられているので、ここでは内容上の問題点に焦点を当てることにする。

最も顕著な内容上の問題点として、“複数の同義語が一つのグループにまとめられているが、各々の語句は完全に同じ語義をもつわけではない。それにもかかわらず、同一グループ中の語句のニュアンスの違いをユーザに理解させるような情報を扱っていない”ということが挙げられる。これが問題となる具体的な状況例を以下に説明する：

1. ある文章を英作文している際に、“彼は教えることを自分に最も適した職業と考えていた”という文を訳さなければならなくなった。そこで、“He regarded the teaching profession as ????” という対訳を考えたが、“自分に最も適した職業”を一単語で表現したい。しかし、職業という語義をもつ単語としては、漠然と profession, occupation 等しか思いつかない。
2. そこで、profession を検索キーワードとして、英語シソーラスから適切な語句を探すことにする。そして、語義 “Activity pursued as a livelihood.” に関する同義語として、occupation, job, work, profession, vocation 等があることがわかった。どの語句がこの文脈では最も適当かを知りたいが、従来の Roget's II が扱う情報で判断できるのはここまでである。この中に望ましい語句があると思われるが、それを特定することはできない。

### 3 単語の意味論的進化情報の利用

前節で挙げられた Roget's II の内容上の問題点を解消するために、由来した語句とその意味を示す、単語の意味論的進化情報を扱うことを考える。本研究では、文献 [5] 等で既にその意義が認められている、印欧語根まで遡る進化情報を扱う。

印欧語根は、印欧原語の語根である。印欧原語は、英語の起源を有史前まで遡っていくと最終的にたどりつくと考えられる、比較言語学によって確立された推定先祖語である。C. Watkins らの研究によって、文献 [5] には 1414 個の印欧語根が再構成されており、その意味的説明、由来する現代英単語、ならびにその発展経過などが詳しく説明されている。なお全ての現代英単語について、その元となった印欧語根が明らかにされているわけではない。しかし、ある学習用参考書に使われた単語に関する調査では、その 87% について印欧語根が明らかになっている [1]。このことから、印欧語根を英語シソーラスの扱う情報とすることは、十分実用性のあることであると判断される。

単語の意味論的進化情報を利用して、単語間のニュアンスの違いを調べる例として、前節で示した状況を再考する。例えば、occupation と vocation の二つについて、単語の意味論的進化情報を調べると、それぞれ以下のような結果が得られる。まず vocation は、しっかりつかむ→強くつかむ→ある種の役職に就く→仕事に従事する、と意味が変化して現在の意味になったことから、元来は定常的に就く職業のことであったことがわかる。それに対し vocation は、話すこと→呼ぶこと→神に呼ばれること、と意味が変化してきたことから、元々、神に命ぜられた職業という意味合いが強かったことがわかる。以上から、本例のように“自分に最も適した職業”というニュアンスを含む英語句を選びたい場合、職業でも特に天職というニュアンスをもつ vocation が適当であることがわかる。

### 4 システムの構築

意味論的進化情報が、その有用性にも関わらず従来の英語シソーラスで扱われてこなかった原因の一つとして、紙を媒体とする辞書における情報量の制限が考えられる。この問題点は、計算機上にシソーラスデータベースを構築することで解決することができる [2]。

本シソーラスデータベースは、Roget's II で解説されている見出し語に関する情報と見出し語間の関係についての情報、並びに文献 [5] に収録されている語句の進化情報を扱うことにする。

一般に、データベースシステムの構築は、対象分析、システム設計そして実装の 3 フェーズに分けられる。本研究では、問題対象である英語シソーラスと意味論的進化情報の対象分析に、定評あるオブジェクト指向システム分析法の一つである Shlaer&Mellor 法 [3] を採用した。同法は、分析結果として、情報モデル、状態モデル及びプロセスモデルを生成する。

図 1 に本シソーラスデータベースの情報モデルを示す。全ての見出し語は、主見出し語として概念スキーマ上に表現され、見出し語が同一の意味記述をもつ場合、それらの見出し語は同義であると定義される。さらに、従来のシソーラスには存在しなかった単語の進化情報を扱えるような構造が与えられていることがわかる。

次に、これらのモデルを基に、C++ を用いてオブジェクト指向 DBMS VERSANT<sup>1</sup> 上に実装を行った。通常のシステム開発では、上述の各フェーズ間の行き来が頻繁に発生する。本研究でもそのことが予想されたため、それぞれのフェーズをオブジェクト指向のパラダイムで統一し、フェーズ間の行き来を容易にした。実装に用いた計算機は、NeXT station Turbo Color<sup>2</sup>、OS は NEXT STEP 3.2J<sup>3</sup> である。

<sup>1</sup>VERSANT は Versant Object Technology Corporation の登録商標である。

<sup>2</sup>NeXT は米国 NeXT Computer Inc. の商標である。

<sup>3</sup>NEXT STEP は米国 NeXT Computer Inc. の商標である。

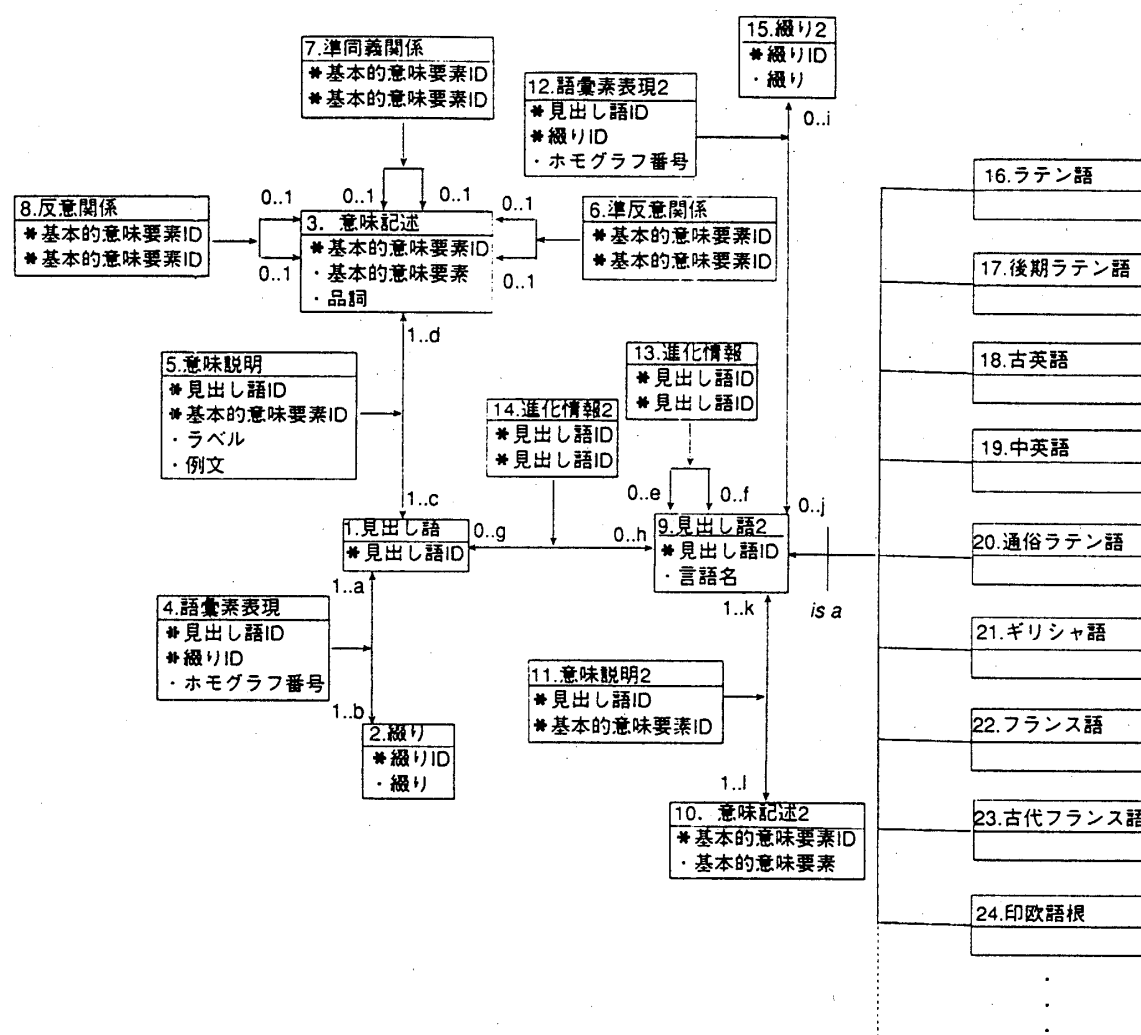


図 1: 英語シソーラスと意味論的進化情報の統合情報モデル

## 5 システムの使用例

実装されたシステムを用いて実験を行う。実際の検索実行例を図2に示す。下線部はユーザ入力を示す。検索の大まかな流れは、0. プログラム起動→1. 同義語を知りたい語句の入力→2. 入力された語句がもつ語義の出力→3. 同義語を知りたい語義を選択→4. 選択された語義に関する同義語を全て出力→5. 進化情報を知りたい語句を選択→6. 進化情報の出力→7. 他の語句についても知りたければ4へ戻り、そうでなければ終了である。以上の手続きによって適切な語句を選ぶための情報が順次ユーザに示される。

## 6 結論と今後の展望

従来の英語シソーラスが取り込めていなかった、“同義語としてまとめられた各語句のニュアンスの違い”を示せる英語シソーラスデータベースシステムを構築した。今後は、オブジェクト指向ユーザインタフェースへの拡張と辞書データの充実化を目指す。現在、同義語のニュアンスの違いを示す情報を与えているのはシステムであるが、最終的に最適な語句がどれかを判断しているのは人間である。そこで、これまでの研究で得られた知見を基に、現在人間が行っている判断を積極的に支援するシステム機能の拡充も併せて検討していく予定である。

## 情報知識学会 A-3

```

shikibu> ./find etwo
Input a spell you want to search
>occupation
Modern English
0: (noun) Activity pursued as a livelihood
1: (noun) The holding of something ,such as position
input number what you want to know its synonyms
>0
0: occupation
1: vocation
.. 中略 ..
input number what you want to know its origin
>0
The word has 1 ancestor
(occupation,Modern English) has (occupacioun,Middle English,occupation)
.. 中略 ..
(occupacion,Old French) has (occupare,Latin,to seize)
.. 中略 ..
(capere,Latin) has (kap-,Indo-European root,to grasp)
.. 中略 ..
Continue? y/n
>y
0: occupation
1: vocation
.. 中略 ..
input number what you want to know its origin
>1
The word has 1 ancestor
(vocation,Modern English) has (vocatio,Latin,a calling)
The word has 1 ancestor
(vocatio,Latin) has (vocare,Latin,to call)
The word has 1 ancestor
(vocare,Latin) has (wek(w)-,Indo-European root,to speak)
Continue? y/n
>n
shikibu>

```

図 2: シソーラスデータベースシステムの検索実行例

## 参考文献

- [1] 佐藤 和彦, 池辺 八洲彦, 藤代 一成: 印欧語根と英単語学習参考書, 第 45 回情報処理学会全国大会講演論文集, 5F-5, 1992
- [2] 藤代 一成, 池辺 八洲彦, 竹川 弘志: 計算機シソーラスシステムの概念スキーマ設計, 第 40 回情報処理学会全国大会講演論文集, 7F-7, 1989
- [3] Sally Shlaer & Stephen J. Mellor: *OBJECT LIFECYCLES Modeling the World in States*, Prentice-Hall, 1992.
- [4] Anne H. Soukhanov et al. (eds.): *Roget's II The New Thesaurus*, Houghton Mifflin, 1988.
- [5] C.Watkins: "Indo-European and the Indo-Europeans," in *The American Heritage Dictionary 1st college edition*, William Morris et al. (eds.), Houghton Mifflin, 1969.

---

山崎 達也:筑波大学大学院工学研究科 (Tatsuya Yamazaki: Doctoral Program of Engineering, University of Tsukuba). 藤代 一成:お茶の水女子大学理学部 (Issei Fujishiro: Faculty of Science, Ochanomizu University). 池辺 八洲彦:筑波大学電子・情報工学系 (Yasuhiko Ikebe: Institute of Information Sciences and Electronics, University of Tsukuba).