

# 場面情報に基づく英語名詞の語義の優先づけ方法と評価\*

3R-8

角田 達彦<sup>†</sup> 田中 英彦<sup>‡</sup>  
 東京大学 工学部

## 1 はじめに

意味解析時における解の探索空間の広がりによる処理の爆発と局所的制約の不足による誤りの問題を避けるため、効率的で適切な文の意味処理には語義の曖昧性の文脈による絞り込みが必要である。言語外の知識に基づく文脈情報に依存したテキスト解釈の手法として Schank らの場面に基づく文脈理解の研究 [1] があるが、知識源が不明確であり、また恣意的でないテキストに対しての評価・考察がなされていない。本稿では場面知識を視覚辞書から構成し利用した場合の英単語の多義性解消の効率化に対する効果の評価および検討を行なう。構文解析と意味解析は通常の記号処理手法を用い、格検証までの全探索を想定する (図 1(a))。正しい解に至るまでの試行回数の、場面特定による減少の効果を実文にて評価する。

## 2 場面情報による語義の優先づけ

場面情報の近似として OXFORD-DUDEN の Pictorial English Dictionary (OPED) を用いる。Kitchen などの日常生活に出てくる場面を 384 カテゴリに分け絵に描き、登場する物の名前を対応づけて列挙してある辞書であり、言語では明示的には通常扱われない常識の一部を含む (図 1(b),(c))。この絵を空間の情報としての場面の近似であるとして扱う。語の意味は Roget 5th ed. International Thesaurus の分類に従い、OPED の各場面と組合せ場面情報を構築する (図 2(a))。場面情報の表現として (a) 単語と意味の対をテーブルとして持つ方法と、(b) 全体の意味の頻度情報による分布として持つ方法とがある。(a) は Thesaurus の最も細かい項目分類の中から手で選び、その語の意味として登録する。(b) は (a) の頻度情報を利用するが、Thesaurus は図 1(d) のように、項目分類 (例: 728.6) とより大きな分類であるカテゴリ分類 (例: 728) があるが、これらには補完と意味干渉のトレードオフの関係があるため、実験の結果により有利な方を使用する。

場面が一つに決まった場合に、多義性解消の対象とする単語すなわち目標単語の候補となる複数の意味の中から試行のために取り出す順番を決める順序列を定める。まず目標単語が場面の (a) のテーブルに登録されているか調べ、登録されている場合はその意味を結果として用いる (図 2(c))。登録されていない場合には (b) の意味分布 (表 1 など) を調べ、目標単語の意味の候補リストの中で尤度の最も高いものから順番に取り出して意味解析部に渡す (図 2(d))。意味解析部はボトムアップパーザである LangLab を用い、トライ辞書を動的に書き換える (図 1(a))。

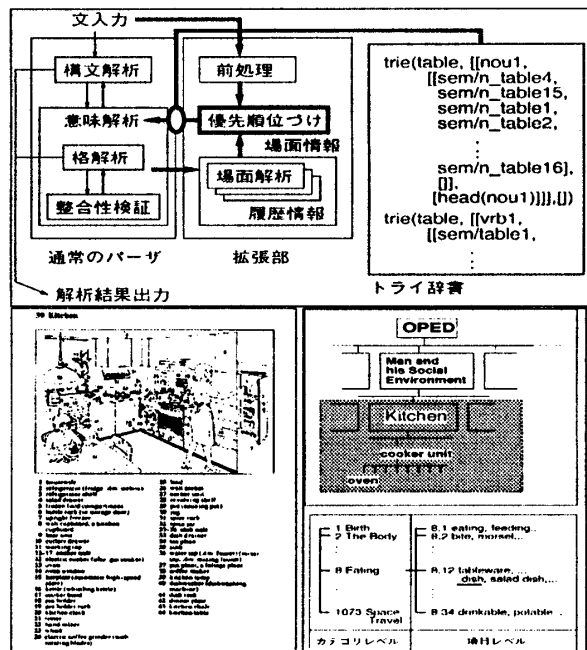


図 1: (a) システムの全体構成とトライ構造意味辞書 (b) OPED での台所のシーン (c) OPED の構造 (d) Roget 5th ed. International Thesaurus の構造

## 3 実験結果と考察

モンゴメリー作の「赤毛のアン」の英語原作の中から台所の場面を切りだし、台所に登場すると思われる名詞を選定した。基本情報を表 2 に示す。

選定した名詞でシソーラスに登録されている名詞の多義性の度数分布、およびランダムに意味分類を取り出した場合の、正解までの試行回数の分布を図 2(e) に示す。台所の場面情報に登録されている名詞、選定した全名詞、そして台所の場面情報に登録されていない名詞に対してそれぞれ、場面情報を使わずにランダムに取り出した場合の試行回数の平均、およびテーブルレベル、項目レベル、カテゴリレベルの試行回数の平均を表 3 に、分布を図 2(f) に示す。以上の結果から、(1) 場面情報に登録されている名詞に対してはテーブル引きによる優先順位づけを行なう (2) 場面情報に登録されていない名詞に対してはカテゴリレベルによる優先順位づけを行なうという方法をとる。その結果を表 3(d) に示す。図 2(e),(f) から、最初の試行での多義性解消率は 46% 以上であり、場面情報がない場合の 3 倍以上である。2 回目までの多義性解消率は 74% 以上である。

シソーラスでの意味の分類に冗長性が生じ、試行回数が無

\*Method of Preference Ordering of English Noun Sense based on Scene Information and its Evaluation

<sup>†</sup>日本学術振興会特別研究員

<sup>‡</sup>Tatsuhiko TSUNODA, Hidehiko TANAKA  
 Faculty of Engineering, University of Tokyo  
 {tsunoda,tanaka}@mtl.t.u-tokyo.ac.jp

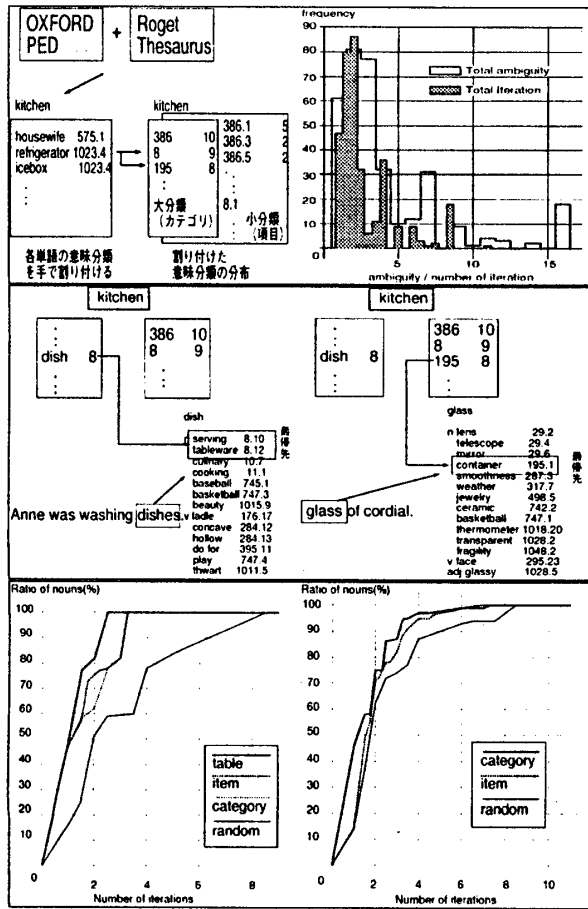


図 2: (a) 場面情報の表現方法 (b) 実文中の台所名詞のシソーラス上の多義性の分布および正解までの試行回数の分布 (場面情報なし) (c) 語義の優先づけ方法 - 1 (d) 語義の優先づけ方法 - 2 (e) 台所の場面情報に登録されている名詞の各試行回数までの正答率 (f) 選定した全名詞の各試行回数までの正答率

表 1: 台所の場を構成するカテゴリの分布

順位	category	category name	概念	頻度
1	386	Store, Supply	保管、供給	10
2	8	Eating	食事	9
3	195	Container	容器	8
3	239	Channel	水路(管)	8
5	11	Cooking	料理	7
6	742	Ceramics	陶磁器	6
7	79	Cleanness	清潔さ	4
7	293	Closure	閉鎖	4
7	1023	Refrigeration	冷却	4
-	others	-	-	55
	total			115

表 2: 選定した名詞の基本情報

台所に関する全名詞	357 個
シソーラスに登録されているもの	341 個
多義数の平均	4.13 個
正解までの試行回数の平均(ランダム)	2.71 回
台所の場面情報に登録されている名詞	114 個
台所の場面情報に登録されていない名詞	227 個

表 3: 平均試行回数 (a) 台所の場面情報に登録されている名詞の (b) 選定した全名詞の (c) 台所の場面情報に登録されていない名詞の (d) ランダムに採した場合(上段)と本手法適用結果(下段)

場面情報の表現方法	平均試行回数			
	(a)	(b)	(c)	(d)
場面情報なし(ランダム)	3.43	2.71	2.35	2.71
単語 - 意味対	1.47	--	--	--
項目による意味分類	1.72	2.08	2.26	--
カテゴリによる意味分類	1.82	1.83	1.83	1.71

意味が大きくなっているものがある。意味間での近さの違いが生じる現象に対しては観点という概念を導入し、その観点に応じて同じものとみなしたり区別したりする必要があると思われる。場面情報に登録されている名詞と登録されていない名詞とで、場を構成するために十分な意味の分布の統計量が得られていないため、項目レベルとカテゴリレベルの平均試行回数の優劣が異なる。このように不完全な場面情報しか得られない場合には、シソーラスのより上位のカテゴリレベルでの補充が効果を持つものと思われる。また逆に補充による意味の干渉も問題となる。例えば 'cream' は「清潔さ」の意味の干渉により洗剤のクリームを解答してしまっているが、文中の意味では「ケーキのソフトクリーム」であった。また今回の文章での解析では現れなかったが、同じ単語の形をとるが互いに意味が異なる複数の物が一つの場面に同時に現れる場合があり、この場合はテーブル引きをしても効果が減少する。例えば今回用いた台所の場面情報の中に 'plate' という単語が2つ含まれているが、1つは「皿」の意味であり、もう1つは「ホットプレート」の天板の意味であった。ただしそのような事例が出現する確率は非常に小さいものと思われる。

#### 4 おわりに

英語名詞の多義性解消の効率化の問題に対し、空間的連想を用いた文脈情報として場面が実文に対し有意な結果を与えることを示した。視覚的情報に基づく辞書とシソーラスを組み合わせる場合の構成方法の比較を行ない、考察を行なった。例として台所という場面に固定した場合に登場する名詞に対するの評価の結果、58%の平均試行回数の速度向上が見られている。また最初の解の多義性解消率は46%以上、2番目までの解では74%以上である。今後の課題は(1)場面を決定する機構との組合せ[2, 3](2)画像処理による実世界の場面情報の獲得(3)時間発展など他の連想関係の処理(4)動詞 - 名詞などの共起関係の処理の制御方法(5)単語の意味間の直接の依存関係の整合性のとり方があげられる。

なおこの研究の一部は文部省科学研究費の助成による。

#### 参考文献

- [1] R.C.Schank. *Dynamic memory*. Cambridge University Press, 1982.
- [2] 角田達彦, 田中英彦. 結束構造および一貫性に基づく場面構造の解析. 情報処理学会第49回全国大会, Vol. 3, pp. 121-122, 1994.
- [3] T.Tsunoda and H.Tanaka. Analysis of scene identification ability of associative memory with pictorial dictionary. In *Proceedings of COLING-94*, Vol. 1, pp. 310-316, 8 1994.