

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

内元 清貴[†] 宇津呂 武仁[‡] 長尾 眞[†]

[†] 京都大学工学部 電気工学第二教室

[‡] 奈良先端科学技術大学院大学 情報科学研究科

自然言語コーパスからの語彙的知識獲得における未解決の問題として、語の持つ多義性の類別の問題がある。本論文では、動詞の語彙的知識獲得において動詞の多義性の類別に焦点を当て、これを類義語の用例を用いて行う方法について述べる。まず一つの動詞について、格要素の名詞の意味が近い用例をクラスターにまとめる。次に、各クラスターに対して、類義語の用例集合から格要素の名詞の意味が近いものを見つけ出し、動詞の意味の解釈として類義語と最も近い意味を選ぶ。動詞の複数の意味の区別は動詞シソーラスに基づいて行う。類別の効率より類別の精度を重視するような設定で実験を行い、精度のよい類別の結果を得た。

Synonym-based Sense Discrimination in Lexical Acquisition of Verbs

Kiyotaka UCHIMOTO[†] Takehito UTSURO[‡] Makoto NAGAO[†]

[†]Dept. of Electrical Engineering, Kyoto University

[‡]Graduate School of Information Science, Nara Institute of Science and Technology

This paper proposes a synonym-based method for discriminating multiple senses of verbs in acquiring lexical knowledge of verbs from corpus. In the method, first, examples of a target verb are divided into clusters of similar examples. Then, for each cluster, an example containing a synonym of the target verb and similar case element nouns is selected and the sense ambiguity of the target verb is resolved to be closest to the sense of the synonymous verb. Sense distinction of verbs is based on an existing hand-compiled thesaurus of verbs. In experiments, we preferred precision to recall and achieved high precision.

1 はじめに

自然言語処理の分野において計算機で利用可能な大規模な意味辞書を構築するためには、人間のための辞書や大規模コーパスに含まれる自然言語の文を解析して、そこから意味辞書を自動的に構築する技術を確立することが重要となる。自然言語コーパスからの語彙的知識獲得においては、動詞の語彙的知識の獲得に関する研究が比較的多く行われており、例えば英語の動詞については、コーパスから動詞の表層格パターンを抽出しようというもの [1, 5], あるいは、人手の助けを借りて動詞の深層格構造を抽出しようというものがある [4]。また、対訳コーパスを用いれば、単言語文の統語的・意味的曖昧性解消に有利であるという観点から、日英の対訳コーパスを用いて、日本語の動詞の表層格フレームを獲得しようという研究も行われている [8]。

このような自然言語コーパスからの動詞の語彙的知識獲得における未解決の問題として、一つの動詞が持つ多義性の類別の問題がある。これはすなわち、ある動詞の語彙的知識を獲得する際に、その動詞の持つ複数の意味を分けて、それぞれの意味ごとに語彙的知識を獲得しなければならないという問題である。例えば、過去の研究についてみると、英語の動詞について行われた上述の研究 [1, 4, 5] においては、動詞の多義性は全く考慮されていなかった。また、日英対訳コーパスから日本語動詞の表層格フレームを獲得する研究 [8] においては、英語の情報を手がかりにすることによって日本語動詞の多義性の類別がうまく行えたが、単言語でどこまで類別が行えるかは明らかにしていない。

そこで、本研究では、自然言語コーパスからの動詞の語彙的知識獲得において、日本語における動詞の多義性の類別を行う方法について研究を行い、検討する。

まず名詞の意味的近さの尺度に関しては既存のシソーラスに頼り、同じ動詞の用例集合において、ある格の格要素の名詞がシソーラス中で意味的に近いものを一つのクラスターとしてまとめる。その際、クラスター内に動詞の意味の異なる用例が混在しないように、シソーラス中で意味的に近いとする範囲を制限する。

次に、各クラスター中の動詞がどのような意味であるかを決定する。このとき、動詞の意味・用法上の意味区別については既存の動詞シソーラスに頼るものとする。具体的には、クラスター内の用例に対して、類義語の用例集合から、ある格の格要素の名詞が類似している用例を見つけ出し、動詞の意味の解釈として類義語の意味に最も近い意味

をとるという方法で、多義性の類別を行う。また、クラスター内の全用例をその類別された意味で統一することとする。

実際に自然言語コーパスから動詞の用例を集め、動詞の多義性を類別する際には、適切な類似用例が必ず含まれるとは限らないという問題があるが、本研究では、コーパス中の用例の不足を人間との質問・応答によって補うことにより、類別された結果の精度を優先しながら多義性の類別の効率も上げるという立場をとった。この際、多義性の類別については、コーパス中の類義語の用例を用いる方法と基本的に同じ枠組で行った。

以下、2節、3節で本研究で用いた動詞句用例、シソーラスの構造について述べる。4節で類似用例のクラスタリング、5節で類義語の用例を用いた多義性の類別について述べ、6節でその具体例を挙げる。さらに、7節、8節では、人間との質問・応答の利用について述べる。9節では実験及び実験結果について述べ、考察する。また、10節で結論を述べる。

2 動詞句用例

本論文では、動詞句用例を動詞の表層格構造によって表現する。一般に、日本文の表層格構造は次のような素性構造で表すことができる。

$$\left[\begin{array}{l} \text{pred: } V \\ \text{sem: } SEM_0 \\ p_1 : \left[\begin{array}{l} \text{pred: } N_1 \\ \text{sem: } SEM_1 \end{array} \right] \\ \vdots \\ p_n : \left[\begin{array}{l} \text{pred: } N_n \\ \text{sem: } SEM_n \end{array} \right] \end{array} \right]$$

ここで、 V は動詞、 p_1, \dots, p_n は日本語の格ラベル¹を、 N_1, \dots, N_n は格要素の名詞を表す。ある名詞 N_i が複数の意味を持つ場合、この名詞はシソーラス中の複数の(葉の位置の)意味カテゴリをもつ。動詞 V の場合にも同様である。そこで、 SEM_0 および各 SEM_i をシソーラス中の意味カテゴリの集合によって表現し、これを動詞 V および各名詞 N_i の意味ラベルと呼ぶ。

$$SEM_0 = \{Sem_{v1}, \dots, Sem_{vk}\}$$

$$SEM_i = \{Sem_{i1}, \dots, Sem_{in}\}$$

¹ここで格ラベルとは、表層格ラベルのことである。表層格ラベルとなり得るものは、格助詞・係助詞・述語接続助詞・引用助詞などである。

動詞句用例間で類似しているかどうかは各々の意味カテゴリに依存し、格要素の名詞の表記には依存しない。

3 シソーラス

現在のところ、我々は、計算機上で利用可能な日本語シソーラスとして、「分類語彙表」[6]を用いている。「分類語彙表」は6層の階層構造から構成され、階層構造の葉の部分に総数約60,000語の単語が分類されている。一つの単語が多義性を持つ場合には、その語にはシソーラス中の複数の意味カテゴリが割り当てられる。「分類語彙表」は、現在入手可能な日本語のシソーラスの中では、最も正確で収録語数の最も多いシソーラスである。

「分類語彙表」の体の類(名詞シソーラス)の部分には、総数約45,000語の名詞が分類されており、根節点の一つ下のレベルは、「抽象的關係」、「人間活動の主体」、「人間活動—精神および行為」、「生産物および用具」、「自然物および自然現象」という5つの意味カテゴリからなる。さらにその下は、細かく5レベルの深さに細分類されており、この木構造の葉の部分に名詞が分類されている。

「分類語彙表」の用の類(動詞シソーラス)の部分には、葉の位置の意味カテゴリが約8,700個あり、総数約7,000語の動詞が分類されている。根節点の一つ下のレベルは、「抽象的關係」、「精神および行為」、「自然現象」という三つの意味カテゴリからなる。さらにその下は図1に示すように細かく5レベルの深さに細分類されており、この木構造の葉の部分に動詞が分類されている。

4 類似用例のクラスタリング

ある動詞 v の動詞句用例全体の集合 E_G において、用例間で、ある格に着目し²、格要素の名詞が類似している用例を一つのクラスターとしてまとめる³(図2参照)。その結果、 E_G はクラスター

$$CL_1, CL_2, \dots, CL_n \quad (\text{ただし, } CL_i \subset E_G)$$

に分割される。

二つの名詞が類似しているのは、名詞シソーラス中で共有する親ノードが L_n レベル以下にある時とする。 L_n の

²どの格に着目するべきかは動詞によって異なるはずで、作成されたクラスター内に動詞の意味の異なる用例が混在しないような格を用例集合から自動的に選べる必要がある。現在はその段階まで至っておらず、着目すべき格については人間が判断して与えることにする。

³その際、名詞の多義性、副動詞の曖昧性は類似用例を用いて解消したものを扱う[7]。

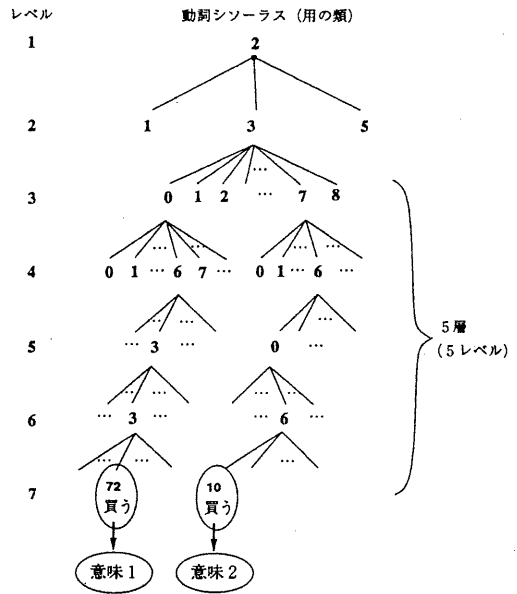


図1: 動詞シソーラスの木構造

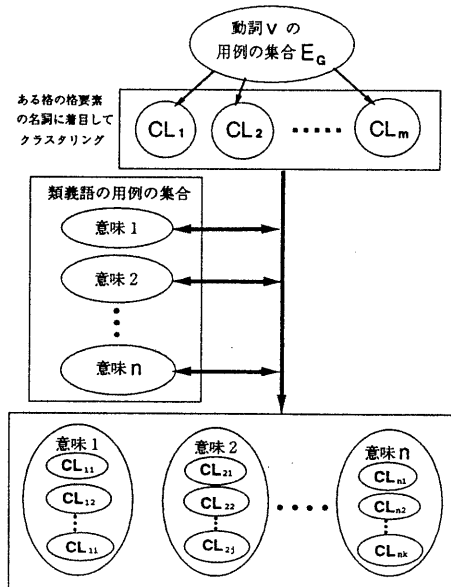


図2: 類義語を用いた多義性の類別の概念図

具体的な値の設定については、6節で述べ、実験の結果から9.2節で適切な値について考察する。その際、一つのクラスター内に動詞の意味の異なる用例が混在しないようにレベルを設定する。そのためにはシソーラス中で意味的にかなり近いもののみを信用し、 L_n の設定を厳しくする必要がある。

5 類義語の用例を用いた多義性の類別

本研究では、動詞の用法上の意味区別は動詞シソーラスにあるとし、クラスター内の一つの動詞句用例について、まず、動詞が動詞シソーラス中で L_v レベル以下に共有する親ノードを持つものを類義語とする。次に類義語の用例を集め、その中から、着目すべき格の格要素の名詞が名詞シソーラス中で L_n レベル以下に共有する親ノードを持つ用例を選択する。そこで、動詞句用例の動詞の意味の解釈として、選択した類義語の用例の動詞の意味に最も近い意味をとるという方法で多義性を類別する。さらに、類別された意味をクラスター内の全用例の意味とすることによって、クラスターに明確な意味を与える。これにより、同じ意味を持つクラスターをまとめることができる(図2参照)。 L_v の具体的な値の設定については、6節で述べる。

名詞シソーラスのみを用いて同じ意味を持つクラスターをまとめるのが難しいのは、例えば、図2における意味1の CL_{11} が、名詞シソーラス中では、意味1の CL_{12} より意味2の CL_{21} に近く、クラスター作成の際、 L_n の設定を緩めてゆくと、意味1の CL_{11} と意味1の CL_{12} が一つのクラスターとなる前に、意味1の CL_{11} と意味2の CL_{21} が一つのクラスターとなってしまうことがあるためである。そこで本研究では、動詞の用法上の意味区別を動詞シソーラスに求め、それぞれのクラスターごとに意味を決定するという方法をとった。

ここで、動詞シソーラスとしては「分類語彙表」を用いる。「分類語彙表」によれば、ここに収められた語は、「同義語義の関係で一つにまとめられており、同じ文脈の同じ地位に現われる」という観点で類別されたものだから、類別の際、類義語の用例を用いるのは妥当な方法だと考えられる。

6 例

実験は次のような設定で行う。4節、5節に述べた L_n 、 L_v については、 $L_n = 6$ 、 $L_v = 5$ とする。類義語の用

例としては、 L_n 、 L_v を満たすものの中から、格要素の名詞が最も類似しているものを選択する。名詞の類似度が同じものについては動詞のより類似しているものを選択する。二つの語は、シソーラス中で共有する親ノードが下のレベルにあるほど類似していると考える。

以下に具体的な類別の例を挙げる。

6.1 動詞「買う」の例

図1に示すように、動詞「買う」には、動詞シソーラスの木構造において、7レベルに二つの(葉の位置の)意味カテゴリが割り当てられている。これを意味1(‘評価する’の意味)、意味2(‘売り買い’の意味)とする。

動詞「買う」の用例数は75文であり、作成されたクラスターの数は、39個であった。クラスタリングは「を」格に着目して行った。

クラスター 1	
例	[を, 手腕]
類義語の用例	[動詞, 疑う],[を, 手腕]
クラスター 2	
例	[が, 私],[で, 本屋],[を, 本]
例	[が, 私],[を, 書物],[で, 推薦]
類義語の用例	[動詞, 買い与える],[を, 本]
クラスター 3	
例	[を, 不興]
例	[が, 行動],[を, 騒動]
類義語の用例	なし

クラスター1及びクラスター2の類義語「疑う」、「買い与える」に割り当てられている(葉の位置の)意味カテゴリは動詞シソーラス中ではそれぞれ意味1、意味2に近い。動詞句用例の動詞の意味の解釈として、選択した類義語の意味に最も近い意味をとるという方法で多義性を類別するため、クラスター1は意味1(‘評価する’の意味)に、クラスター2は意味2(‘売り買い’の意味)に類別される。クラスター3は類義語の用例がないため、類別されない。

類別されなかった用例には次の二通りの解釈ができる。一つは、類義語の用例が疎らなために類似用例がない場合であり、もう一つは動詞シソーラスの不完全さにより、本来類別されるべき意味がない場合である。前者は、類義語の用例が十分あれば必ず類別されるもので、後者は、動詞シソーラスが補完されない限り類別されないものである。クラスター3は後者の場合である。

6.2 動詞「当たる」の例

動詞「当たる」には、動詞シソーラスの木構造において、7レベルに四つの(葉の位置の)意味カテゴリが割り当てられている。これを意味1(‘命中・適中する’の意味)、意味2(‘相当する’の意味)、意味3(‘中毒になる’の意味)、意味4(‘(敵に)向かう / 従事する’の意味)とする。

動詞「当たる」の用例数は62文であり、作成されたクラスターの数は、46個であった。

クラスターの作成の際、動詞によって着目すべき格が異なる。着目すべき格が複数あるものについては格に優先順位を与えて、優先順位の高い格から順に着目してクラスタリングをする。例えば、動詞「当たる」の場合、着目すべき格が「に」格と「が」格の二つあり、優先順位は「に」格、「が」格の順である。この時、次のようなクラスターが作成される。「に」格がない用例は、「が」格を見て分けている。

クラスター 1		
が格	例	[が, 山]
	例	[が, 見込み]
	類義語の用例	[動詞, 適中する],[が, 見込み]
クラスター 2		
に格	例	[に, 事業]
	例	[が, 会社],[に, 事業]
	類義語の用例	[動詞, 取り組む],[が, 会社],[に, 事業]

また、類義語の用例を選択する際、クラスター内に着目すべき格を全て持つ用例があれば、それをクラスターの代表例として取り挙げ、類義語の用例としては着目すべき格全てについて近い用例を選択する。これは、例えば動詞「当たる」のように「に」格と「が」格に着目すべき動詞においては、「に」格だけ着目してもどちらの意味か判断できない場合でも、「が」格も着目すれば判断できることが多いためである。例えば、「彼女が私の祖母に当たる」、「弾が足に当たる」などがそれに当たる。具体的には、上の例では、クラスター2の代表例として「会社が事業に当たる」が取り挙げられ、類義語の用例としては「に」格と「が」格の両方が近い「会社が事業に取り組む」が選択された。

各クラスターの類義語「適中する」、「取り組む」に割り当てられている(葉の位置の)意味カテゴリは動詞シソーラス中ではそれぞれ意味1、意味4に近い。従って、クラスター1は意味1(‘命中・適中する’の意味)に、クラスター2は意味4(‘(敵に)向かう / 従事する’の意味)に類別される。

7 人間との質問・応答の利用

十分な数の類義語の用例を自動的に集めることは容易でない。従って、集めた類義語の用例は疎らなことが多く、その結果、類義語の用例がないクラスターが多く現われてくる。クラスターに類義語の用例がないような場合、多義性の類別は行われない。そこで、多義性の類別が行われなかったクラスターについて、コーパス中の類義語の用例の不足を人間との質問・応答によって補うことにより多義性の類別を行う方法について述べる。

具体的には、まず多義性の類別されなかったクラスターから用例の一つ取り出してくる。その用例において、クラスタリングする際に着目した格の格要素の名詞に注目し、いくつかの類義語に対して、その名詞が同じ格の格要素になり得るかどうかという質問を人間に対して行う。人間はyes/noで答えればよい。

例えば、動詞「買う」について、多義性の類別が行われなかった次のようなクラスターがある。

クラスター 4		
例	[を, 新車]	[を, 中古車]
	[を, 自動車]	[を, 外車]

システムは、クラスター内の一例の「を」格の格要素「新車」を取り出して、例えば、動詞「買う」の類義語「判断する」、「買い求める」の「を」格の格要素となり得るかどうかを質問してくる。この時、類義語の動詞に多義性のある動詞がいくつかの意味で解釈できてしまうため、類義語は多義性のないものを選択する。また、人間は「新車を買う」というもとの用法と近いかどうかを判断して、yes/noで答える。

質問	yes/no
「新車を判断する」	no
「新車を買求める」	yes

この質問・応答のシステムにおいても、多義性の類別は5節の方法と基本的に同じ枠組で行える。従って、類別の結果は、クラスター4の類義語の用例として「新車を買求める」が存在した時と同じことになる。

8 人間との質問・応答における複数の動詞の多義性の同時類別

前節では、多義性の類別が行われなかったクラスターについて、人間との質問・応答によってその多義性を類別し

表 1: 類義語及び用例数 (かっこ内)

動詞「買う」 — 合計 168 文
みなす (1), もくする (1), 当てる (10), うなづく (4), 売りさばく (4), 予期する (1), 売る (22), 求める (4), 迷う (3), 揺るぐ (1), おさめる (19), 買い与える (2), 見抜く (1), まける (7), 計る (3), 見付ける (1), 推す (2), 否定する (1), 仕入れる (2), 売りつける (2), 察する (4), 注文する (5), ならむ (6), 確定する (1), 疑う (1), 売れる (1), 買い占める (5), 売り切る (1), 売り渡す (2), 買い切る (1), もらう (2), 判断する (6), 決める (20), 定まる (2), 決定する (4), 取り決める (1), 規定する (1), 定める (3), 売り出す (1), 判定する (1), 決する (4), きまる (3), 決心する (2).
動詞「売る」 — 合計 163 文
裏切る (5), 認める (18), 申し込む (1), 買いあさる (1), 要請する (2), 買い切る (1), 買う (74), 売りさばく (2), 売れる (1), 要求する (15), 買い込む (1), 注文する (5), 買い占める (4), 断る (1), 買い与える (2), 願う (2), 売り切る (1), 売り渡す (2), 買い上げる (1), 購入する (3), 申し出る (1), 買い求める (2), 仕入れる (3), 訴える (13), 売り出す (1).
動詞「当たる」 — 合計 53 文
尋ねる (3), ばれる (1), 下る (1), 適中する (2), 突き当たる (1), 差す (1), 宛てる (1), 充てる (1), 合う (23), 触れる (1), 向かう (2), ぶつかる (1), 取り組む (1), 衝突する (1), いたむ (1), 当てる (12).
動詞「読む」 — 合計 110 文
歌う (3), 書き改める (1), 理解する (3), 書き損じる (1), 書き入れる (2), 読みこなす (2), 砕く (1), 読み込む (1), 書く (80), 書き落とす (1), 書き込む (1), 読み返す (1), 汲み取る (1), 書き上げる (3), 書き留める (3), 読み上げる (2), 分かる (1), 読みふける (2), 読み取る (1).

た。類別が行われないクラスターは、類義語の用例が疎らであるほど多くなり、また、それぞれの動詞の用例については少なくとも、様々な動詞の用例についてみれば莫大な数になる。その結果、質問・応答の数も多くなる。しかし、本研究の趣旨から、動詞が異なっても、類義語の関係にある時には、格要素の名詞が類似していれば動詞の意味も類似していると考えてよい。そのため、類義語同士の間で、同じ格の格要素の名詞が類似しているクラスターを一つにまとめてよい。これにより、多義性の類別が行われなかったクラスターを複数同時に解消することができ、質問・応答の効率上がる。

例えば、次のような、動詞「買う」、「売る」のクラスターは、ともに類義語の用例として「本を購入する」があれば類別される。従って、このようなクラスターを一つにまとめておけば質問・応答の数は半分で済む。

クラスター 1				
動詞	「買う」	例	[を, 本]	[を, 書物]
クラスター 2				
動詞	「売る」	例	[を, 本]	[を, 書物]

9 実験及び考察

9.1 実験

実験は、計算機上で利用可能な和英辞典 [9] から抽出した動詞句用例約 700 文について行った。この約 700 文には動詞「買う」、「売る」、「読む」、「当たる」の動詞句用例と、その類義語の用例が含まれている。類義語は、動詞ソーラス中で、共有する親ノードが 5 レベル以下 (5 節で、 $L_0 = 5$) のものである。表 1, 表 2 にそれぞれ類義語及び用例数、動詞の動詞ソーラス中での意味を示す。

また、これらの用例は全て形態素解析及び係り受け解析により表層格構造を抽出した結果を人手で修正し用いた。

結果を表 3 の実験結果 (1) に挙げる。また、人間との質問・応答の結果を表 3 の実験結果 (2) に挙げる。

9.2 考察

和英辞典中の用例のみを用いた場合 (表 3 の (1)), 類別の効率あまりよいとは言えないが、正解率は動詞「買う」で 97%, 「売る」で 100% 「当たる」で 80%, 「読む」で 100% と、動詞によってばらつきはあるもののよい結果が得られている。これは、本研究の方法の正当性を証明するものであると言える。

表 2: 動詞の動詞シソーラス中での意味

動詞	シソーラス中での意味	意味の数
買う	‘売り買い’, ‘評価する’	2
売る	‘売り買い’, ‘裏切る’	2
当たる	‘命中・適中する’ ‘相当する’, ‘中毒になる’ ‘(敵に) 向かう / 従事する’	4
読む	‘読み書き’, ‘詠む’ ‘見透かす’, ‘数える’	4

類別の効率があまりよくない原因について考える。原因の一つとして、現在利用可能なコーパスに十分な数の類義語の用例が含まれていなかったことが挙げられる。これは、表 3 の実験結果 (2) を見ても分かるように人間との質問・応答によりかなり補うことができる。

もう一つの原因としては、動詞シソーラスの不完全さが挙げられる。本論文では、動詞の複数の意味の区別は動詞シソーラスに基づいて行うが、日本語基本動詞用法辞典 [3] や、計算機用日本語基本動詞辞書 (IPAL) [2] における動詞の意味の分類に比べると「分類語彙表」の動詞シソーラスの分類は不完全なため、本論文の研究方法では動詞の意味が動詞シソーラスに分類されていない用例は類別することができないのである。例えば、動詞「買う」の場合「不興を買う」、「行動が響きを買う」などがそれに当たる。これらの用例は、動詞シソーラスが補完されれば類別できる。

次に、類別を誤った例について考察する。例えば、次のような例

例 1 ばちが当たる

について、類義語の用例としては、

例 2 天罰が下る

がある。ところが、動詞シソーラスには、動詞「当たる」と動詞「下る」に「罰を与える」という意味はなく、その代わりともに‘中毒になる’という別の意味で近いため、誤って類別されてしまう。このような例は、シソーラスが補完されても複数の意味に類別されてしまい、もちろん片方は間違った意味となる。この時、動詞「下る」以外に多義でない類義語の用例があれば、間違えて類別された意味をノイズとして取り除くことは可能であると考えられる。

このように複数の意味に間違えて類別されてしまうのは、類義語が多義である場合が多い。そこで、類義語が多義で

表 3: 実験結果

動詞	買う	売る	当たる	読む
(1): 和英辞典中の用例のみを用いた場合 ($L_v = 5, L_n = 6$)				
類別の効率 1	35/75 (47%)	27/36 (75%)	20/62 (35%)	14/22 (64%)
類別の効率 2	35/64 (55%)	27/33 (82%)	20/52 (42%)	14/22 (64%)
類別の正解率	34/35 (97%)	27/27 (100%)	16/20 (80%)	14/14 (100%)
クラスター比 1	14/39 (36%)	18/27 (67%)	14/46 (30%)	11/17 (65%)
クラスター比 2	14/34 (41%)	18/24 (75%)	14/40 (35%)	11/17 (65%)
(2): 人間との質問応答を利用した場合 ($L_v = 5, L_n = 6$)				
類別の効率 1	64/75 (85%)	33/36 (92%)	48/62 (81%)	22/22 (100%)
類別の効率 2	64/64 (100%)	33/33 (100%)	48/52 (92%)	22/22 (100%)
類別の正解率	63/64 (98%)	33/33 (100%)	44/48 (92%)	22/22 (100%)
クラスター比 1	34/39 (87%)	24/27 (89%)	36/46 (78%)	17/17 (100%)
クラスター比 2	34/34 (100%)	24/24 (100%)	36/40 (90%)	17/17 (100%)
(3): 類義語が多義でない時 ($L_v = 5, L_n = 5$)				
類別の効率 1	23/75 (31%)	18/36 (50%)	7/62 (11%)	19/22 (86%)
類別の効率 2	23/64 (36%)	18/33 (55%)	7/52 (13%)	19/22 (86%)
類別の正解率	23/23 (100%)	18/18 (100%)	7/7 (100%)	17/19 (89%)
クラスター比 1	9/30 (30%)	11/23 (48%)	3/37 (8%)	10/13 (77%)
クラスター比 2	9/27 (33%)	11/20 (55%)	3/31 (10%)	10/13 (77%)

$$(\text{類別の効率 1}) = \frac{(\text{多義性の類別された用例数})}{(\text{動詞句用例の総数})}$$

$$(\text{類別の効率 2}) = \frac{(\text{多義性の類別された用例数})}{(\text{シソーラスに意味のある用例数})}$$

$$(\text{類別の正解率}) = \frac{(\text{正しく類別された用例数})}{(\text{多義性の類別された用例数})}$$

$$(\text{クラスター比 1}) = \frac{(\text{多義性の類別されたクラスター数})}{(\text{全体のクラスター数})}$$

$$(\text{クラスター比 2}) = \frac{(\text{多義性の類別されたクラスター数})}{(\text{シソーラスに意味のあるクラスター数})}$$

ないものから用例を選択することにして実験を行ってみた。ただし、類別の効率が著しく悪くなるのを防ぐため、格要素の名詞が類似しているがどうかの基準をゆるめて5レベル以下(4節, 5節で, $L_n = 5$)に共有する親ノードを持つ時と設定した。結果を表3の実験結果(3)に挙げる。表3の実験結果(1)と比較してみると、正解率は動詞によってよくなったものもあれば悪くなったものもあり、類別の効率は悪くなってしまっているので、一概にどちらがよいとは言えない。

さらに、クラスター作成の際、4節での L_n の値を $L_n = 5$ として、類義語が多義であるものからも用例を選択することにして実験を行ったところ、作成されたクラスターそのものには意味は混在していないが、類別の正解率が悪くなった。これは、同じ動詞の同じ格の格要素となるものという視点で見ると、名詞シソーラス中の広い範囲に近い意味となるものがあるが、類義語の同じ格の格要素となり得るものという視点で見ると、近い意味として許容できる範囲が狭くなることを示している。そこで、クラスター作成の際、4節での L_n の値を $L_n = 5$ とし、類義語の用例は着目すべき格の格要素の名詞が共有する親ノードが6レベル以下(5節での L_n の値を $L_n = 6$)となるものから選択することにして実験を行ったところ、正解率も類別の効率も向上することが分かった。

10 おわりに

本研究では、自然言語コーパスからの動詞の語彙的知識獲得において、動詞の多義性の類別に焦点を当て、これを類義語の用例を用いて行う方法について実験、検討を行った。具体的な方法としては、まず名詞シソーラスを用いて同じ動詞の用例をクラスタリングし、作成されたクラスター内の代表例について、類義語の用例を用いて動詞の多義性を類別し、クラスター内の全用例をその意味で統一することにより、クラスターに明確な意味を与えた。多義性の類別の際、動詞の複数の意味の区別は動詞シソーラスに基づいて行い、そのもとで、意味的に近い類義語の用例を選択し用いた。実験は、類別の効率より類別の精度を重視するような設定で行った。

本研究では、動詞シソーラスとして「分類語彙表」を用いたが、動詞シソーラスに分類されていない動詞の意味がいくつかあり、そのような意味の用例については多義性を類別することができなかった。今後、動詞シソーラスが補完されると、このような用例の多義性の類別は可能となる

であろう。また、現在利用可能なコーパス中に十分な数の用例がないため、適切な類義語の用例が得られない場合があった。このような場合については、人間との質問・応答によって用例の不足を補う方法をとった。

本研究の方法では、多義性の類別の際、人間が着目すべき格を与えたが、どの格に着目すべきかを自動的に判定する方法については検討する必要がある。また、類別の正解率がよくない動詞の用例についてどのような工夫が必要であるかを検討するとともに、より多くの動詞の用例について本研究の方法でどのような結果が得られるかを実験・検討する必要がある。

参考文献

- [1] M. R. Brent. From grammar to lexicon: Unsupervised learning of lexical syntax. *Computational Linguistics*, Vol. 19, No. 2, pp. 243-262, 1993.
- [2] 情報処理振興事業協会技術センター. 計算機用日本語基本動詞辞書 IPAL(Basic Verbs) 説明書, 1987.
- [3] 小泉, 船城, 本田, 仁田, 塚本(編). 日本語基本動詞用法辞典. 大修館書店, 1989.
- [4] R-L. Liu and V-W. Soo. An empirical study on thematic knowledge acquisition based on syntactic clues and heuristics. In *Proceedings of the 31th Annual Meeting of ACL*, pp. 243-250, June 1993.
- [5] C. D. Manning. Automatic acquisition of a large subcategorization dictionary from corpora. In *Proceedings of the 31th Annual Meeting of ACL*, pp. 235-242, June 1993.
- [6] 国立国語研究所. 分類語彙表. 秀英出版, 1964.
- [7] 内元清貴. シソーラスに基づく動詞句用例の自動意味分類. 学士論文, 京都大学工学部, 1994.
- [8] T. Utsuro. *Lexical Knowledge Acquisition from Bilingual Corpora*. Doctorial Thesis, Kyoto University, 1993.
- [9] 清水護, 成田成寿(編). 和英辞典. 講談社学術文庫, 1979.