

## 動詞の多義性解消における 格の弁別能力と集中度の有効性について

藤井 敦\* 秋山 典丈 徳永 健伸 田中 穂積

東京工業大学 工学部

本論文は、システムに蓄えられた例文と入力文との間で、対応する格が取る名詞の類似度に基づいて動詞の多義性を解消する従来の手法に、「格の弁別能力」と「格の集中度」という二つの尺度を導入した新しい手法を提案する。「格の弁別能力」とは格が動詞の語義を弁別する能力を表す尺度であり、取り得る名詞の例(事例)が語義によって異なる格ほど語義の弁別能力は大きくなる。動詞の多義性解消において弁別能力の大きい格をより重視する。「格の集中度」とは、格が取る事例の意味的な広がりを表す。広がり大きい格は様々な名詞を取りやすいことから、その格が持つ事例と入力文中の名詞との間に、より大きな類似度を割り当てる。いくつかの動詞について新聞記事を中心とするコーパスを用いて実験を行い、二つの尺度を考慮した場合に動詞の多義性を解消する精度が向上することが確認された。

### 1 はじめに

動詞の多義性解消は自然言語解析における重要な処理の一つである。機械翻訳における訳語選択や、係り受け解析における名詞句の係り先の決定においても動詞の多義性解消が重要な役割を担う。

動詞の語義ごとに格フレームを記述したIPAL動詞辞書(以下IPALと略す)[1]の例文とシソーラスを用いて動詞の多義性を解消する新しい手法が黒橋らによって提案されている[2]。これは、入力文とIPALに登録された例文の間で対応する格が取る名詞の類似度を計算して最も類似度の高い格フレームを選択する方法である。本研究では、この方法に二つの新しい尺度を導入して動詞の多義性を解消する精度(解析精度)の向上を試みた。

第一に、動詞の語義を弁別する能力はガ格やヲ格などの各々の格によって異なることが指摘されている[3]。また、同様の概念として「かかりの広さ」が提案されている[4, 5]。これは、かかり要素がどの程度述部を限定するかという概念であり、述部を限定する程度が強いほどかかりの広さは狭くなる。しかし述部の持つ意味的曖昧性までは解消しない。本手法では、格が動詞の語義を弁別する能力を表す尺度として「格の弁別能力」を導入する。格の弁別能力とは、「かかりの広さ」を動詞の語義の弁別にまで拡張したものであり、動詞の語義を限定する力が強い格ほど弁別能力は

大きくなる。

直観的には、格が取り得る名詞が、動詞の語義によって差が無い格ほど弁別能力は小さくなる。例えばIPALにおいて動詞「かける」には34個のサブエントリがある<sup>1</sup>。「彼が眼鏡をかける」という文の「かける」は、そのうちの一つ「ひも状のような物を身体に付ける」のサブエントリに属する。この例文の動詞の語義を決めるのは「彼」よりも「眼鏡」であると考えられる。これは、「彼」という単語が「かける」の別の語義(水をかける、橋をかけるなど)のガ格にも現れることができるので、「彼」だけでは動詞の語義を決めるのが決められないのに対して、「眼鏡」は他の語義のヲ格には現れにくいので動詞の語義を決定しやすいのだと考えられる。つまり、この例では動詞「かける」のヲ格はガ格よりも弁別能力が高いと考える。

第二に、格が取り得る名詞間の意味的な広がりを表す尺度として「格の集中度」を導入する。名詞間の意味的な類似度が高いほど集中度が大きくなり、逆に類似度が低いほど、つまり広がり大きいほど集中度は小さくなる。集中度が小さい格は様々な意味の名詞を取る傾向があるといえる。

黒橋らの手法では、システムが持っている格ごとの名詞の例(以下、事例と記述する)の数が不十分である場合に、入力文中の名詞が、複数の語義の(対応する格の)事例との間で同じ類似度となり、多義性が解消できないという問題が生じる。例えば、動詞「くわえ

\*fujii@cs.titech.ac.jp

<sup>1</sup>漢字の表記の違いを含む。

る」について IPAL では5つのサブエントリが用意されている。そのうちの二つ「ある事に対して何らかの見解を付け足す」(語義1とする)と「これまであった物にさらに新しく何かを付け足す」(語義2とする)のヲ格について考えてみる。IPAL ではヲ格の事例として、それぞれ「配慮・解釈・検討・批判・説明」、「塩・新人・彼・(ゴッホの)絵」が挙げられている。今「得点をくわえる」(語義1に属する)という入力があったとき、「得点」と先に挙げた語義1、語義2の事例との類似度を分類語彙表を用いて計算すると、どちらの語義に対しても類似度が0となる<sup>2</sup>。

格の集中度を考慮すると、語義1の事例は互いに意味的な類似度が高いのに対して、語義2の事例には意味的な類似度が低く<sup>3</sup>、集中度が小さい。そこで、語義2のヲ格は様々な意味の名詞を取る傾向があると考え、「得点」と語義2の事例との間に、より大きい類似度を割り当てる。

本稿では、2節で黒橋らのアルゴリズムについて概説し [2]、3節で格の弁別能力について、4節では格の集中度について説明する。さらに5節ではコーパスを用いた評価実験について述べる。

## 2 基本アルゴリズム

動詞の多義性解消に関して、黒橋らは、格フレーム辞書として IPAL、シソーラスとして分類語彙表 [6] を用いた方法を提案している [2]。そのアルゴリズムを概説する。

- (1) 入力文の格要素と格フレームの事例の対応付けを行う。
- (2) 対応付けられたそれぞれの格要素について、入力文の名詞と格フレームの事例との間の類似度を計算する。類似度の値は分類語彙表における二つの名詞の分類コードの一致するレベルによって決まる。一致するレベルと類似度との関係を表1に示す。
- (3) 式1に従って格フレームと入力文との対応の評価値 (score) を計算し評価値の最も高い格フレームを選択する。

$$score = \begin{cases} 0 & (\text{if } l > n) \\ sum \times \sqrt{\frac{1}{n}} \times \sqrt{\frac{n}{m}} & (\text{otherwise}) \end{cases} \quad (1)$$

$n$ : 対応付けされた格要素数

$l$ :  $n +$  (入力文側の対応付けられていない格要素における必須要素数)

<sup>2</sup>黒橋らの手法による

<sup>3</sup>分類語彙表による

$m$ :  $n +$  (格フレーム側の対応付けられていない格要素における必須要素数)

$sum$ : 対応付けされた格要素の類似度の和

$l, m, n$  を図1に示す。長方形は名詞を表す。ガ・ニ・ヲ・ヘ・ヨリ格を必須要素とする (ただしニ・ヨリを伴う格要素が時を表す場合は任意要素とする)。

表1: 名詞  $X$  と  $Y$  の間の分類コードの一致レベルと類似度との関係

一致するレベル	0	1	2	3	4	5	6	一致
類似度 ( $sim(X, Y)$ )	0	0	5	7	8	9	10	11

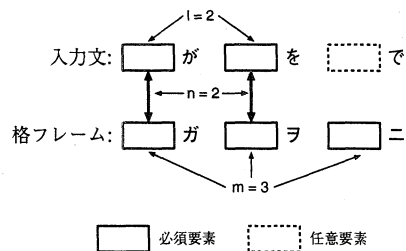


図1: 入力文と格フレームの対応付け

## 3 格の弁別能力

本論文で扱う記号を示す。

- $N_{s,k}$ : 語義  $s$ , 格  $k$  の事例の集合
- $N_{s,k}^i$ :  $N_{s,k}$  の  $i$  番目の要素
- $|N_{s,k}|$ :  $N_{s,k}$  の要素数
- $I_k$ : 入力文中の格  $k$  が取る名詞
- $sim(X, Y)$ : 名詞  $X$  と  $Y$  の間の分類コード (表1) に基づく類似度
- $d(k)$ : 格  $k$  の弁別能力
- $c(s, k)$ : 語義  $s$ , 格  $k$  の集中度

格の弁別能力とは格が動詞の語義をどの程度限定するかを表す尺度であり、格が取る事例が語義によって差が無いほど弁別能力は小さくなる。格の弁別能力は各動詞の格ごとに計算する。

$n$  個の語義を持つ動詞の格  $k$  の弁別能力  $d(k)$  を式2で与える。ただし、分類コードがレベル5まで一致する単語どうしを共通要素 ( $N_{i,k} \cap N_{j,k}$ ) とする。

$$d(k) = \frac{1}{n C_2} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{|N_{i,k}| + |N_{j,k}| - 2|N_{i,k} \cap N_{j,k}|}{|N_{i,k}| + |N_{j,k}|} \quad (2)$$

$N_{s,k}$ の中で $I_k$ との間の類似度が最も高い事例を $E_{s,k}$ とする。入力文に現れた格 $i$  ( $i = 1 \dots K$ ) について類似度の加重和を計算し、これを式1における $sum$ とする。

$$sum = \sum_{i=1}^K (exp(\alpha \cdot d(i)) \times sim(I_i, E_{s,i})) \quad (3)$$

$\alpha$  は定数である。評価値 ( $score$ ) の計算は式1と同じである<sup>4</sup>。

5節の実験に用いた動詞について、IPALの事例を用いて計算した格の弁別能力の値を表2に示す<sup>5</sup>。

表 2: 格の弁別能力

動詞	ガ	ニ	ヲ	大小関係
のる	0.895	0.950	—	ガ < ニ
わかる	1.00	0.480	—	ニ < ガ
とる	0.661	—	0.989	ガ < ヲ
おさめる	0.463	1.00	0.992	ガ < ヲ < ニ
くわえる	0.625	0.939	0.972	ガ < ニ < ヲ
あたえる	0.876	0.944	0.978	ガ < ニ < ヲ

## 4 格の集中度

格の集中度とは格が取る複数の事例が意味的にどの程度類似しているかを表す尺度であり、各語義の格ごとに計算する<sup>6</sup>。

$N_{s,k}$ 中の事例の全ての組について類似度を計算し、その平均を語義 $s$ 、格 $k$ の集中度と考え、式4で与える。 $m = |N_{s,k}|$ とする。

$$c(s, k) = \begin{cases} \frac{\sum_{i=1}^{m-1} \sum_{j=i+1}^m sim(N_{s,k}^i, N_{s,k}^j)}{mC_2} & (\text{if } m > 1) \\ \text{最大値} & (\text{if } m = 1) \end{cases} \quad (4)$$

$m = 1$ すなわち事例が一つしかない場合、集中度は類似度の最大値 (表1では11) とする。

集中度が小さい格は、より多くの名詞を取りやすいので、分類語彙表の分類コードを用いて計算した類似度 (2節参照) を格の集中度によって補正し、集中度の小さい格により高い類似度を与える。 $n$ を語義数とするとき、 $I_k$ と $E_{s,k}$ の間の補正後の類似度 $sim'(I_k, E_{s,k})$ を式5で計算する。 $\beta$ は定数である。

<sup>4</sup>ただし、入力文中の全ての格の名詞が例文の名詞と完全に一致した場合は格の弁別能力や集中度は考慮しない。

<sup>5</sup>必須要素のみ示す

<sup>6</sup>弁別能力は語義ごとではなく動詞ごとに計算することに注意。

$$sim'(I_k, E_{s,k}) = sim(I_k, E_{s,k}) + \beta \frac{1/c(s, k)}{\sum_{i=1}^n 1/c(i, k)} \quad (5)$$

## 5 評価

### 5.1 実験

本研究で提案した手法を評価するために、新聞記事を中心とするコーパスに含まれる動詞の多義性解消の実験を行った。コーパスはあらかじめ人手によって係り受けと語義を決定し、データとして用意する。実験に用いたデータを表3に示す。

表 3: 実験に用いたデータ

動詞	語義数 <sup>7</sup>	データ数
のる	10 (8)	126
わかる	5	60
とる	29 (18)	84
おさめる	8	108
くわえる	5 (4)	168
あたえる	4	528

これらのデータを動詞ごとに6等分し、そのうちの一つをテストデータ、残りの5つを訓練データ群とする。初回はIPALの例文を与えてテストデータの解析精度を調べる<sup>8</sup>。以降、訓練データ群をシステムに一つずつ与えて、事例・格の弁別能力・集中度を更新しながらテストデータの解析精度を調べる。図2に実験の概略を図示する。

テストデータを変えて同様の操作を6回行い、訓練回数ごとに平均を計算する。以上の操作によって、各動詞ごとに訓練回数と解析精度との関係を求める。

実験に用いた全ての動詞の多義性解消の精度 (解析精度) を平均したグラフを図3に示す。グラフの横軸はシステムの訓練回数、縦軸は解析精度 (%) を表す。比較のために以下の4通りの方法で実験を行った。グラフの番号はこれらの手法の番号に対応している。

1. 格の弁別能力・集中度の両方を考慮
2. 格の弁別能力のみを考慮
3. 格の集中度のみを考慮
4. どちらも考慮しない (黒橋らの手法)

<sup>7</sup>IPALにおける語義数で、「乗る」と「載る」のような漢字の表記の違いも含む。括弧内の数字は実験に使用したコーパスに出現した語義数。

<sup>8</sup>語義を一意に特定でき、尚かつそれが正解であるものを正解とし、(解析精度) = (正解数) ÷ (テストデータ数) で計算する。

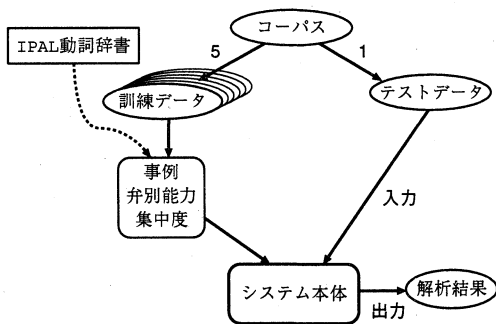


図 2: 実験の手順

個々の動詞ごとに0~5回の訓練における解析精度の平均を計算した例を表4に示す。1から4までの番号は図3における各手法と対応している。最後の列は手法1と4の解析精度の差を表している。

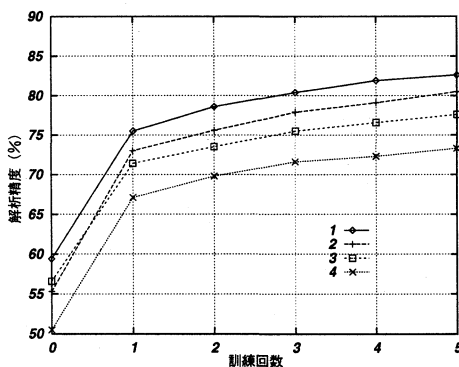


図 3: 訓練回数と解析精度 (%) との関係

表 4: 動詞ごとの解析精度 (%)

動詞	4	3	2	1	1-4
のる	73.8	79.5	74.9	79.6	5.8
わかる	50.8	56.9	55.8	65.0	14.2
とる	47.8	50.0	63.7	64.9	17.1
おさめる	61.4	61.6	65.3	65.4	4.0
くわえる	73.8	79.2	78.8	81.2	7.4
あたえる	70.1	75.0	76.9	79.5	9.4

## 5.2 考察

今回の実験によって以下のことが確認された。

- 本研究で提案した格の弁別能力・集中度の導入によって動詞多義性解消の解析精度が向上する。表4において、最大で17.1%精度が向上した（動詞「とる」）。
- 本手法は、事例データが少ない初期の段階から解析精度を向上させることができる。図3において、手法1では1回の訓練で手法4の最終的な解析精度を上回っている。事例ベースのシステムにおいて事例データが十分得られないような場合に本手法は有効であると考えられる。
- 事例データの増加とともに解析精度が向上する。

## 6 おわりに

本研究では、従来の動詞多義性解消の手法に格の弁別能力と集中度を導入し、評価実験を行うことでその有効性を示した。

今後の研究課題として、本論文の $\alpha$ 、 $\beta$ 等の定数の選び方も含め、弁別能力や集中度の計算式についてさらに検討する予定である。

コーパスでは格の省略が頻繁に起こる。格の省略をどのように扱うかについても検討の余地がある。

また、分類語彙表は名詞の多義性についてあまり考慮されておらず、規模としても大きくないことから、より大規模なシソーラスを利用して名詞の類似度をより精密に計算することなども今後の課題である。

## 参考文献

- [1] 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, 1987.
- [2] Sadao Kurohashi and Makoto Nagao. A Method of Case Structure Analysis for Japanese Sentences Based on Examples in Case Frame Dictionary. *IEICE TRANSACTIONS on Information and Systems*, Vol. E77-D, No. 2, pp. 227-239, February 1994.
- [3] 益岡隆志. 命題の文法—日本語文法序説. くろしお出版, 1987.
- [4] 佐伯哲夫. 現代日本語の語順. 笠間書院, 1975.
- [5] 徳永健伸, 岩山真, 乾健太郎, 田中穂積. 日本語語順の推定モデルとその応用. 情報処理学会 自然言語処理研究会, Vol. 81, No. 2, pp. 9-16, 1991.
- [6] 国立国語研究所(編). 分類語彙表. 秀英出版, 1964.