類推に基づく語の重み付け学習を用いた動詞の多義解消

福本文代† 鈴木良弥†

本稿では，類推により語義の特徴を抽出した結果を用いて，文中に含まれる多義動詞の曖昧さを解消する手法を提案する．本手法では，多義であるかどうかを判定しながら意味的なクラスタリングを行うことで多義解消に必要な情報を抽出する．そこで，表層上は1つの要素である多義動詞を，その動詞が持つ各意味がまとまった複数要素であるととらえ，これを1つ1つの意味に対応させた要素に分解したうえでクラスタを作成するという手法を用いた．我々の手法における学習とは，多義語を含む動詞グループに対し，クラスタリングを行う過程で，類推を用いることで語義の特徴を示す語（重要語と呼ぶ）を正しく判定することである．その結果，重要語には高い重み付けを行い，重要でない語に対しては低い重み付けを行う．この処理は動詞グループ中のすべての動詞が正しくクラスタリングされるまで繰り返される．本手法の精度を検証するために *Wall Street Journal* から抽出した多義語を含む9,706文に対して実験を行った結果，7,572文の正解が得られ，正解率は78.6%に達した．

Term Weight Learning Based on Analogy for Word Sense Disambiguation

FUMIYO FUKUMOTO† and YOSHIMI SUZUKI†

In this paper, we propose a method for disambiguating verbal word senses using term weight learning based on analogy. The characteristics of our approach are (1) our algorithm explicitly introduces new entities called *hypothetical verbs* when an entity is judged polysemous (2) *keywords* which characterise every sense and co-occur with polysemous verb are extracted using analogy of the relation of co-occurrence of two words. For the results, term weight learning is performed. Parameters of term weighting are then estimated so as to maximise the keywords and minimise the other words which co-occur with polysemous verb. The results of experiment demonstrate the effectiveness of the method.

1. はじめに

自然言語処理における重要な問題の1つに，形態・構文・意味といった言語に関する様々な曖昧さの問題がある．一般に，意味的な曖昧さを解消するためには，意味に関する様々な情報を規則化し記述しておく必要がある．しかし，意味は文脈に依存して決まるため，あらゆる文脈に対応できるすべての意味をあらかじめ規則として網羅的に記述しておくことは難しい．英語では，Collins English Dictionary，Roget のシソーラス，また日本語では分類語彙表，結合価パターン辞書など，機械可読辞書として電子化されたものがあるが，辞書の記述は語の定義が言語学者によりまちまちであるため，現実の文に対処できる有用な意味情報が得られるかどうか，検証の余地がある．

† 山梨大学工学部コンピュータ・メディア工学科
Department of Computer Science and Media Engineering, Faculty of Engineering, Yamanashi University

近年，電子化された大量のコーパスが流通するようになったことを背景に，コーパスから自動的に抽出した情報を用いて語義の曖昧さを解消する研究が多く行われている[1),3),7),9),10),15),16),20),23),25),26)]．コーパスを用いて語義の曖昧さを解消する場合，その精度は，どのようにして多義語が持つ各意味を特徴付け，コーパスから自動的に抽出するかに依存する．

Brown ら[3)]は対訳テキストを用い，一方の言語の語義の曖昧さを他方の語の情報を利用することで解消する手法を提案し，実際に英仏機械翻訳システムに適用し，検証を行っている．しかし，彼らは問題点として，(1) 多義語が持つ意味をあらかじめ2つに限定している，(2) 膨大な対訳テキストを必要とする，をあげている．

Zernik[30)]や Schütze ら[24)]は，単一言語コーパスを用いて多義語を解消する手法を提案している．彼らは，多義語が持つ各意味をそれと共起する名詞を用いて特徴付けている．すなわち，クラスタリングアルゴリズ

ムを用いることで多義語と共起する名詞の集合を分類している．しかし，分類された名詞の集合と多義語が持つ各意味とを対応させる処理は人手で行っているため，多義語の解消は人間の言語的な直観に頼ることになってしまう[13]．

Yarowsky らは，Roget のシソーラスカテゴリを利用し，統計手法を用いることでテキスト中に現れる多義語の曖昧さを解消する手法を提案した．彼らは多義語が属するシソーラスカテゴリを用いることで多義語が持つ各意味を特徴付けている．彼らの手法は，統計情報を用いてシソーラスカテゴリに出現する単語に重み付けを行った後，その結果を利用して多義語の周辺語の重みの和から多義語がどのシソーラスカテゴリに属するかを決定するというものである．この手法を12の多義語名詞に適用し実験を行った結果，平均解消率92%という高い正解率が得られることが報告されている[27]．しかし，Yarowsky らのシソーラスを用いる問題として，データスパースネスの問題が指摘されている[22]．すなわち，シソーラスカテゴリに示されている語が，文書の種類によっては出現しない場合がある．

丹羽らは，文脈を構成する単語をベクトルで表現し，文脈をそれらベクトルの和で表した．任意の文脈 A における単語の意味は，多義語の各意味を表す文脈例を各意味に応じてあらかじめ用意しておき，各々の例と文脈 A における単語の意味との類似度（内積）を計算し，その値が最も大きい文脈が示す意味であるとした．この手法を語義数2の名詞の多義判定に適用した結果，平均80%の正解率が得られている[21]．しかし，多義語の各意味を示す文脈例，すなわちトレーニングデータは人手によりあらかじめ用意しておかなければならないため，大量のデータに対して解消を行う場合には労力を要する．Ng らはこの問題に対処するため，例示による学習を取り入れた多義解消手法を提案している[20]．このアプローチはトレーニングデータとして共起関係，多義語の周辺語の情報に加え，文の構文的な関係を用いることで曖昧さの解消を行っている．しかし，一般に統計手法を用いてコーパスから情報を得ようとする場合，大量のデータが必要となる．Ng らのようにコーパスとして構文解析結果を利用する場合，ロバストな解析システムが必要となる．

本稿では，語義の特徴を示す語に対し，重み付けの学習を行った結果を用いて文中に含まれる多義語の曖昧さを解消する手法を提案する．本手法では，Schütze や Zernik らの手法と同様，単一言語コーパスを用い，多義語が持つ各意味をそれと共起する名詞を用いて表現する．しかし，彼らの手法がクラスタリングにより分類された名詞の集合と，多義語が持つ各意味とを人手により対応させているのに対し，多義語が持つ各意味は，それと対応する動詞に割り当てているため，意味を判断するうえで人間の介在を必要としない．本手法では，多義であるかどうかを判定しながら意味的なクラスタリングを行うことで多義解消に必要な情報を抽出する．そこで，表層上は1つの要素である多義動詞を，その動詞が持つ各意味がまとまった複数要素であるととらえ，これを1つ1つの意味に対応させた要素に分解したうえでクラスタを作成するという手法を用いた．

本手法では，動詞と名詞の意味的な関係の度合いは相互情報量を用いて表される．一般に相互情報量の値はそれが高い値を持つ場合には2語間に強い意味的な関係が存在する．しかし，高い値を設定し，その値以上の動詞と名詞の共起関係を抽出すると，Yarowsky らと同様，データスパースネスの問題が生じる．すなわち，解消を行おうとする多義語と共起する名詞がトレーニングデータに出現しない場合が生じる．一方，低い値も含めてクラスタリングを適用すると，多義語と，それと共起する名詞との間に意味的な関係が保証されないため，結果的に精度が下がる．我々はこの問題に対処するために，Ng らの手法と同様，類推により低い相互情報量の値を持つ名詞が動詞の意味を特徴付けるか否かを判定するという手法を用いた．我々の手法における学習とは，多義語を含む動詞グループに対し，クラスタリングを行う過程で，類推を用いることで語義の特徴を示す語（重要語と呼ぶ）を正しく判定することである．その結果，重要語には高い重み付けを行い，重要でない語に対しては低い重み付けを行う．この処理は動詞グループ中のすべての動詞が正しくクラスタリングされるまで繰り返される．

以下，2章では，クラスタリングの観点から多義語を定義し，3章では類推を用いて重要語か否かを判定する手法について述べる．4章ではコーパスから多義解消に必要な情報を抽出する手法について述べる．5章では得られた情報を用いて文中に含まれる多義語の曖昧さを解消する手法について述べる．6章では本手法の有効性を検証するために行った実験とその結果について報告する．

2. 仮想動詞

一般に，意味的に近い2つの動詞は同じ名詞と共起して現れる[4),5),12),30)]．(1)～(2') は *Wall Street Journal* から抽出した例文を示す[14]．

(1) In the past, however, coke has typically taken

a minority stake in much ventures.

(1') Guber and Peters tried to buy a stake in Mgm in 1988.

(2) That process of sorting out specifies is likely to take time.

(2') We spent a lot of time and money in building our group of stations.

(1)〜(2') において，(1) および (1') に現れる take と buy はともに stake と共起して現れ，両者はほぼ同じ意味を持つ．同様に (2) および (2') に現れる take と spend はともに time と共起して現れ，両者は同じ意味を持つ．したがって，多義語 take が持つ複数の意味は，動詞 buy と共起して現れる名詞 stake，および spend と共起して現れる名詞 time と特徴付けて考えることができる．すなわち，多義語を含む文において，もし多義語と共起する名詞のうち少なくとも 1 つが多義語の意味を特徴付ける名詞と同じ（あるいは名詞の集合に属する）ならば，文中の多義語の意味はその名詞と共起する動詞の意味に同定することができる．我々は，文中に現れる多義語の曖昧さを，その語と共起する名詞を用いることで解消した．

本手法では，語義を判定しながら意味的なクラスタリングを行うことで多義語の曖昧さ解消に必要な情報，すなわち，多義語の意味を特徴付ける重要語の集合を抽出する．そこで，表層上は 1 つの要素である多義語を，多義語が持つ各意味がまとまった複数要素であるととらえ，これを 1 つ 1 つの意味に対応させた要素（**仮想動詞**ベクトルと呼ぶ）に分解したうえでクラスタを作成するという手法を用いた．

我々は，動詞をベクトルととらえ，動詞と共起する n 個の名詞を軸とする n 次元名詞空間上でこれを表した．軸 i $(1 \leq i \leq n)$ に射影した動詞ベクトルの長さは，i 軸で示される名詞と動詞の相互情報量の値を用いた．仮に 2 つの動詞に多義性がなく，かつこの 2 つの動詞が意味的に近いとすると，これらの動詞はこの空間上で互いに距離が近いため，同一のクラスタに含まれることになる．一方，(1) と (2) に現れる take は多義語であるため，各意味を表す動詞ベクトル buy と spend のいずれともクラスタを構成しなければならない．そこで，ベクトル take を各軸に従って（この場合，stake と time の 2 軸）分割することを考える．ベクトル take を stake と time の軸に従って分割した結果を**図 1** に示す．

図 1 において，ベクトル take は，stake と time の軸上でベクトル **take1** と **take2** に分割されている．**take1** と **take2** を**仮想動詞**ベクトルと呼ぶ．図 1 は仮想動詞ベクトルを導入することで，各々意味的に近い要素から成る 2 つのクラスタ {**take1**,buy},{**take2**,spend} が得られることを示す．

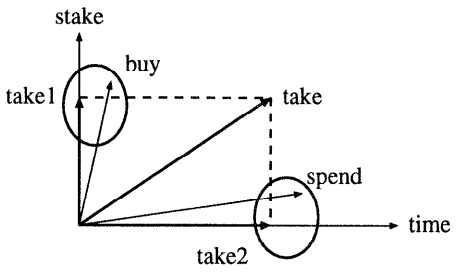


図 1 動詞 take の仮想動詞への分割
Fig. 1 The decomposition of the verb take.

表 1 {**close**, open, end} のクラスタリング結果
Table 1 The result of the set {**close**, open, end}.

v_i	n	Mu(v_i,n)
close1	account	2.116
(open)	acquisition	1.072
	banking	2.026
	book	4.427
	bottle	3.650
	...	...
close2	announcement	1.692
(end)	connection	2.745
	conversation	4.890
	period	1.876
	practice	2.564
	...	...

3. 類推に基づく重要語の判定

m 個から成る動詞グループを $\mathrm{VG} = \{v_1, \cdots, v_m\}$ とすると，クラスタリングアルゴリズムは動詞グループの意味的な偏差を比較し，偏差の少ない順にクラスタを生成する．仮に，v_i が多義語でない場合には v_i を含むクラスタが作成される．一方，v_i が多義語である場合には，v_i は仮想動詞に分割され，各仮想動詞は少なくとも 1 つのクラスタに属するようなクラスタが作成される．**表 1** は {**close**, open, end} のクラスタリング結果を示す．

表 1 において，'open'，'end' は，'**close**' の語義を示す．$\mathrm{Mu}(v_i, n)$ は動詞 v_i と名詞 n の相互情報量の値を示し，式 (1) で示される．

$$\mathrm{Mu}(x, y) = \log_2 \frac{P(x, y)}{P(x)P(y)} \tag{1}$$

$P(x)$ および $P(y)$ は，x の頻度数 $f(x)$ および y の頻度数 $f(y)$ をそれぞれコーパスに出現する語の総数 N で正規化したものであり，$P(x, y)$ は x と y の共

begin
(a) **for all** $n_q \in N\_Set_1$ - N_Set_3 *such that* $\mathrm{Mu}(w_p, n_q) < 3$
$\mathrm{Mu}(w'_{pi}, n_q) \geq 3$ となるような w'_{pi} $(1 \leq i \leq s)$ を抽出する．ここで，s は名詞 n_q と共起する動詞の個数とする．
for all w'_{pi}
if w'_{pi} exists *such that* $Sim\ (w_p, w'_{pi}) \geq 0$
then n_q は重要語であり，v と n_q，および w_p と n_q の相互情報量のパラメータを α $(1 < \alpha)$ とする．
end_if
end_for
end_for
(b) **for all** $n_r \in N\_Set_3$ *such that* $\mathrm{Mu}(w_p, n_r) \geq 3$ and $\mathrm{Mu}(w_1, n_r) \geq 3$
$\mathrm{Mu}(w'_{pi}, n_r) \geq 3$ となるような w'_{pi} $(1 \leq i \leq t)$ を抽出する．ここで，t は名詞 n_r と共起する動詞の個数とする．
for all w'_{pi}
if w'_{pi} exists *such that* $Sim\ (w_p, w'_{pi}) \geq 0$ and $Sim\ (w_1, w'_{pi}) \geq 0$
then n_r は重要語ではなく，v と n_r，w_p と n_r，および w_1 と n_r の相互情報量のパラメータを β $(0 < \beta < 1)$ とする．
end_if
end_for
end_for
end

図 2 類推に基づく重要語の判定
Fig. 2 Recognition of keywords based on analogy.

起頻度数 $f(x,y)$ を N で正規化したものである．

文中に現れる動詞の多義解消は名詞 n を用いて行われる．すなわち，表 1 に示されている名詞が文中に現れる動詞（**close**）と共起するとき，文中の動詞（**close**）は，その名詞と共起する仮想動詞の意味となる．

我々の手法における学習とは，多義語を含む動詞グループに対し，クラスタリングを行う過程で，類推を用いることで語義の特徴を示す語（重要語と呼ぶ）を正しく判定することである．その結果，重要語には高い重み付けを行い，重要でない語に対しては低い重み付けを行う．この処理は動詞グループ中のすべての動詞が正しくクラスタリングされるまで繰り返される．重要語の判定は，クラスタリングの入力である各動詞と，それと共起する名詞との相互情報量が一定値☆よりも低い名詞に対して行われる．これは，動詞と名詞との相互情報量の値が小さい場合には，それが対象としているコーパスに出現しなかったか，あるいは意味的な関係が成り立たないかが明らかでないためである[6]．我々は多義語と，それと共起する名詞に対して類推を用いることでこの判定を行った．x と y に共起関係が存在することを (x,y) で示す．今，(w_p, n_q) かつ，(w'_{pi}, n_q) であるとする．仮に w'_{pi} と n_q に意味的な関係があり，(w_p, n_q) と (w'_{pi}, n_q) が意味的に類似しているとき，w_p と n_q は意味的な関係があると見なす．

今，動詞 v は 2 つの意味，w_p と w_1 を持つとする．N_Set_1 を v と共起する名詞と w_p と共起する名詞の共通集合とする．同様に，N_Set_2 を v と共起する名詞と w_1 と共起する名詞の共通集合とし，N_Set_3 を N_Set_1 と N_Set_2 の共通集合とする．類推に基づく重要語の判定アルゴリズムを**図 2** に示す☆☆．

図 2 において，(a) は低く重み付けされている重要語を抽出する処理を示し，(b) は高く重み付けされている重要でない語を抽出する処理を示す．図 2 の $Sim(v_i, v_{i'})$ は v_i と $v_{i'}$ の類似度を示し，式 (2) で表される．

$$Sim(v_i, v_{i'}) = \frac{v_i \cdot v_{i'}}{|\, v_i \,||\, v_{i'} \,|} \tag{2}$$

$$\text{ここで } v_i = (\mathrm{Mu}(v_i, n_1), \cdots, \mathrm{Mu}(v_i, n_k)) \tag{3}$$

ただし

$$\mathrm{Mu}(v_i, n_j) = \begin{cases} \mathrm{Mu}(v_i, n_j) & \text{if } \mathrm{Mu}(v_i, n_j) \geq 3 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

式 (3) において，k は v_i および $v_{i'}$ と共起する名詞の異なり数を示す．$\mathrm{Mu}(v_i, n_j)$ は v_i と n_j の相互情報量の値を示す．

図 2 の (a) において $\mathrm{Mu}(w'_{pi}, n_q) \geq 3$ のとき，w'_{pi} と n_q は意味的な関係があるとした．同様に

☆ 実験では相互情報量が 3 よりも小さい値とした．

☆☆ 図 2 において w_1 の場合には，w_p と w_1，$n_q \in N\_Set_1$ - N_Set_3 と $n_q \in N\_Set_1$ - N_Set_2，そして，$Sim(w_p, w'_{pi}) \geq 0$ と $Sim(w_1, w'_{pi}) \geq 0$ をそれぞれ置き換える．

```
begin
  ICS := Make-Initial-Cluster-Set(VG)

  ( VG = {v_i | i = 1, ..., m}    ICS = {Set_1, ..., Set_{m(m-1)/2}}
    ただし Set_p = {v_i, v_j} と Set_q = {v_k, v_l} ∈ ICS (1 ≤ p < q ≤ m) は
    Dev(v_i, v_j) ≤ Dev(v_k, v_l) を満たす.
  for i := 1 to m(m-1)/2 do
    if CCS = φ
        then Set_γ := Set_i
      i.e. Set_i は新たに得られるクラスタとしてCCSに蓄積される
    else if Set_α ∈ CCS exists such that Set_i ⊂ Set_α
        then Set_i がICSから排除され, Set_γ := φ となる.
    else if
        for all Set_α ∈ CCS do
          if Set_i ∩ Set_α = φ
            then Set_γ := Set_i  i.e. Set_i は新たに得られるクラスタとしてCCSに蓄積される.
          end_if
        end_for
    else Set_β := Make-Temporary-Cluster-Set(Set_i,CCS)
         ( Set_β := Set_α ∈ CCS such that Set_i ∩ Set_α ≠ φ
           (Set'_i,Set'_β) := Recognition-of-Keywords(Set_i,Set_β)
           Set_γ := Recognition-of-Polysemy(Set'_i,Set'_β)
    end_if
    end_if
    end_if
    if Set_γ = VG
      then for_loopを抜ける.
    end_if
  end_for
end
```

図3 アルゴリズムの流れ
Fig. 3 Flow of the algorithm.

$Sim(w_p, w'_{pi}) \geq 0$ のとき (w_p, n_q) と (w'_{pi}, n_q) は意味的に類似しているとした．類推を用いて重要語を抽出した結果，w_p が重要語の場合には高い重み付け $(\alpha \times \mathrm{Mu}(w_p, n_q), 1 < \alpha)$ を行い，重要語でない場合には低い重み付け $(\beta \times \mathrm{Mu}(w_p, n_q), 0 < \beta < 1)$ を行った☆．

4. 学習に基づくクラスタリング

クラスタリングアルゴリズムは動詞グループの意味的な偏差を比較し，偏差の少ない順にクラスタを作成する．今，m 個から成る動詞グループを VG=$\{v_1, \cdots, v_m\}$ とすると，VG の偏差 Dev(VG) は式 (5) で示される．ただし，n は動詞と共起する名詞の個数とする．

$$Dev(\mathrm{VG}) = \frac{1}{\mid \bar{g} \mid (\beta * m + \gamma)} \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} (v_{ij} - \bar{g}_j)^2} \quad (5)$$

式 (5) の $\bar{g}_j$ は j 軸での重心の値を示す．また，$\mid \bar{g} \mid$ は重心ベクトルの長さを示す．式 (5) の v_{ij} は式 (4) で示される値とする．β と γ は，動詞の偏差を示す値が動詞の個数に比例して増加することを防ぐために最小2乗法を用いて行った正規化である☆☆．式 (5) はその値が小さいほど，より偏差が少ないことを示す．

図3 はクラスタリングアルゴリズムの流れを示す．図3 の‘(’はその上で示される関数の処理を示す．図3 において，関数 **Make-Initial-Cluster-Set** は，動詞グループ VG を入力とし，VG の任意の動詞対の組合せに対し，意味的な偏差の値を計算し，任意の動詞対と偏差の値をその値が昇順になるように出力する．この結果を ICS（Initial Cluster Set）と呼ぶ．

CCS（Created Cluster Set）は作成されたクラスタの集合を示す．関数 **Make-Temporary-Cluster-Set** は Set_i のどちらか一方の動詞を含むクラスタを CCS から抽出する．Set_i と **Make-Temporary-Cluster-Set** の結果である Set_β に対し，重要語を判定する関数 **Recognition-of-Keywords** が適用さ

☆ 実験では α の増分，β の減分をともに 0.001 とした．

☆☆ *Wall Street Journal* を用いた実験では，$\beta = 0.964$, $\gamma = -0.495$ を得た．

れ，その結果が関数**Recognition-of-Polysemy**に渡される．関数**Recognition-of-Polysemy**は動詞が多義語か否かを判定する関数である．

今 Set'_i と Set'_β の両方に属する動詞を v とする．v が多義語であり，w_p（ただし，w_p は Set'_i の要素とする）と w_1（ただし，w_1 は Set'_β の要素とする）の意味を持つか否かを判定するために，式(6)，(7)で示されるクラスタが作成される．

$$\{v_1, w_p\}, \{v_2, w_1, \cdots, w_n\} \quad (6)$$

$$\{v, w_1, \cdots, w_p, \cdots, w_n\} \quad (7)$$

ここで，v と w_p は動詞を示し，$w_1, \cdots, w_n$ は動詞か仮想動詞を示す．式(7)の $w_1, \cdots, w_p, \cdots, w_n$ は $Dev(v, w_i) \leq Dev(v, w_j)$ $(1 \leq i \leq j \leq n)$ を示す．式(6)の v_1 と v_2 は v の語義に対応した仮想動詞を示す．

式(6)および式(7)に対して，式(8)，(9)をともに満たすように α と β が推定される．

$$Dev(v_1, w_p) \leq Dev(v, w_1, \cdots, w_p, \cdots, w_n) \quad (8)$$

$$Dev(v_2, w_1, \cdots, w_n) \leq Dev(v, w_1, \cdots, w_p, \cdots, w_n) \quad (9)$$

この処理は新しく得られるクラスタ Set_γ がVGと等しくなるか，あるいはICSの要素がなくなるまで適用される．

5. 多義の解消

文中の多義語 v の意味は，クラスタリング結果を示す表1を用いて決定される．今 v の語義を $v_1, v_2, \cdots, v_m$ とする．v の意味は $\sum_j^t Mu(v_1, n_j), \cdots, \sum_j^t Mu(v_m, n_j)$ の中で $\sum_j^t Mu(v_i, n_j)$ の値が最も大きいとき，v_i となる．ただし t は文中に存在する名詞のうち，動詞 v と共起する名詞の個数を示す．

6. 実　　験

実験では，40の動詞グループに対し，クラスタリングアルゴリズムを適用した結果得られた情報を用い，曖昧さがどの程度解消できるか検証した．

6.1 デ ー タ

トレーニングデータは，1989年のタグ付けされた *Wall Street Journal* であり，182,992文，総数2,878,688語から成る．このコーパスからウィンドウサイズを5語にとり，総数5,940,193個から成る任意の2語対を得た．ここで単語 x と y のウィンドウサイズが5語であるとは，x の出現位置から x の後方5語以内に現れる単語 y と x との対を示す．実験では動詞と名詞の対を使用した．これは5語という比較的小さいウィンドウサイズでは動詞と目的語という観点から動詞と名詞の意味的な関係が顕著に現れると考えられるためである．名詞と動詞に対し，派生形は原形に置き換えた．たとえば，名詞の'books'は'book'に置き換え，動詞の'buys'，'bought'，'buying'は'buy'に置き換えた．その結果，479,672個から成る2語対を得た．これらすべてに対し，相互情報量を計算した．

テストデータとして我々は，2，3および4の語義数を持つ動詞グループをそれぞれ，20，10および10個作成し，合計40の動詞グループに対し，クラスタリングアルゴリズムを適用した．語義の選定は，Wordnet 1.5[18]を使用し，各動詞に対し，その語義の区別が容易に判断できるものを抽出し，動詞の語義として割り当てた．

クラスタリングの実験では，語の重み付け手法の有効性を検証するために，重み付けを行わずにクラスタリングを行った場合と本手法とを比較した．

Wall Street Journal の各記事はそれぞれ78種類の分野に分類されている．曖昧さの解消実験では，語義の偏りを防ぐため，均等かつランダムに各分野からそれぞれ最大5文ずつ抽出し，総数9,706文をテスト文として用いた．

6.2 実 験 結 果

実験結果をそれぞれ**表2**，**表3**および**表4**に示す．表2，3および4において $\{v, w_p, w_1\}$，$\{v, w_p, w_1, w_2\}$，$\{v, w_p, w_1, w_2, w_3\}$ は，それぞれ動詞グループを示す．v は多義語を示し，w_p，w_1，w_2 および w_3 は v が持つ各意味を示す．たとえば，(1)の'**cause**'は'effect'と'produce'の語義を持つ．'Sentence'は多義語を含む文の総数を示し，w_p，w_1，w_2 および w_3 は多義語を含む文においてそれぞれの語義で用いられている文数とその割合を示す．v はトレーニングデータにおける多義動詞の総数を示す．$v \cap W$ において，W は w_p，w_1，w_2，あるいは w_3 と共起する名詞の集合を示し，$v \cap W$ は v と共起する名詞と W との共通集合の要素数を示す．'Correct'は正解数を示す．'Total (2 senses)'，'Total (3 senses)'，'Total (4 senses)'および表4の最下行の'Total'における'Sentence'はそれぞれ語義数が2，3，4および動詞グループ40の総文数を示し，w_p，w_1，w_2 および w_3 は，各動詞グループにおいて文中に出現する語義数が最も多い語義の総平均を表す．最下行の'Total'における'Correct'は全正解数を示す．'Learning'における'○'は，重み付けの学習を行わずに正解が得られたことを示し，×は得られなかったことを示す．

表2 語義の解消結果（語義数2）
Table 2 The result of disambiguation experiment (two senses).

					本手法		重み付けなし
Num	$\{v, w_p, w_1\}$	Sentence	w_p / w_1	v	$v \cap W$	Correct (%)	
(1)	{**cause**, effect, produce}	232	89 (38.3) / 143 (61.7)	464	254	158 (68.1)	×
(2)	{**claim**, require affirm}	202	158 (78.2) / 44 (21.8)	545	328	160 (79.2)	×
(3)	{**close**, open, end}	206	92 (44.6) / 114 (55.4)	2,025	1,498	162 (78.6)	×
(4)	{**fall**, decline, win}	278	182 (65.5) / 96 (34.5)	697	393	215 (77.3)	○
(5)	{**feel**, think, sense}	280	178 (63.5) / 102 (36.5)	210	53	156 (55.7)	×
(6)	{**hit**, attack, strike}	250	156 (62.4) / 94 (37.6)	452	122	181 (72.4)	○
(7)	{**leave**, remain, go}	183	41 (22.4) / 142 (77.6)	1,409	942	160 (87.4)	×
(8)	{**lose**,win, get}	301	70 (23.3) / 231 (76.7)	2,244	1,920	242 (80.3)	○
(9)	{**manage**, accomplish, operate}	216	23 (10.6) / 193 (89.4)	648	306	165 (76.3)	×
(10)	{**occur**, happen, exist}	199	137 (68.9) / 62 (31.1)	324	180	175 (87.9)	○
(11)	{**order**, request, arrange}	240	206 (85.8) / 34 (14.2)	1,256	831	202 (84.1)	○
(12)	{**pass**, adopt, succeed}	301	175 (58.1) / 126 (41.9)	378	247	259 (86.0)	×
(13)	{**post**, mail, inform}	274	63 (22.9) / 211 (77.1)	91	68	231 (84.3)	○
(14)	{**produce**, create, grow}	231	129 (55.8) / 102 (44.2)	640	370	184 (79.6)	○
(15)	{**push**, attack, pull}	223	93 (41.8) / 130 (58.2)	202	153	189 (84.7)	×
(16)	{**save**, keep, rescue}	216	149 (68.9) / 67 (31.1)	396	354	168 (77.7)	×
(17)	{**ship**, put, send}	218	92 (42.2) / 126 (57.8)	361	241	176 (80.7)	○
(18)	{**stop**, end, move}	244	52 (21.4) / 192 (78.6)	993	886	210 (86.0)	○
(19)	{**add**, append, total}	184	35 (19.0) / 149 (81.0)	1,164	778	147 (79.8)	×
(20)	{**keep**, maintain, protect}	378	231 (61.1) / 147 (38.9)	1,266	905	349 (92.3)	×
Total (2 senses)		4,856	3,332 (68.6)			3,889 (80.0)	

6.3 考　　察

- 手法の有効性

表2，3および4によると9,706文のうち7,572文が正解であり，正解率は78.6%に達した．語義数による正解率の内訳は，2，3，4語義でそれぞれ，80.0%, 77.7%, 76.4%であり，正解率に顕著な差がみられなかった．

w_p，w_1，w_2 および w_3 の Total，すなわち語義の出現頻度数のみで語義を決定した場合の平均正解率が58.2%であり，本手法の平均が78.6%であることから，本手法による効果は，語義を出現頻度だけで決定する場合と比較すると20.4(78.6−58.2)%となる．

'Learning' の結果において40の動詞グループのうち，31の動詞グループは語の重み付けの学習を適用しなければ語義の判定が正しく行われなかった．

表5は，各動詞グループにおける類推の結果を示す．

表5は動詞 v と W の相互情報量の値が $\mathrm{Mu}(v, W) < 3$ を満たす動詞と名詞の組に対し，類推を行った結果である．表5において $\mathrm{Mu}(v, W) < 3$ は，動詞と名詞の相互情報量が3より小さい名詞の個数を示す．表5によると $v \cap W$ の総数22,148に対して，類推を行う前は，$\mathrm{Mu}(v, W) < 3$，すなわち，相互情報量の値が3より小さい動詞と名詞の対が9,444組存在した．類推の結果，正しく類推が行われた組は7,672組存在し，このうち重要語と判定された組は，5,308組得られたことから，相互情報量が3以上の組12,704（22,148-9,444）に対し，多義語が持つ各意味を特徴付ける名詞が，さらに5,308個抽出で

表3 語義の解消結果（語義数3）
Table 3 The result of disambiguation experiment (three senses).

Num	$\{v, w_p, w_1\ w_2\}$	Sentence	w_p / w_1 / w_2	v	本手法 $v \cap W$	本手法 Correct (%)	重み付けなし
(21)	{**catch**, acquire, grab, watch}	240	120 (50.0) 21 (9.0) 199 (41.0)	447	432	180 (75.0)	×
(22)	{**complete**, end, develop, fill}	365	107 (29.3) 242 (66.3) 16 (4.4)	727	450	280 (76.7)	×
(23)	{**gain**, win, get, increase}	334	47 (14.0) 228 (68.2) 59 (17.8)	527	467	270 (80.8)	×
(24)	{**grow**, increase, develop become}	310	68 (21.9) 132 (42.5) 110 (35.6)	903	651	241 (77.7)	×
(25)	{**operate**, run, act, control}	232	76 (32.7) 83 (35.7) 73 (31.6)	812	651	187 (80.6)	×
(26)	{**rise**, increase, appear, grow}	276	51 (18.4) 137 (49.6) 88 (32.0)	711	414	198 (71.7)	×
(27)	{**see**, look, know, feel}	318	128 (40.2) 162 (50.9) 28 (8.9)	1,785	934	263 (82.7)	×
(28)	{**want**, desire, search, lack}	267	66 (24.7) 53 (19.8) 148 (55.5)	590	470	208 (77.9)	×
(29)	{**lead**, cause, guide, precede}	183	139 (75.9) 38 (20.7) 6 (3.4)	548	456	138 (75.4)	×
(30)	{**carry**, bring, capture, behave}	186	142 (76.3) 39 (20.9) 5 (2.8)	474	440	142 (76.3)	×
Total (3 senses)		2,711	1,573 (56.5)			2,107 (77.7)	

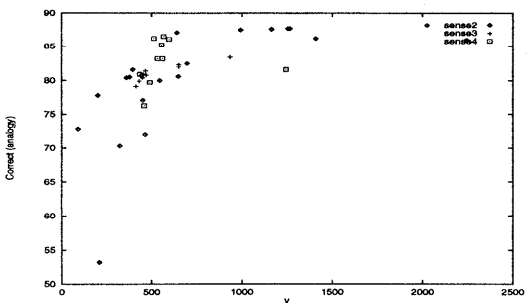


図4 v の個数と類推の正解率との関係
Fig. 4 The relationship between frequency of v and correct ratio of analogy based estimation.

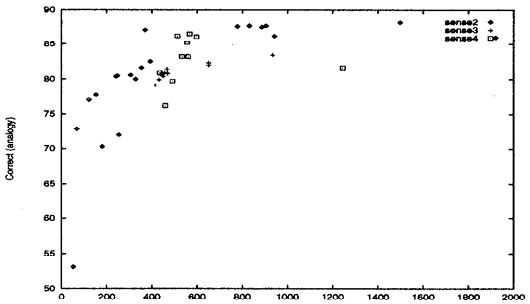


図5 $v \cap W$ と類推の正解率との関係
Fig. 5 The relationship between $v \cap W$ and correct ratio of analogy based estimation.

きたことになる．

多義動詞 v の個数と類推の正解率との関係を**図4**に示す．

図4において横軸は v の個数を示し，縦軸は類推の正解率を示す．図4によると，語義数が2の場合，v が500個以下の場合には正解率との間にばらつきがみられる．一方，500以上の場合および語義数が3および4の場合には，v の個数が多くなると類推の正解率も高くなる．類推の正解率と $v \cap W$ との関係を**図5**に示す．

図5において横軸は $v \cap W$，すなわち多義動詞 v と共起する名詞と W との共通名詞の個数を示し，縦軸は類推の正解率を示す．図5によると，語義数が2であり，かつ $v \cap W$ が300以下の場合を除くと，類推の正解率の増加にともない，$v \cap W$ の個数も増える．図4および図5より，動詞の頻度が比較的大きい場合には，それと共起す

表4 語義の解消結果（語義数4）
Table 4 The result of disambiguation experiment (four senses).

					本手法		重み付けなし
Num	$\{v, w_p, w_1, w_2, w_3\}$	Sentence	w_p w_1 w_2 w_3	v	$v \cap W$	Correct (%)	
(31)	{**develop**, create, grow, improve, expand}	187	117 (62.5) 34 (18.1) 4 (2.1) 32 (17.3)	922	597	155 (82.8)	×
(32)	{**face**, confront, cover, lie, turn}	222	54 (24.3) 103 (46.3) 12 (5.4) 53 (24.0)	859	567	184 (82.8)	×
(33)	{**get**, become, lose, understand, catch}	302	88 (29.1) 98 (32.4) 34 (11.2) 82 (27.3)	762	513	229 (75.8)	×
(34)	{**go**, come, become, run, fit}	217	101 (46.5) 66 (30.4) 36 (16.5) 14 (6.6)	732	435	145 (66.8)	×
(35)	{**make**, create, do, get, behave}	227	123 (54.1) 28 (12.3) 58 (25.5) 18 (8.1)	783	555	178 (78.4)	×
(36)	{**show**, appear, inform, prove, express}	227	121 (53.3) 16 (7.0) 40 (17.6) 50 (22.1)	996	560	181 (79.7)	×
(37)	{**take**, buy, obtain, spend, bring}	246	20 (8.1) 123 (50.0) 42 (17.0) 61 (24.9)	2,742	1,244	179 (72.7)	×
(38)	{**hold**, keep, carry, reserve, accept}	145	7 (4.8) 53 (36.5) 2 (1.5) 83 (57.2)	727	459	111 (76.5)	×
(39)	{**raise**, lift, increase, create, collect}	204	2 (1.1) 81 (39.7) 86 (42.1) 35 (17.1)	746	491	151 (74.0)	×
(40)	{**draw**, attract, pull, close, write}	162	78 (48.1) 13 (8.0) 43 (26.5) 28 (17.4)	798	533	123 (75.9)	×
Total (4 senses)		2,139	1,063 (49.6)			1,636 (76.4)	
Total		9,706	5,998 (58.2)			7,572 (78.6)	

る名詞の個数も多くなるため，類推の正解率が向上し，結果的に $v \cap W$，すなわち，多義語の特徴を示す名詞が多く抽出できることが分かる．

次に，$v \cap W$ と多義解消の正解率との関係を**図6**に示す．

図6において横軸は $v \cap W$ を示し，縦軸は正解率を示す．図6によると，語義数が2の場合には，$v \cap W$ が400以下のときは正解率との間にばらつきがみられる．一方 $v \cap W$ が750を超えると $v \cap W$ の増加にともない正解率もほぼ増加する．語義数が3の場合には $v \cap W$ の個数は400～500に集中し，正解率は70～81%であった．さらに語義数が4の場合には $v \cap W$ は450～600に集中し，正解率は70～83%であった．以上のことから，多義動詞の頻度が比較的大きい場合には語義数にかかわらず類推の正解率も高くなり，多義語が持つ各意味を特徴付ける名詞を多く抽出できる．結果的に語義解消の正解率が上がるため，その場合には本手法が有効であることが分かる．

- 手法の問題点

表4によると，全体として2,134文に含まれる語義が正しく解消できていない．表2，3および4によると，正解率が80%以上得られている動詞グループは表5において(10)，(13)，(15)を除き，類推の正解率も約80%以上に達している．一方，表2，3および4において，最も悪い結果は{**feel**, think, sense}の動詞グループであり，正解率は55.7%であった．表5においても，{**feel**, suppose,

表5 類推結果
Table 5 The result of analogy.

Num.	動詞グループ	$v \cap W$	$\mathrm{Mu}(v, W) < 3$	Correct (%)
(1)	{**cause**, effect, produce}	254	118	85 (72.0)
(2)	{**claim**, require affirm}	328	89	71 (80.0)
(3)	{**close**, open, end}	1,498	136	120 (88.2)
(4)	{**fall**, decline, win}	393	395	325 (82.5)
(5)	{**feel**, think, sense}	53	39	20 (53.1)
(6)	{**hit**, attack, strike}	122	83	63 (77.0)
(7)	{**leave**, remain, go}	942	132	113 (86.2)
(8)	{**lose**, win, get}	1,920	329	282 (86.0)
(9)	{**manage**, accomplish, operate}	306	89	71 (80.6)
(10)	{**occur**, happen, exist}	180	136	95 (70.3)
(11)	{**order**, request, arrange}	831	138	121 (87.7)
(12)	{**pass**, adopt, succeed}	247	52	41 (80.5)
(13)	{**post**, mail, inform}	68	44	32 (72.9)
(14)	{**produce**, create, grow}	370	296	258 (87.1)
(15)	{**push**, attack, pull}	153	126	97 (77.7)
(16)	{**save**, keep, rescue}	354	115	93 (81.6)
(17)	{**ship**, put, send}	241	92	74 (80.4)
(18)	{**stop**, end, move}	886	193	169 (87.5)
(19)	{**add**, append, total}	778	164	129 (78.6)
(20)	{**keep**, maintain, protect}	905	267	234 (87.7)
(21)	{**catch**, acquire, grab, watch}	432	124	99 (79.9)
(22)	{**complete**, end, develop, fill}	450	240	193 (80.4)
(23)	{**gain**, win, get, increase}	467	187	152 (81.4)
(24)	{**grow**, increase, develop become}	651	372	305 (82.0)
(25)	{**operate**, run, act, control}	651	311	255 (82.3)
(26)	{**rise**, increase, appear, grow}	414	372	294 (79.1)
(27)	{**see**, look, know, feel}	934	497	414 (83.4)
(28)	{**want**, desire, search, lack}	470	198	159 (80.8)
(29)	{**lead**, cause, guide, precede}	456	274	221 (80.9)
(30)	{**carry**, bring, capture, behave}	440	207	167 (80.7)
(31)	{**develop**, create, grow, improve, expand}	597	253	218 (86.1)
(32)	{**face**, confront, cover, lie, turn}	567	178	154 (86.5)
(33)	{**get**, become, lose, understand, catch}	513	424	365 (86.2)
(34)	{**go**, come, become, run, fit}	435	374	302 (80.9)
(35)	{**make**, create, do, get, behave}	555	435	370 (85.2)
(36)	{**show**, appear, inform, prove, express}	560	258	214 (83.2)
(37)	{**take**, buy, obtain, spend, bring}	1,244	829	677 (81.6)
(38)	{**hold**, keep, carry, reserve, accept}	459	394	300 (76.2)
(39)	{**raise**, lift, increase, create, collect}	491	341	272 (79.7)
(40)	{**draw**, attract, pull, close, write}	533	143	119 (83.2)
Total		22,148	9,444	7,672 (81.2)

sense}の動詞グループに対する類推結果が最も悪く，正解率は53.1%であった．動詞'sense'の頻度は49と少なかったことから，'sense'と共起する名詞の数が相対的に減少した．結果的に，類似度が計算できない場合が生じたことから，動詞と名詞の共起関係を用いる本手法の限界であることが分かる．さらに，類推により重要語が正しく判定でき，語義の特徴を示す名詞の集合が得られたにもかかわらず，多義語の後方5語以内に存在する名詞がこれらに含まれなかったため，正解が得られなかった場合が生じた．

本手法では，重要語を抽出するために類推が用いられている．今後この問題に対処するため，クラスタリングにより得られた(x', y')（x'は多義判定の結果得られた語義を示し，y'はそれと共起する名詞を示す）と，多義語を含む文における(x, y)（xは多義語を示し，yはそれと共起する名詞を示す）に類推を適用する必要がある．

その他，今後の課題として以下の3点があげられる．

(1) 動詞と共起する名詞の多義性を考慮した多義解消

本手法は，名詞との共起関係を用いて動詞の多義を解消している．しかし，動詞と共起する名詞が多義である場合を考慮していない．Yarowskyら

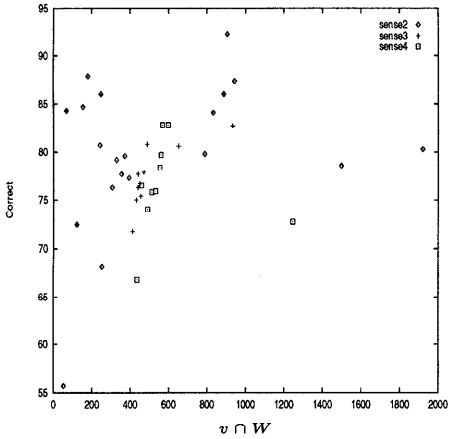


図6 $v \cap W$ と正解率との関係
Fig. 6 The relationship between $v \cap W$ and correct ratio.

は名詞が多義である場合にその出現頻度の補正を行うことで動詞の多義を解消している[28),29)]. 今後, この考え方を適用し, 多義解消の精度を向上させる必要がある.

(2) 多数の語義数を持つ動詞への適用

語義数が2, 3および4の場合におけるクラスタの重心間の距離と正解率との関係を**図7**に示す.

図7において横軸はクラスタの重心間の距離を示し, 縦軸は正解率を示す. 図7において語義数が3および4の場合には各クラスタの重心間の距離の平均を用いた. 図7によると, 語義数がいずれの場合にも重心間の距離が小さくなるにつれ, 正解率が減少している. これはクラスタの重心間の距離が小さくなるとクラスタ間の区別がつきにくくなるため, 正解率が減少していると考えられる.

語義数の違いによる重心間の距離を比較すると, 語義数が2の場合, 重心間の距離は30~70に集中した. 一方, 語義数が3および4の場合には, ともに10~40に集中し, 両者に顕著な区別がつかなかった. 今後は多数の語義数を持つ動詞に対する本手法の有効性を検証するため, 4つ以上の意味を多義動詞に割り当てることで定量的に評価を行う必要がある.

(3) クラスタリングアルゴリズムの改良

本手法は, overlappingのクラスタリングアルゴリズムである B_k 手法に基づき動詞の多義を解消している. しかし, 動詞数 n に対してその計算量は $O(n^2)$ であるため, 語義数が多くなるとその処理時間が問題となる. Needhamら[19)]は

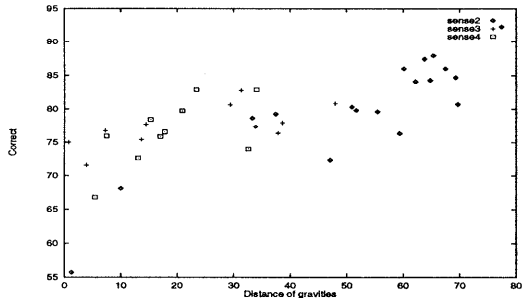


図7 クラスタの重心間の距離と正解率との相関
Fig. 7 The relationship between distance of gravities and correct ratio.

この問題に対処するため, 分割手法 (Partition method) を提案している. 今後, Needhamらの手法を適用することで語義数が多い動詞に対しても実時間で処理できるようアルゴリズムの改良を行う必要がある.

7. おわりに

本稿では, 類推により語義の特徴を抽出した結果を用いて, 文中に含まれる多義語の曖昧さを解消する手法を提案した. 本手法では, 多義を判定しながら意味的なクラスタリングを行うことで多義解消に必要な情報を抽出する. そこで, 表層上は1つの要素である多義動詞を, 多義語が持つ各意味がまとまった複数要素であるととらえ, これを1つ1つの意味に対応させた要素に分解したうえでクラスタを作成するという手法を用いた. 我々の手法における学習とは, 多義語を含む動詞グループに対し, クラスタリング過程で, 類推を用いることで重要語を正しく判定することである. その結果, 重要語には高い重み付けを行い, 重要でない語に対しては低い重み付けを行う. この処理は動詞グループのすべての動詞が正しくクラスタリングされるまで繰り返される. 本手法の精度を検証するために *Wall Street Journal* から抽出した多義語を含む文9,706文を用いて実験を行った結果, 7,572文の正解が得られ, 正解率は78.6%に達した. 今回の実験では, 多義動詞に対したかだか4つの意味を割り当て, 各々の意味で用いられている文をテスト文として用いた. 今後は, 4つ以上の意味を多義動詞に割り当て, 定量的に評価を行うと同時に, 6.3節で述べた問題と課題にも対処する予定である.

謝辞 本研究の一部は, 平成10年度文部省科学研究費補助金, および立石科学技術振興財団, 国際交流助成の援助を受けています. ここにそれらを記し謝意に代えさせていただきます.

参考文献

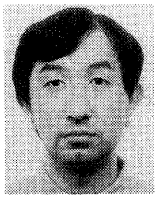
1) Ananiadou, S. and Tsujii, J.: Term Disambiguation by Adding Structural Constraints to Lexically Based Context Matching Techniques, *Proc. Natural Language Processing Pacific Rim Symposium 1997*, Phuket, Thailand, pp.195–200 (1997).
2) Brill, E.: A simple rule-based part of speech tagger, *Proc. 3rd Conference on Applied Natural Language Processing*, pp.152–155 (1992).
3) Brown, P.F., et al.: Word-Sense Disambiguation Using Statistical Methods, *Proc. 29th Annual Meeting of the ACL*, pp.264–270 (1991).
4) Calzolari, N. and Bindi, R.: Acquisition of Lexical Information from a Large Textual Italian Corpus, *Proc.13th COLING*, pp.1–6 (1990).
5) Church, K.W., et al.: Using Statistics in Lexical Analysis, *Lexical acquisition: Exploiting on-line resources to build a lexicon*, Uri, Z. (Ed.), pp.115–164, Lawrence Erlbaum Associates, London (1991).
6) Dagan, I., et al.: Contextual Word Similarity and Estimation from Sparse Data, *Proc. 31st Annual Meeting of the ACL*, pp.164–171 (1993).
7) Dagan, I., et al.: Word Sense Disambiguation Using a Second Language Monolingual Corpus, *Computational Linguistics*, Vol.20, pp.563–596 (1994).
8) Fukumoto, F. and Tsujii, J.: Automatic Recognition of Verbal Polysemy, *Proc. 15th COLING*, Kyoto, Japan, pp.762–768 (1994).
9) Gale, W.K., et al.: A Method for Disambiguating Word Senses in a Large Corpus, *Computers and the Humanities*, Vol.26, pp.415–439 (1992).
10) Guthrie, J., et al.: Subject Dependent Co-occurrence and Word Sense disambiguation, *Proc. 29th Annual Meeting of the ACL*, pp.146–152 (1991).
11) Grishman, R.: *Computational Linguistics: An Introduction*, Cambridge University Press (1986).
12) Hindle, D.: Noun Classification from Predicate-Argument Structures, *Proc. 28th Annual Meeting of the ACL*, pp.268–275 (1990).
13) Ide, N. and Véronis, J.: Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art, *Computational Linguistics*, Vol.24, No.1, pp.1–40 (1998).
14) Liberman, M. (Ed.): *CD-ROM I*, Association for Computational Linguistics Data Collection Initiative, University of Pennsylvania (1991).
15) Lin, D.: Using Syntactic Dependency as Local Context to Resolve Word Sense Ambiguity, *Proc. ACL-EACL '97*, pp.64–71 (1997).
16) Luk, A.K.: Statistical Sense Disambiguation with Relatively Small Corpora Using Dictionary Definitions, *Proc. 33rd Annual Meeting of ACL*, pp.181–188 (1995).
17) McLeod, W.T.: *The new Collins dictionary and thesaurus in one volume*, HarperCollins Publishers, London (1987).
18) Miller G.A., et.al.: WordNet: An on-line lexical database, *International Journal of Lexicography*, Vol.3, No.4, pp.235–244 (1990).
19) Needham, R.M.: Automatic classification in linguistics, *The Statistician*, No.17, pp.45–54 (1967).
20) Ng, H.T. and Lee, H.B.: Integrating Multiple Knowledge Sources to Disambiguate Word Sense: Examplar-Based Approach, *Proc. 34th Annual meeting of the Association for Computational Linguistics*, Santa Cruz, California, pp.40–47 (1996).
21) Niwa, Y. and Nitta, Y.: Co-occurrence vectors from corpora vs. distance vectors from dictionaries, *Proc. 15th COLING*, Kyoto, Japan, pp.304–309 (1994).
22) Niwa, Y. and Nitta, Y.: Statistical Word Sense Disambiguation Using Dictionary Definitions, *Proc. Natural Language Processing Pacific Rim Symposium '95*, Seoul, Korea, pp.665–670 (1995).
23) Oflazer, K.: Error-tolerant tree matching, *Proc. 16th International Conference on Computational Linguistics*, Copenhagen, Denmark, Vol.2, pp.860–864 (1996).
24) Schütze, H.: Dimensions of meaning, *Proc. Supercomputing*, pp.787–796 (1992).
25) Uramoto, N.: Disambiguation with distinctive features extracted from an example base, *Proc. 2nd Natural Language Processing Pacific Rim Symposium*, Fukuoka, Japan, pp.44–50 (1993).
26) Watanabe, H.: A method for abstracting newspaper articles by using surface clues, *Proc. 16th International Conference on Computational Linguistics*, Copenhagen, Denmark, Vol.2, pp.974–979 (1996).
27) Yarowsky, D.: Word sense disambiguation using statistical models of Roget's categories trained on large corpora, *Proc. 14th COLING*, Nantes, France, pp.454–460 (1992).
28) Yarowsky, D.: One Sense Per Collocation, *Proc. ARPA Human Language Technology Workshop*, Princeton (1993).
29) Yarowsky, D.: Unsupervised Word Sense Disambiguation Rivaling Supervised Meth-

ods, *Proc. 33rd Annual Meeting of the ACL*, pp.189–196 (1995).
30) Zernik, U.: Train1 vs. Train2: Tagging Word Senses in Corpus, *Lexical acquisition: Exploiting on-line resources to build a lexicon*, Uri, Z. (Ed.), pp.91–112, Lawrence Erlbaum Associates, London (1991).

(平成10年 1 月8日受付)
(平成10年10月2日採録)

福本 文代（正会員）

1986年学習院大学理学部数学科卒業．同年沖電気工業（株）入社．総合システム研究所勤務．1988年より1992年まで（財）新世代コンピュータ技術開発機構へ出向．1993年マンチェスター工科大学計算言語学部修士課程修了．同大学客員研究員を経て1994年より山梨大学工学部助手，現在に至る．自然言語処理の研究に従事．理学博士．ACL，言語処理学会各会員．



鈴木 良弥

1986年山梨大学工学部計算機科学科卒業．1988年同大学大学院工学研究科計算機科学専攻修了．同年木更津工業高等専門学校助手．1993年東京工業大学大学院総合理工学研究科博士後期課程修了．1994年より山梨大学工学部助手，現在に至る．音声言語処理の研究に従事．工学博士．電子情報通信学会，日本音響学会，言語処理学会各会員