

語義曖昧性解消の誤り分析

新納 浩幸¹⁾ 白井 清昭²⁾ 村田 真樹³⁾ 福本 文代⁴⁾
藤田 早苗⁵⁾ 佐々木 稔⁶⁾ 古宮 嘉那子⁷⁾ 乾 孝司⁸⁾

茨城大学 工学部 情報工学科^{1,6,7)}
北陸先端科学技術大学院大学 情報科学研究科²⁾
鳥取大学大学院 工学研究科³⁾
山梨大学大学院 総合研究部⁴⁾
NTT コミュニケーション科学基礎研究所⁵⁾
筑波大学大学院 システム情報工学研究科⁸⁾

1 はじめに

Project Next NLP は自然言語処理 (NLP) の様々なタスクの横断的な誤り分析により、今後の NLP で必要となる技術を明らかにしようとするプロジェクトである。プロジェクトでは誤り分析の対象のタスクが十数個設定され、「語義曖昧性解消」はその中の 1 つである。プロジェクトではタスク毎にチームが形成され、チーム単位でタスクの誤り分析を行っている。本論文では、我々のチーム（「語義曖昧性解消」のチーム）で行われた語義曖昧性解消の誤り分析について述べる。特に、誤り分析の初期の段階で必要となる誤り原因のタイプ分けに対して、我々がとったアプローチとその結果について述べる。

誤り分析を行う場合、(1) 分析対象のデータを定める、(2) その分析対象データを各人が分析する、(3) それら分析を統合し分析対象データの誤り原因をタイプ分けする、という手順が標準的である。我々もこの手順で誤り分析を行ったが、各自の分析結果を統合することが予想以上に困難であった。本来、誤りの原因は一意に特定できるものではなく、しかもそれを各人が独自の視点でタイプ分けしているため、分析結果の違いが大きすぎるためであった。そこで我々はこの統合のために、各自の誤り原因をクラスタリングすることを試みた。クラスタリングを行っても、自動で妥当なタイプ分けが行えるわけではないが、ある程度共通している誤り原因を特定でき、それらをベースにクラスタリング結果を調整することで誤り原因のタイプ分けが行える。

具体的には、各自の設定した誤り原因を対応する事例を用いてベクトル化し、それらのクラスタリングを行った。そのクラスタリング結果からベースとなる誤り原因を特定し、クラスタリング結果の微調整によって最終的に 9 種類の誤り原因に統合した。この 9 種類の中の主要な 3 つの誤り原因により、語義曖昧性解消の誤りの 9 割が生じていることが判明した。また考察ではタイプ分け間の類似度を定義することで、本論文で構築した 9 種類の誤り原因のタイプ分けが、各自の誤り原因のタイプ分けを代表していることを示した。これによって本論文により得られた誤り原因のタイプ分けの妥当性も確認できた。

2 分析対象データ

誤り分析用のデータは SemEval-2 の日本語 WSD タスクから作成した [6]. SemEval-2 のデータは対象単語が 50 単語あり, 各対象単語に対して 50 個の訓練用例と 50 個のテスト用例が存在する.

まず SemEval-2 のコンペの際に baseline とされたシステムを構築した. 学習アルゴリズムは SVM であり, 以下の 20 種類の素性を利用する.

e1= 二つ前の単語, e2= 二つ前の品詞, e3= その細分類,
e4= 一つ前の単語, e5= 一つ前の品詞, e6= その細分類,
e7= 問題の単語, e8= 問題の単語の品詞, e9= その細分類,
e10= ひとつ後の単語, e11= ひとつ後の品詞, e12= その細分類,
e13= 二つ後の単語, e14= 二つ後の品詞, e15= その細分類, e16= 係り受け,
e17= ふたつ前の分類語彙表の値 (5 桁), e18= ひとつ前の分類語彙表の値 (5 桁),
e19= ひとつ後の分類語彙表の値 (5 桁), e20= ふたつ後の分類語彙表の値 (5 桁)

baseline のシステムでは分類語彙表 ID の 4 桁と 5 桁を同時に使う形になっていたが, ここでは 5 桁のみとした. また一般に一つの単語に対しては複数の分類語彙表 ID が存在するので, e17,e18,e19,e20 に対する素性は複数になる場合もある. SVM の学習は libsvm の線形カーネルを用いた. 指定できるパラメータは全て default のままである.

SemEval-2 のデータはコーパスの形で配布されており, 配布された形のままでは解析できない. そのため各単語の訓練用例とテスト用例をコーパスから取り出し, それら用例を上記のフォーマットに従い, 素性リストの形式と libsvm 形式に変換した. このように作成されたデータは他の学習器や別素性を試すのが容易なデータとなっている.

SVM により識別した結果, テスト事例 2500 のうち, 誤りは 577 事例であった*¹. ここから新語義と未出現語義の事例を除くと 543 事例となった. ここからランダムに 50 個の事例を選出し, この 50 事例を誤り分析の対象事例とした. この 50 事例は付録 1 に記した.

3 各人の分析結果

前述した 50 事例の分析対象データに対して, 我々のチームのメンバー 7 名が独自に誤り分析を行った. 分析結果として, 各人は分析対象の 50 事例に対して, 各自が設定した誤り原因の記号をつけた. 表 1 がその結果である.

*¹ 平均正解率は 76.92% であり, これは SemEval2 の参加システム中, 最高値であった

表 1: 50 事例に対する各自の分析結果

事例 ID	村田	白井	福本	新納	藤田	佐々木	古宮
1	f,d,s	18,24	1-d	1-a,1-b	*	a,b	M
2	d	30	1-b	1-a,1-b,2-b	*	c	(M)
3	t	31	2-a-ii	4	*	d	?
4	n	25	2-a-ii	3-a	*	a,b,c	M
5	f,w,s	24	1-e	3-a, 2-c	*	a,b,c	M
6	p	19	1-e	1-a	Difficult	b,c	M
7	p	29	1-e	1-a	Difficult	a,b,c	M
8	d	13	2-a-ii	1-a,1-b,2-b	Difficult	c	F
9	t,d	13	1-b	2-a	*	a,d	M
10	i	26	1-d	1-d	TooShort	f	M
11	f,t	26	1-e	2-c	*	b	M
12	n	12,22	2-a-iii	2-b,2-c	Difficult	d	M
13	n	24,27	1-d	4	*	c	F,?
14	f,d	18,23	2-a-i	1-a, 1-b	Difficult	c	M
15	f,d	20	1-e	4	*	z	(M),?
16	f,d	20	1-e	4	Difficult	z	F,?
17	n	27,33	1-e	2-a	*	z	F
18	n	29	1-e	4	*	c	M
19	n	18,20	1-e	2-b	*	c	M
20	f,d	13	1-e	1-a, 1-b	*	b,c	M
21	d	13,31	1-d	1-c	*	b	F
22	u	13	1-e	1-c	*	c	M
23	d,c	14,33	1-e	2-c	*	b	F,?
24	c	15,27	1-b	2-c	Kakari,SRL	c,d	?
25	n	18	2-c-i	3-a	*	a,b	M
26	d	13	2-a-iv	2-b	*	c,d	?
27	t,c	13,31	1-e	1-a	BothAreOK	c,d	M
28	c,d,p	16,31	1-b	2-b, 2-c	Difficult	a,c,d	F
29	d,p	18,24,27	1-b	4	*	c,d	?
30	d	13	2-b-i	2-a	*	c	M,F
31	f	13,28	1-a	2-b	*	z	F
32	f	13	1-b	2-c	*	a,b,c,d,e	M,F
33	u	13,28	1-b	2-b	*	a,b,c,d	(M),?
34	d	27,31	1-b	4	*	z	F
35	w	32	1-e	1-a	GuessIsCorrect	d	F
36	d	12,25,32	1-d	2-c	*	d	M,F,?
37	f	12,19,31	1-b	2-c	*	a,d	F?
38	u	34	1-b	2-a	*	e	F
39	p	26	1-c	1-a,1-b	FeaMakingError	b,c,d	M
40	n	26	1-d	1-a, 1-b	*	b,c,f	F
41	r	17	2-c-i	2-c	FewFea	a,b,d	M
42	n	25	2-a-i	3-a	Ancient	c	M,F
43	n	14,24	1-d	1-a,1-b	Kakari,SRL	a,b	M,F
44	d	13	1-c	2-b	Difficult	a,d	M
45	d	12,14,23	1-c	2-c	SRL	a,b,d	M,F
46	d,t	25	1-b	1-a, 1-b	*	c,d	M,F
47	n	25	1-d	1-a, 1-b	Difficult	c,d	F,(M)
48	u	13,21	1-c	1-c	*	c	M
49	n	12,25	1-d	1-a,1-d	*	a,b,f	M
50	n	12,25	1-d	4	*	e	M,F,?

各自の記号の意味やどのような観点で分析したかを以下に述べる。

3.1 村田の分析：解き方に着目

普遍的な誤り分析を目指して、以下の誤り分析のフレームワークを用いる。

- 誤り事例を人手で考察し人ならそれを正しく解くにはどう解くかを考えて、その事例の解析に有効な特徴(解き方)を見つける。その特徴が学習データにあるかを確認する。
- 誤り分析の際には、正解に至るまでの誤り原因をすべて網羅して調べる。(これは、複数の誤り原因が存在する場合があります、一つの原因だけを見つけるのでは誤り分析としては不十分な場合があるためである。)

まず各事例の対象単語の品詞を調べた。品詞の出現数を表 2 に示す。表の「記号」の列はその品詞のデータに付与した記号である。

表 2: 品詞の出現数

品詞	記号	頻度
動詞	v	29
名詞	n	17
形容詞	a	4

次に各事例の解き方を調べた。解き方はある程度対象語の品詞に依存する。このため対象語の品詞を考慮しながら、解き方を考える。各事例に解き方のタグを付与する。解き方(解析に有効な特徴)の出現数を表 3 に示す。表の「記号」の列は、実際に事例に付与したタグの記号である。タグは一つの事例に複数重複してふられる場合がある。「解き方未判定」は、難しい事例で解き方が思いつかなかったものである。「文パターン」は、例えば「対象語の直前に『て』がある文パターンの場合語義 X になる」という説明が辞書にある場合があり、そのような文パターンを利用して解く方法である。「表現自身」は、例えば「対象語において漢字 X を使う場合は語義 Y になる」という説明が辞書にある場合がありそのような情報を利用して解く方法である。

表 3: 解き方の出現数

解き方	記号	頻度
格に取る名詞(対象語が用言の場合)	n	20
共起語(主に対象語が名詞の場合)	c	15
解き方未判定	タグなし	8
言い換え	p	6
文パターン	g	5
推論	r	2
表現自身	e	2
動詞(主に対象語が名詞の場合)	v	1

次に各事例の誤り原因を調べた。各事例で付与した解き方のタグを参考にして、誤り原因を調べた。誤り原因の出現数を表 4 に示す。表の「記号」の列は、実際に事例に付与したタグの記号である。タグは一つの事例に複数重複してふられる場合がある。

表 4: 誤り原因の出現数

誤り原因	記号	頻度
学習データの不足	d	19
分析が困難	n	13
素性の種類の不足	f	10
シソーラスの不備	t	5
言い換え技術が必要	p	5
素性も学習データもあるのに解けていない	u	4
格解析が必要	c	4
学習データ, テストデータの誤り	w	2
構文解析が必要	s	2
推論技術が必要	r	1
入力文の情報が少なすぎる	i	1

表の「分析が困難」は、分析が困難で分析を行っていないものを意味する。より綿密な作業により分析ができる可能性がある。「シソーラスの不備」は、シソーラスの不備の他、シソーラスでの多義性の解消が必要な場合を含む。「素性も学習データもあるのに解けていない」は、解くときに役立つ素性も存在し、その素性を持つ学習データもあるのに解けていない場合である。その素性を持つ学習データの事例数が少ないか、他の素性や学習データが悪さをした可能性がある。「格解析が必要」は、能動文、受け身文、連体などの正規化や、格の把握が必要な場合である。「入力文の情報が少なすぎる」は、入力文だけでは文が短く、その文だけでは語義解消ができない場合である。前後の文など、より広範囲の文脈の情報の入力が必要な場合である。

解き方の分類に基づき、いくつか誤り分析の事例を示す。

「格に取る名詞(対象語が用言の場合)」が解き方の場合を示す。対象語が用言の場合、格に取る名詞が語義解消に役立つ表現となりやすい[5]。格に取る名詞を中心に眺めて誤り分析を行う。

3件目の誤り事例を考察する。対象文は「…悲鳴をあげながら…」で、対象単語は「あげる」である。動詞「あげる」の格になっている「悲鳴」が語義解消に役立つ個所となる。現在のデータでは対象データの「悲鳴」が分類語彙表の情報を持たない。他のバージョンの分類語彙表には「悲鳴」の情報がある。「悲鳴」の類似事例「声」が多数学習データにある。シソーラスの情報をよりよく利用することで改善できる事例である。誤りの分類としては、「t:シソーラスの不備」を与えている。

意味ソート[4]を使うと、学習データに類似事例があるかどうかを簡明に知ることができる。「悲鳴」が存在する分類語彙表を利用して意味ソートを行った。意味ソートとは、単語群を分類語彙表の意味の順に並べる技術である。「あげる」の学習データにおいて、「あげる」のヲ格の単語を取り出し、その単語の意味ソートを行った。注目している単語「悲鳴」の近くの単語群での意味ソート結果は「13030010201:顔(545-0-1-1), 13031010101:声(545-0-1-2), 13031021203:歓声(545-0-1-2), 13031050102:叫び声(545-0-1-2), 13031050304:悲鳴*(545-0-1-2), 13041060106:顔(545-0-1-1), 13061110102:声(545-0-1-2)」である。単語の後ろの括弧にはその単語を含むデータの文の分類先を示し、単語の前の数字はその単語の分類語彙表の番号である。今解析している単語には「*」の記号を付与している。意味ソートの結果では、「叫び声」が類似事例としてあることがすぐにわかる。

「共起語(主に対象語が名詞の場合)」が解き方の場合を示す。対象語が名詞の場合、同一文の共起語が語義解消に役立つ表現となりやすい[5]。同一文には共起語が多く存在するため、この場合の誤り分析は基本的に困難である。14件目の誤り事例を考察する。対象文は、「…事件で、鶴見署は二十一日現場で…」であり、対象単語は「現場」である。対象単語は名詞であるので、共起語が役立ちやすく、この例では、「事件」「署」が

語義解消に役立つ。今の素性では対象単語の前方 2 形態素、後方 2 形態素しか素性に用いておらず、同一文の単語すべては素性に使っていない。今の素性では、「事件」「署」が使えない。共起語の素性を使えるように素性を拡張する必要がある。学習データを見たところ、共起語が重なる事例がなさそうであり、学習データ不足の問題もあるようだった。この事例には、誤りの分類としては、「f:素性の種類の不足」「d:学習データの不足」を与えている。

「言い換え」が解き方の場合を示す。7 件目の誤り事例を考察する。対象文は、「…自己防衛の意味でも…」であり、対象単語は「意味」である。正解語義は「表現や行為の意図・動機。」であり、システム出力の誤り語義は、「その言葉の表す内容。意義。」である。対象語の「意味」を「動機」に言い換えることが可能であることを認識できれば、正解語義「表現や行為の意図・動機。」と推定できるようになると思われる。この事例には、誤りの分類としては、「p:言い換え技術が必要」を与えている。^{*2}

3.2 白井の分析

誤りの要因を図 1 のように分類した。大きくは手法の問題、前処理の問題、知識の問題、データの不備、問題設定の不備に分類し、これらをさらに細かく分類した。図中の () はそれぞれの要因に該当する誤り事例の数、[] は分析対象とした 50 事例に占める割合である^{*3}。枠内の数字は付録 2 に記載されている誤り原因 ID に対応する。

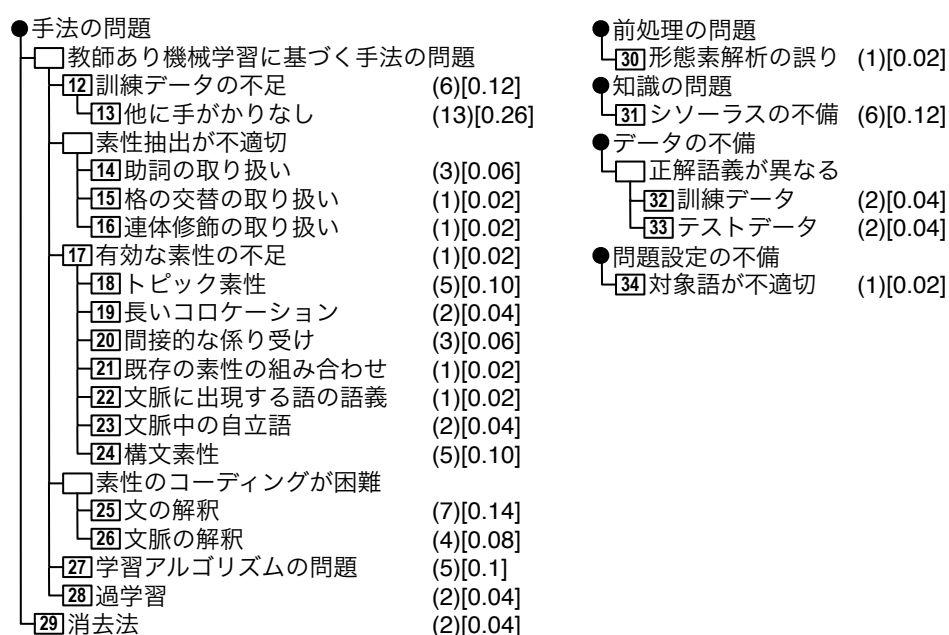


図 1: 白井による誤りの要因の分類

【手法の問題】については、教師あり機械学習手法の問題点の整理を試みた。【訓練データの不足】は、他に手がかりとなる情報がある場合^{*4}と、テスト文に類似した事例が訓練データにないと語義を判別しようがな

^{*2} 言い換え技術での処理方法として、以下が考えられる。「動機」「内容」を含む文を収集し、それを「意味」の語義「動機」の場合の学習データ、「意味」の語義「内容」の場合の学習データとして利用して解く方法である。これは文献 [3, 9] と類似した考え方になる。

^{*3} 1 つの誤り事例に対して複数の要因が割り当てられることもあるので、() 内の数字の和は 50 を越える。

^{*4} 【訓練データの不足】に分類した事例は、必ず他の要因にも分類している。

い場合 (【他に手がかりなし】) に分けた。後者の多くは定型的な言い回しで語義が決まる事例である。例えば、「指揮を*とる*」は決まり文句に近く、この文が訓練データにないと「とる」の語義を決めるのは難しい。【素性抽出が不適切】は表 5 のような文の正規化をした上で素性を抽出すべき事例である。【有効な素性の

表 5: 【素性抽出が不適切】の細分類

分類	元の文	正規化後の文
【助詞の取り扱い】	異動希望も*だし*ていました	異動希望を*だし*ていました
【格の交替の取り扱い】	退去勧告が*出さ*れている。	退去勧告を*出す*。
【連体修飾の取り扱い】	*出る*熱を	熱が*出る*

不足】は、WSD の手がかりとなる情報が素性として利用されていない場合である。今回のベースラインでは最小限の素性しか使用していないため、トピック素性 (スポーツや事件といったトピックの文内に出現することで語義が決まる事例があった)、文脈中の自立語、構文素性など、先行研究で既に使われている素性の不足も分類されている。また、【長いコロケーション】とは、ベースラインでは前後 2 単語を素性としていたが、対象語からの距離が 3 以上の単語で語義が決まる場合である。【素性のコーディングが困難】とは、語義を決める手がかりは発見できたが、高度な言語処理や推論を必要とし、機械学習の素性として表現することが難しい事例である。文の深い解釈が必要な場合 (【文の解釈】) と文章全体の解釈が必要な場合 (【文脈の解釈】) に分けた。【学習アルゴリズムの問題】とは、WSD に必要な素性は抽出できていて、類似用例も訓練データに存在するが、SVM で学習された分類器では正解を選択できなかった事例である。他の機械学習アルゴリズムなら正しく解ける可能性がある。【消去法】とは、該当しない語義を除外することで正解の語義がわかる事例を指す。例えば「かえって医師の処方を経ないで入手できる*市場*が生じている」という文での「市場」は、21128-0-0-1 の意味 (野菜などを得る市場) でもなければ 21128-0-0-3 の意味 (株式市場) でもないことから、21128-0-0-2 の意味 (売行き先) とわかる。このような事例は教師あり機械学習とは別の枠組で解く必要があるかも知れない。

【前処理の問題】は前処理の誤りに起因する事例である。【知識の問題】は外部知識の不備が誤りの原因となっているものである。【データの不備】はタグ付けされた語義の誤りである。【問題設定の不備】に分類したのは、対象語の解析対象文における品詞と辞書見出しにおける品詞が一致せず、そもそも対象語として不適切であった事例である。今回の分析では上記は少数の事例しか該当しなかったが、多くの外部知識を用いたり、文節の係り受け解析など多くの前処理を必要とするシステムでは、これらの要因ももっと細かく分類する必要があるだろう。

教師あり機械学習に基づく手法を用いるという前提で、今後 WSD の正解率を向上させるには、【訓練データの不足-他に手がかりなし】に分類した事例が多いことから、訓練データを自動的または半自動的に拡充するアプローチが有望である。また、【素性抽出が不適切】や【有効な素性の不足】に分類した要因に対応することも考えられる。ただし、表 5 に示すような正規化の処理を導入しても誤った解析結果が得られたり、単純に素性を追加しても素性数が多すぎて過学習を引き起こすなど、単純な対応だけでは WSD の正解率の向上に結びつかない可能性もあり、深い研究が必要であろう。また、【素性のコーディングが困難】に分類した事例は、現時点での言語処理技術では対応が難しい事例だが、誤り要因の 20% 程度を占めており、軽視できない。これらの事例に対応することは、チャレンジングではあるが、必要であると考えられる。

3.3 福本の分析

50 事例について誤りの原因を分析し解消に必要と考えられる知識の観点から (1) 語義解消タスク内の問題と (2) 語義解消タスク外の問題とに大別した。さらに各々を以下のように分類した。括弧は各誤り原因に相当する事例数とその割合 ((1) と (2) での割合, 及び各詳細項目での割合) を示す。また, “*” で囲まれた単語は語義解消となる対象単語を示す。

1. 語義解消タスク内の問題 (40 事例, 80%)

(a) 語義の記載がない. (1 事例, 2.5%)

「くもりを*取る*」というテスト事例において、「くもり」に関する語義情報が分類語彙表に存在しないため、「くもり」と「取る」での共起による語義判定が難しく、「取る」が訓練事例数の多い語義に判定されている。

(b) テスト事例と類似した事例が、訓練事例中に存在しない. (11 事例, 27.5%)

訓練事例不足の問題である。例えば「見せて*あげる*事ですね。」のように、動詞連用形+「て」と「あげる」のパターンが訓練事例中に存在していないために、誤って判定されている。

(c) テスト事例の語義が、訓練事例中では低頻度で出現している. (4 事例, 10%)

語義の分布に片寄りがあるものの、対象としているテスト事例中の語義の特徴と高頻度の語義を持つ特徴との区別が困難であるために、低頻度の語義であるテスト事例の語義が正しく解消できない。例えば「私の*場所*だ!」であるテスト事例が相当し、「ところ. 場」の意味の訓練事例は 49 事例、正解語義である「居るところ」は 1 事例であるために、「ところ. 場」に誤って判定されている。

(d) 解消に必要な情報が欠如している. (10 事例, 25%)

この誤り原因に相当する事例として、例えばテスト事例「*相手*をすべて倒した」において、「倒した」(行為)の対象が「*相手*」(人)であることから、共起関係を利用することにより「*相手*」が「自分と対抗して物事を争う人」に判定可能である。しかし語義の前後 2 単語というウィンドウサイズの制限により、判定に必要な「倒した」に関する情報(素性)が欠如している。

(e) 語義同士の意味が互いに類似しているために、解消が非常に難しい. (14 事例, 35%)

この誤りは、誤り原因の中で最も多くの事例が相当した誤りである。「発音を*教え*てください。」などのように、「*教え*」が「知識や技能を身につけるように導く」という語義か、正解である「自分の知っていることを告げ示す」か、両者の語義が類似しているために判定が難しい。

2. 語義解消タスク外の問題 (10 事例, 20%)

(a) 形態(語義を含む). (7 事例)

i. 形態素解析における品詞推定誤り. (2 事例, 20%)

語義解消の対象単語と共起して出現する単語の品詞が誤って判定されているために、品詞、及び共起関係の情報が利用できないという問題である。例えば、ひらがな表記の「神のみ*まへ*の」において形態素解析において「御前」と認識されていない。

ii. テスト事例の単語について、その同義語・類義語に関する情報が辞書に掲載されていない. (3 事例, 30%)

この誤りは、例えば「悲鳴を*あげ*ながら」というテスト事例において、訓練事例中に存在する「歓声」が「悲鳴」と意味的に類似していることが分類語彙表に記載されていれば、「悲鳴」と「あげる」との共起関係により判定が可能であると考えられる。

- iii. 語義解消の対象となっている単語と共起している単語に曖昧さが存在している. (1 事例, 10%)
 例えば「レベリングは*技術*がいる」というテスト事例において、「技術」と共起関係にある「いる」は「必要である」という語義と「豆などを煎る」という語義が存在する. 分類語彙表の情報として「豆などを煎る」が素性としてテスト事例に付与されているため, 共起の語彙情報を利用することができない.
- iv. 慣用句表現の認識 (1 事例, 10%)
 「めどが*立つ*。」が相当する. 語義解消の対象となっている単語を慣用句表現として認識する必要がある.
- (b) 構文 (1 事例, 10%)
 - i. 複合名詞の認識
 「国際*電話*」の事例のように, 複合名詞が正しく認識されず, 語義解消の対象単語である「*電話*」が複合名詞の一部として出現している.
- (c) 文脈 (2 事例, 20%)
 - i. 省略語の補完
 例えば「*開い*たときに請求書ご案内が上に来るように入れます。」のように, 語義解消の対象単語である「開く」の主語が省略されているため, 共起関係など, 解消に必要な情報が利用できない.

語義解消タスク内の誤り原因に相当する事例は 40 事例であり, タスク外の事例は 10 事例であったことから, 誤りの多くは語義解消の処理方法に問題があると考えられる. 語義解消内の誤り原因のうちの 6 事例は, 既存の学習手法や統計手法の工夫により語義を正しく解消できた. 一方, 例えば上述した (1) における (e) の「*教え*てください」や, 「島がびっしょり濡れているようにさえ*見え*た」における「見え」が (a) 「目にうつる」, (b) 「そう感じ取れる」において, (a) と判定するために必要となる素性が何かを明かにすることが難しい事例も存在した. 文内に限定した語彙・語義情報を用いた手法の限界であり, 今後は文外に存在する情報, 例えば分野に依存した主要語義に関する情報とも組み合わせることにより, 語義解消を行う方法なども考えられる. 今後のさらなる調査と検討が必要である.

3.4 新納の分析：手法・意味・知識・領域

語義曖昧性解消での誤りの原因を (1) 手法の問題, (2) 意味の問題, (3) 知識の問題, 及び (4) 領域の問題, の 4 タイプに大別した. タイプによっては更に詳細化した. 以下各タイプがどのような誤りかと, それをどのように判定したかを述べる.

3.4.1 手法の問題

分析対象のデータは, 学習手法として SVM を使った場合の誤りであり, 他の手法を用いた場合には誤りにならないこともある. ここでは最大エントロピー法 (ME), Naive Bayes 法 (NB), 決定リスト (DL), 及び最大頻度語義 (MFS) の 4 つを試した.

まず各手法の SemEval-2 のデータに対する正解率を表 6 に示す.

表 6: 各手法の SemEval-2 の正解率 (%)

	SVM	ME	NB	DL	MFS
正解率	76.92	76.00	75.08	74.28	68.96

SVM が最も正解率が高いが、他の手法の正解の事例を完全にカバーしているわけでない。表 7 に正解の事例の差分を示す。表 7 は行が誤りを、列が正解を表している。例えば行 (NB-×)、列 (ME-○) の要素は 98 であるが、これは NB で誤りであった事例のうち ME では正解であった事例数が 98 存在したことを意味する。表 7 から分かるように、手法 A が手法 B よりも正解率が高いからといって、必ずしも、手法 B が正解していた事例すべてを手法 A が正しく識別できる訳でない。これは手法を選択した際に生じる副作用であり、誤りの 1 つの原因であると考えられる。そして、ここでは SVM では誤りだが、他の 2 つ以上の手法で正解となっていた誤りの事例を「手法の問題」(記号 4) と判定した。

表 7: 手法間の差分

	SVM-○	ME-○	NB-○	DL-○	MFS-○
SVM-×		63	101	117	140
ME-×	86		75	87	92
NB-×	147	98		107	109
DL-×	183	130	127		115
MFS-×	339	268	262	248	

「手法の問題」の誤りは分析対象データの 50 事例中、8 事例であった。

3.4.2 意味の問題

語義曖昧性解消の問題設定自体に誤りの原因があると考えられるものを「意味の問題」と判定した。この下位分類として (a) 辞書の語義が似ていて識別困難 (記号 1-a), (b) 深い意味解析が必要 (記号 1-b), (c) 表現自体からしか識別できない (記号 1-c), 及び (d) テスト文の問題 (記号 1-d), の 4 つを設けた。

語義曖昧性解消の問題設定では、対象単語の語義が固定的に与えられる。ある対象単語が持つ複数の語義は、明確に異なる場合もあるが、非常に似ている場合もある。もしある語義 s1 と s2 が非常に似ている場合、それらを区別することは明らかに困難であり、それらを取り違えた誤りの原因は、問題自体の困難性から生じていると考えた。このようなタイプの誤りを「(a) 辞書の語義が似ていて識別困難」とした。例えば事例 27 は対象単語「強い」の語義 34522-0-0-1 「積極的に働く力にあふれている。」と語義 34522-0-0-2 「抵抗力に富み、簡単には壊れたりくずれたりしない。」を区別する問題だが、どちらの語義も互いの意味を想起させるため、意味的に非常に似ていると判断した。

上記の (a) のタイプであっても深い意味解析が可能であれば解決できるものを「(b) 深い意味解析が必要」とした。例えば事例 1 は対象単語「相手」の語義 117-0-0-2 「物事をするとき、行為の対象となる人。」と語義 117-0-0-3 「自分と対抗して物事を争う人。」を区別する問題である。「争う人」も「行為の対象となる人」であることは明かであり、意味的には非常に近く (a) である。ただしその「行為」が「争い」なのかどうかを深い意味解析で判断できれば解決できるため、「(b) 深い意味解析が必要」のタイプと判定した。(a) のタイプでかつ (b) であるかどうかは、「深い意味解析」の深さの度合いである。技術的に可能なレベルの深さと思え

れば (b) をつけた。

次に「(c) 表現自体からしか識別できない」のタイプであるが、これは語義曖昧性解消の問題として不適と思えるものである。例えば慣用表現中の単語に語義が存在していると考えるのは不自然である。また語義曖昧性解消の問題では、対象単語が自立語であることは暗黙の了解である。つまり単語の品詞自体が名詞や動詞であっても、その単語が機能語に近いものであれば、語義曖昧性解消の問題として不適と考えられる。このようなタイプの誤りを「(c) 表現自体からしか識別できない」とした。例えば事例 21, 21 の対象単語「する」、事例 48 の対象単語「やる」を、このタイプの誤りとした。

最後に「(d) テスト文の問題」のタイプであるが、これは単純にテスト文に手がかりとなる単語がほとんどないために誤るものである。これは「意味の問題」ではないが、問題設定自体に誤りの原因があると捉え、この範疇に含めた。例えば事例 10 の「*教え* て下さい。」などがこのタイプの誤りである。

3.4.3 知識の問題

語義曖昧性解消を教師付き学習により解決するアプローチをとった場合、前述した「手法の問題」「意味の問題」以外の誤りの原因は、システムに何らかの知識が不足していたためと考えられる。そこで「手法の問題」「意味の問題」以外の誤り原因を「知識の問題」と判定した。

不足している知識（解決のために必要としている知識）としては、現状のシステムの枠組みから考え、(a) その表現自体が訓練データに必要（記号 2-a）、(b) 周辺単語に同じ単語が必要（記号 2-b）、及び (c) 周辺単語に類似単語が必要（記号 2-c）、の 3 つを設定した。

例えば事例 9 の「・・・待ち伏せて詫びを *入れる* 振りをしながら・・・」の「入れる」の語義の識別には「詫びを入れる」が訓練データに必要と考え、「(a) その表現自体が訓練データに必要」と判定した。また事例 30 の「・・・朝日新聞からの国際 *電話* に対して・・・」の「電話」の語義の識別も「国際電話」が訓練データに必要だと考えた。

また事例 32 の「どうすればくもりを *取る* ことができますか？」の「取る」の語義は単語「くもり」が対象単語の周辺に存在することが必要と考え、「(b) 周辺単語に同じ単語が必要」と判定した。(a) との違いは微妙だが、(a) は慣用表現に近い表現であり、単語間に別の単語が挿入できない、態が変化できない、などの特徴があるが、(b) は「くもりをきれいに取る」や「きれいに取ったくもり」という表現が可能であり、慣用表現とは異なると考えた。

また事例 45 の「・・・患者はどここの病院でも *診* て貰えない・・・」の「診る」の語義は対象単語の周辺に「病院」と類似の単語が存在することが必要だと考え、「(c) 周辺単語に類似単語が必要」と判定した。

3.4.4 領域の問題

語義曖昧性解消の誤りは上記までの項目のいずれかに該当すると考えられるが、特殊なケースとして訓練データのコーパス内にはまれにしか出現しない表現が、テストデータとして出現したために生じる誤りが存在する。これは領域適応の問題であり、教師付き学習により問題解決を図った場合に必ず生じる問題である [8]。この原因の誤りを「領域の問題」と判定した（記号 3-a）。例えば事例 4 や事例 42 はテスト文が古文であり、学習の対象であった領域とは異なっている。このような誤りを「領域の問題」と判定した。

3.5 藤田の分析：素性に着目

まず、素性の重なりに着目した分析を 50 エラー事例に対して行った（3.5.1 節）。その中から、訓練データの拡張などを行った過去の研究でも共通してエラーとなった 16 例について詳細に調査した（3.5.2 節）。3.5.3 節で、語義曖昧性解消というタスクを考える上で、今後考えるべき問題点について考察する。

3.5.1 素性の重なるの調査

教師あり学習の場合、適切なラベルと素性を得ることができればほぼ正しく分類可能だと考えられる。適切な素性があるにも関わらず誤りになる場合、素性に付与する重みが適切ではないなど、学習器側の問題だと考えることができる。そこで、当初は、適切な素性があるかどうか、あるならば、素性に対する重みの付与などが適切かどうかを調査することを考えた。ただし、そもそも適切な素性が得られていないものが大半だったため、最終的には重みの適切さについての詳細な調査は行わなかった。

まず、システム出力語義（以下、SYS）と正解語義（以下、COR）が付与された訓練データから得られる素性と、対象のテスト文から得られる素性の重なりを調査した。

例えば、事例 ID 13(付録の表を参照)の場合、対象テスト文の 19 種類の素性のうち、10 種類は SYS と COR の両方の訓練データに出現し、8 種類は両方に出現しない。差がある素性は、1 種類（“e17=11950”，2 語前の分類語彙表の値）のみであり、これは、SYS の訓練データのみ出現している。つまり、COR にのみ出現するような特徴的な素性は得られていないことがわかる。

COR の訓練データにだけ存在する（手がかりになる可能性が高い）素性が存在するかどうかに着目すると、50 エラー例のうち、COR の訓練データにだけ出現する素性があるテスト文は 17 例（34%）^{*5}であり、そうした素性がないテスト文が 33 例（66%）を占めた。素性の不足に対応するには、学習データ自体を何らかの方法で増やすか、学習データが変わらない場合には、利用する素性を増やす必要がある。

エラー解析用システムでは、与えられた訓練データだけを用いており、利用している素性も比較的少ない。しかし、事例 ID 13 の場合でも、同一文中に「ライン」や「経験」など、他に素性として有効そうな語があることから、ウインドウ幅を広げたり、BOW を利用することでも正解となる可能性がある。また、本エラー解析用システムでは、例文を訓練データとして利用していないが、例文は重要な手がかりであり、簡便に追加できる訓練データとなり得る。

そこで、次節では、例文などを訓練データに用いた他システム（[2]、以下、システム B）の結果と比較し、両方で共通するエラーを取り出し、エラー分析を行う。

3.5.2 共通エラー

まず、システム B の概要を説明する。システム B は、2 段階に分けられる。Step-1 では、語義ラベルが付与されていない生コーパスの中から辞書の例文を含む文を抽出し、ラベルありデータとして自動獲得する。例えば、語義 15615-0-0-2 の例文「工事現場」を含む文として、例 (1)^{*6}のような文をラベルありデータとして利用できる。特に人間用の紙の辞書の場合、省スペース化のため、例文は非常に短いことが多い。Step-1 では、

^{*5} このうち、3.5.2 節でもエラーとなったのは、3 例（事例 ID 24,42,47）だった。

^{*6} 日本経済新聞 1999 年版より抜粋

例文だけをラベルありデータとして追加するより、より長くて情報量の多い文を自動獲得できることが利点である。

(1) 足場などを組み合わせて建設工事現場や各種工場のラインをつくる。

Step-2では、ラベルありデータとラベルなしデータを訓練データとして、半教師あり学習法(ハイブリッド法, *Maximum Hybrid Log-likelihood Expectation: MHLE*, [1])を適用する。MHLEでは、ラベルありデータで学習させたMEモデル(識別モデル)とラベルなしデータで学習させたNBモデル(生成モデル)を統合して分類器を得る。

素性は、エラー分析用システムで利用している素性以外に、各語の基本形、前後3語以内のbigrams, trigrams, skipbigrams, 各対象語と同一文内に出現する全内容語の基本形、トピック分類の結果*7を利用して。ただし、係り受け情報(e16)と分類語彙表の値(e17-e20)は利用していない。

もちろん、本システムを利用した場合、正解になるばかりではなく、逆にエラー分析用システムでは正解だったテスト文が不正解になる場合もあるが、本節では両者の共通エラーを取り上げる。50エラーの内、システムBでのエラーと共通したのは16例だった。分析結果を表8に示す。

表8: 共通する16エラー例の分析

	タグ	エラー番号	頻度	コメント
[A] テストデータの問題	BothAreOK?	27	1	どちらの語義も解釈可能
	Ancient	42	1	古文用の解析器や訓練データが必要
[B] そもそも判別が困難	Difficult	6,7,8,12,14,16,28,44,47	9	人間でも判別が困難
[C] 素性の問題	Kakari? SRL	24,43	2	係り受け解析誤り。項構造解析が必要
	SRL	45	1	項構造解析が必要
[D] 素性の問題2	FeaMakingError	39	1	違う文からの素性
	FewFea	41	1	対象語が文頭のため素性が半減
合計			16	

ただし、[A]-[D]は参照用の記号。「タグ」はエラー分析で付与したタグ名である。

表8から、[A][B]は両手法で解くことは困難だと考えられる。[C][D]は素性の問題だが、[C]の場合は、両手法で採用していない項構造解析(SRL)を正しく行うことができれば、正解となる可能性がある。なお、これらの対象語はすべて動詞であり、動詞の語義曖昧性解消には、SRLが特に重要であることがわかる。ただし、係り受け解析誤りも含まれるエラー例については、係り受け解析の精度向上により正解できる可能性もある。一方、[D]の場合、利用した素性が不適切だったり、少なすぎたと考えられるので、適切な素性を取り出したり、利用素性を増やすことで正解できる可能性がある。

しかしながら、[D]はシステムBでもエラーとなっている。[D]のエラー例について、3.5.1節と同様、システムBで得られた素性の重なりを調べると、訓練データの追加*8とBOW等の利用にも関わらず、少なくともラベルありデータにおいて、CORにのみ出現した素性はなく、逆にSYSにのみ出現した素性があるという結果だった。なお、両エラー例とも、CORの語義は元の訓練データにも、それぞれ1回と4回しか出現しない低頻度語義である。両対象語は、語義の頻度分布のエントロピー($E(w) = -\sum_i p(s_i|w) \log p(s_i|w)$)。ここで、 $p(s_i|w)$ は、単語 w の語義が s_i となる確率。[7])による難易度分類では、低難易度の語に分類される。つま

*7 Gibbs サンプルングを用いたトピック分類 (<http://gibbslda.sourceforge.net/>) を行い、分類されたトピック番号を利用。

*8 ただし、追加されたラベルありデータは事例ID 39の場合で3文、そのうちCORにあたるものは1文、事例ID 41の場合で57文、そのうちCORにあたるものは4文だった。

り、ある語義が圧倒的に多く出現するため低難易度の語に分類されるが、そうした語の低頻度語義の推定の難しさを示している。

3.5.3 考察

前節の分析結果をふまえ、重要だと考える点について考察する。

まず、従来の語義曖昧性解消の問題設定で今後取り組むべき課題として、以下の項目を上げる。

1. データの質の向上: 人手作成データの一貫性の担保が必要。(表 8,[A])
2. 素性の追加: 特に動詞について、係り受け精度の向上や項構造解析の組み込みが必要。(表 8,[C])
3. ラベルありデータの追加等: 特に低頻度語義に対して対処方法の考案が必要。(表 8,[D])

また、今後の方向性として、現在の語義曖昧性解消の枠組みにこだわらず、他のタスクでも利用されるには、どういった語義、どういった粒度で判別すべきか検討することが重要だと考える。特に、そもそも人間にとっても判別が困難な語義(表 8,[B])の推定が必要なのか、エンタリや語義によって全く語義の粒度が異なるにも関わらず、一律に扱ってよいのかどうか、といった点を考慮すべきだと考えている。

3.6 佐々木の分析

今回の誤り分析では、語義曖昧性解消実験で利用した訓練事例から語義を特定するパターンを考慮し、それらのパターンと誤分類したテスト事例のパターンとの比較を行った。機械学習によって構築された語義識別モデルは、共起する素性の組合せをパターンとして抽出する。語義毎のパターン集合を抽出することで、テスト事例のパターンに対して類似する訓練事例のパターンから、ふさわしい語義を識別することができる。このような機械学習に基づくアプローチで、訓練事例について以下の2つの疑問を持った。

- テスト事例に対し、訓練事例の中で何が不足しているのか
- 訓練事例からどのようなパターンが抽出されているのか

一番目の疑問はテスト事例の語義を識別できるパターンが訓練事例に存在しないからだと考えられる。そのため、訓練事例を大量に用意して、未知の事例がなくなるようにあらゆるパターンを網羅する必要がある。二番目の疑問は識別したいパターンが訓練事例に存在しているが、識別可能なパターンを抽出できていないと考えられる。この場合、語義識別のアルゴリズムを改良してより多くのテスト事例に当てはまるパターンを抽出する必要がある。従って、適切な識別を行うための必要十分な訓練事例を目指すには、訓練事例の内容をどのように改良すれば良いのかを調べることが重要である。また、人手で語義識別する際に使われるパターンの中で、機械学習を利用しても訓練事例から抽出できないパターンを見つけることも、語義識別モデル構築において重要なことだと考えられる。今回の分析では、誤り事例に対して訓練事例集合と比較することで、識別を誤る要因を調査することとした。

テスト事例に対する誤り要因を表 9 に示す。表中の記号はテスト事例に付与した誤り要因のタイプを記号で表現したものである。テスト事例の誤り要因は上記のタイプの1つを選ぶのではなく、複数の要因が重複する場合がある。各誤り要因の内容は次のようになっている。「構文情報の不足」は入力された用例文に対して文の構造から得られる特徴を捉えきれないことを表す。例えば、識別対象となる単語を含む単語間の係り受け関係を考慮した素性が少ない、格関係のような文の意味的構造を表現した素性が少ないといったことが挙

表 9: 誤り要因の分類と出現数

記号	誤り要因	頻度
a	構文情報の不足	15
b	考慮する単語の不足	18
c	パターンの一部が不足	26
d	概念情報の不足	19
e	表記のずれ	3
f	文が短く、手がかりがない	3
z	再実験では正解した例	5

げられる。「考慮する単語の不足」は語義を識別できる特徴的な共起単語が少ないことを表す。テスト事例において対象単語の前後に出現する共起単語では語義を区別することが難しい、または、訓練事例に出現する共起単語の特徴では語義を識別することが難しいという場合をこの要因として設定した。「パターンの一部が不足」は品詞情報など、単語表層以外の特徴的な情報が不足していることを表す。語義を識別できる特徴には名詞や動詞などの特徴的な単語だけではなく、接続する品詞によって語義が決定する場合もある。助詞や助動詞といった品詞を含むパターンが大きく影響して誤りとなるテスト事例はこの誤り分類とした。「概念情報の不足」は手がかりとして使う単語の上位・下位関係にある単語を利用していないことを表す。テスト事例において対象単語の前後に出現する共起単語に対し、単語を表層形で利用すると訓練事例の単語と一致しないが、外部辞書として概念体系を使うと同じ概念として一致する場合がある。同じ概念ではあるが概念体系を利用していないために誤って分類するテスト事例はこの誤り分類とした。「表記のずれ」は訓練事例に記述があるものの異なる表記で書かれ、異なる単語として扱われたことを表す。「文が短く、手がかりがない」は文が短く、特徴が捉えにくいことを表す。「再実験では正解した例」は誤り事例集作成時は異なる語義に分類されたが、再実験を行った結果正しく分類された事例を表す。この誤り要因についてはプログラムのミスなども考えられるため、今後改めて結果が異なる原因を分析する予定である。

次に、個別のテスト事例について訓練事例との比較を行い、誤りの要因を分析する。最も出現数の多い「パターンの一部が不足」が要因である事例のひとつとして、「早く元気な顔を見せて *あげる* 事です。」を紹介する。「見せてあげる」と同様の「～してあげる」というパターンが訓練事例に存在していれば適切に識別できると考えられる。しかし、訓練事例には「あげる事です」に対応する「あげる+普通名詞+助動詞」のパターンが異なる語義の事例に存在するために、この用例は誤った語義に分類されたと思われる。「考慮する単語の不足」が要因である事例として、例えば「あれで 木曜と 木曜の時に 手をぶらぶらさせてる時の 発音を *教え* てください。」を考える。この事例の正しい語義は「知識や技能を身につけるように導く」ことであるが、誤って識別した語義「知っていることを告げ示す」と共起単語を比較した結果、どちらの語義でも「～てください」が続く可能性があるため、語義を区別できる有効な特徴がそれほど存在していなかった。そのため、訓練事例数を数多く用意し、「教える」の前に接続する単語の種類を揃えることで、これらの語義を区別する知識が得られると考えられる。「概念情報の不足」が要因である事例のひとつとして、「海水は思ったより冷たくて、おとうさんも私も悲鳴を *あげ* ながらもぐんぐん進んだ。」がある。この事例の正しい語義は「勢い・資格・価値・程度を高める。」で、その中に「声を（高く）出す。」という項目がある。しかし、システムは誤った語義「取り出して言う。」と識別した。正しい語義の訓練事例には「声」、「叫び声」、「歓声」といった声に関連する単語が含まれているため、テスト事例の「悲鳴」も含めて同じ「声」の概念として捉えることができれば識別可能だと考えられる。「表記のずれ」が要因の事例には、「落札する前に聞いた方が *いい* です

か？」がある。訓練事例には正しい語義の用例として「ほうがいいです」との表記があり、テスト事例の「方」をひらがなの「ほう」に変更して識別を行うと適切に語義を識別することができた。このように、異なる表記をした単語を別の単語として扱うと正しく識別できないことがある。

3.7 古宮の分析：最大頻度語義と素性に注目

機械学習の観点から誤りの原因の分析を行った。古宮の分析の結果を表 10 に示す。なお、「MFS に誤分類」の二つの分類（表の M と (M)）には重複して分類されることはないが、これらと「テスト事例の素性が訓練事例の素性と等しい」（表の F）と「分からない、自信がない」（表の ?）については重複して分類されることがある。

表 10: 古宮の誤り原因の分類とその出現数

誤り原因	記号	頻度
MFS に誤分類（第二語義との用例の数の差が 8 以上）	M	27
MFS に誤分類（第二語義との用例の数の差が 4 以下）	(M)	5
素性が理由（テスト事例の素性が訓練事例の素性と等しい）	F	22
分からない、自信がない（原因の分類困難、または素性が原因か自信がない）	?	12

まず、訓練事例の中での最頻出語義（Most Frequent Sense, MFS）に注目した。訓練事例の MFS ではない語義をもつ事例の誤りは、機械学習によって訓練事例中の MFS に誤って分類された可能性が高いと考えたためである。そこで、「訓練事例中の MFS に誤って分類された」事例と、そうでない事例の分類を行った。すると、50 用例中の 32 用例がそのような事例であることが分かった。また、誤った回答が MFS と等しいが、それほど第二位の比率を持つ語義と差がない（4 用例以内の）ものがそのうちの 5 用例であり、残りの 27 用例は、訓練事例中の MFS が第二位の比率を持つ語義と差がある（8 用例以上の）ものであった*9。

たとえば、最も顕著な例では、「41150」の「場所」という単語は、訓練事例の全データ 50 用例中の 49 用例が「41150-0-0-1」（ところ）であり、「41150-0-0-2」（居るところ）は 1 件しかなかった。その結果、テスト事例はすべて「41150-0-0-1」と回答されており、テスト事例中に 2 例あった「41150-0-0-2」は誤りとなっていた。

このことから、エラーの原因として、機械学習の特質により、訓練事例中の MFS に誤って分類されることが大きいことが分かる。また、この例にも見られるように、今回の訓練事例の少なさから、少量の事例しか持たない語義は十分に学習ができていないことがあったと思われる。なお、MFS に誤分類されたのが誤りの原因と思われる用例には、第二語義との用例の数の差が 8 以上のときは「M」を、差が 4 以下のときは「(M)」を付与した。（差が 7 から 5 の用例は存在しなかった。）

また次に、テスト事例の素性が訓練事例の素性と等しいことで、誤分類されているものを目視で探した。たとえば、「2843」の「意味」という用例のひとつ、「エミヤのように無理して平常を装う「やせがまん」も、これらの単語で*意味*が通じるよ。」の「意味」（正解は「2843-0-0-1」（その言葉の表す内容. 意義.））は、「対象の単語のひとつ後の形態素」が「が」である、という素性が発火したためであると思われる。この素性は「2843-0-0-3」（表現や行為のもつ価値. 意義.）に頻出していたことから、「2843-0-0-3」に誤分類されている。

*9 また、残りの 18 用例のうち、二値分類ではない用例が 12 用例あったが、そのうちの 9 用例が二番目に語義頻度の高い語義に分類されていた。

この例は、「2843-0-0-3」として訓練事例にあった「意味がある」「意味がない」に「意味が通じる」という表現が少し似ていた、と見ることができる。このようなものは 22 用例あった。

このように表現の類似性は、実際に語義曖昧性解消の手掛かりともなるが、逆に誤りの原因ともなっている。なお、このような、素性が誤りの原因と思われる用例に対しては、「F」を付与した。また、訓練事例中に何度も現れる顕著な素性ではない場合には、素性が発火したかどうか分からないため、「F」とともに「？」も付与した。さらに、これらの観点から分類が難しかったものに対しては、単独で「？」を付与した。

また、他にも、クラスタリングのための分類には用いなかったが、この素性が訓練事例にあれば分類可能だと思われる素性が、訓練事例にない場合が二例、目についた。ひとつは、「早く元気な顔を見せて*あげる*事ですね。」であり、正解は「545-0-3-2」（敬語としての用法）だが、手掛かりとなりそうな「ひとつ前の形態素が『て』である」という素性が訓練事例にはなかった。

また、「村の人らは、お宮さんにおまいりして、「どうぞ、ええ水をお*あたえ*くださいませ」てお願いしてたんやと。」の正解は「755-0-0-1」（自分の物を他人に渡し、その人のものとする。）であり、この「おあたえくださる」という表現は典型的であると思われるが、訓練事例に「与えてください」のように「与える」と「くださる」の間に「て」をはさむ用法はあっても、このような用法は存在しなかった。

最後に、分類語彙表の値に曖昧性があり、本来の意味ではない値をとってきたために誤った事例がひとつあった。「コンクリートがゆっくり凝固する際に*出る*熱を冷やしているから、」という用例で、これは、「～（の）際（さい）」という表現が「きわ」として誤って取得されたために誤った例である。「出る」の訓練事例には「きわ」と同じ意味分類を持つ「外」などを二つ前の形態素にもつ用例が二例あった。

4 クラスタリングを用いた分析結果の統合

4.1 誤り原因のクラスタリング

前述した表 1 が各自の分析結果である。誤り分析の次のステップとしては、これらを統合した誤り原因のタイプ分けや、それに基づく考察などを行う必要がある。しかし各自の分析結果を統合することは、実際には、容易な作業ではない。これは誤りの原因に一意の正解がないため、統合するとしたら、結果的に誰かの分析結果をベースに修正していく形になってしまう。誰の分析結果をベースにすればよいかも正解はなく、しかもある人の分析結果をベースにした時点で、他の人の分析結果に含まれるかもしれない重要な情報を捨ててしまう危険性もある。このような観点から、我々は各自の誤り原因のセットを作り、それをクラスタリングすることで、ある程度機械的に誤り原因のタイプ分けを試みた。

各自の分析では分析対象の 50 事例に対して、各自が設定した誤り原因の記号を付与している形になっている。見方を変えて各自が設定した誤り原因の記号の 1 つ 1 つに注目すると、50 個の対象事例のどの事例がその誤り原因に対応しているかを見ることができる。対応する事例に 1 を、対応しない事例に 0 を与えれば、誤り原因は 50 次元のベクトルに変換することができる。そしてこのベクトルの距離が近いほど誤り原因の意味が近いと考えることができるため、ベクトルに変換した誤り原因のクラスタリングが可能となる。

まず各自の誤り原因を取り出すと、全部で 75 個存在した。この 75 個の誤り原因がクラスタリングの対象である。処理のために各誤り原因に ID 番号を付与した。この誤り原因と ID 番号との対応は付録 2 に記した。また付録 2 には誤り原因の意味（簡単な説明）も付与している。以後、誤り原因に対してはこの ID 番号によって参照することにする。75 個の各誤り原因を 50 次元のベクトルに変換し、そのノルムを 1 に正規化した後に Ward 法によりクラスタリングを行った [10]。このクラスタリング結果であるデンドログラムを図 2

に示す。

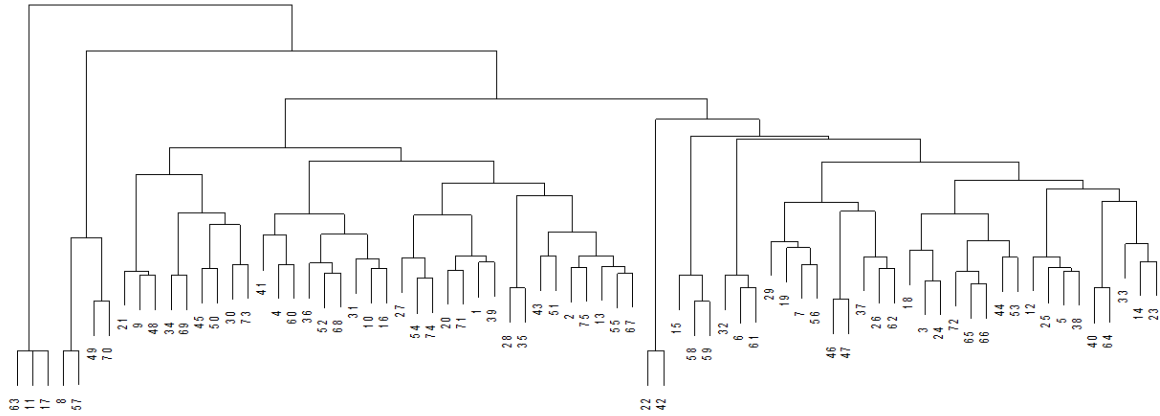


図 2: クラスタリング結果

4.2 クラスタの抽出

誤り原因の総数が 75 個，分析者が 7 人であり，その平均から考え，ベースとなる誤り原因は 10 個前後でまとめるのが適切だと考えた．そこで図 2 のデンドログラムから目視により，図 3 に示す A から M の 13 個のクラスタを取り出した．各クラスタに含まれる誤り原因の ID 番号を表 11 に示す．またクラスタ内の各誤り原因には対応する事例があるので，その総数と種類数も表 11 に示す．

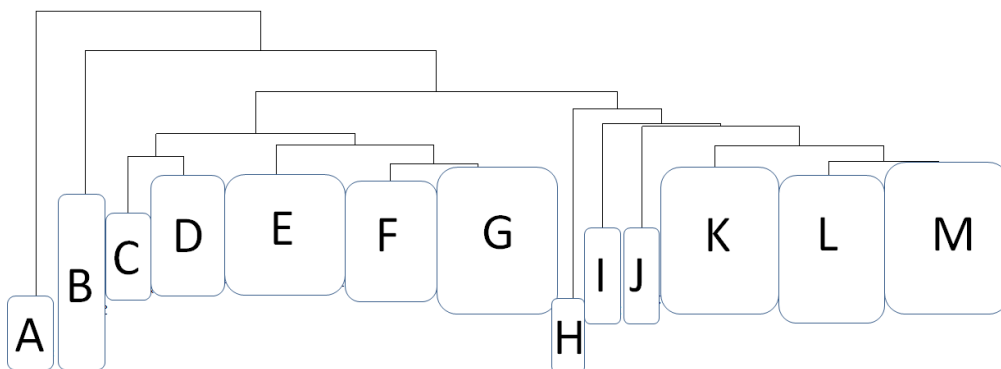


図 3: クラスタの設定

表 11: クラスタ内の誤り原因と対応する事例数

クラスタ	事例数	異なり事例数	誤り原因 ID
A	3	1	63, 11, 17
B	7	3	8, 57, 49, 70
C	8	5	21, 9, 48
D	15	9	34, 69, 45, 50, 30, 73
E	61	28	41, 4, 60, 36, 52, 68, 31, 10, 16
F	57	28	27, 54, 74, 20, 71, 1, 39
G	125	48	28, 35, 43, 51, 2, 75, 13, 55, 67
H	2	1	22, 42
I	6	3	15, 58, 59
J	5	3	32, 6, 61
K	50	25	29, 19, 7, 56, 46, 47, 37, 26, 62
L	78	36	18, 3, 24, 72, 65, 66, 44, 53
M	46	22	12, 25, 5, 38, 40, 64, 33, 14, 23

4.3 各自の誤り分析の統合

クラスタリングによって A から M の 13 個のクラスタを取り出した。次に各クラスタに意味を与える必要がある。この意味を与えることで各自の誤り分析の統合が完成する。ただし各クラスタに正確に 1 つの意味を与えることは困難である。通常、クラスタにある意味を設定した場合、クラスタ内にはその意味とは異なる要素が含まれることが多い。

ここでは各クラスタ内の要素（誤り原因）を精査し、その意味を設定する。意味を与えられたクラスタがベースとなる誤り原因となる。次にその意味から考え、不適な要素を省いたり別クラスタに移動させることで、最終的な統合を行う。

4.3.1 クラスタの意味の付与とクラスタの合併

クラスタに意味を付与するには、クラスタ内の類似している要素に注目し、それらの共通の意味を抽出することでできる。この段階で意味が同じクラスタは合併することができる。以下、各クラスタについてその内容を表にまとめる。その表の「注目」の列に“○”がついているものが意味付けを行うために注目した要素である。

クラスタ A : 【削除】

クラスタ A の内容は以下の通りである。意味付けは困難でありこのクラスタは削除する。

誤り原因 ID	事例数	注目	意味
63	1		対象語が文頭のため素性が半減
11	1		推論技術が必要
17	1		素性の不足 (その他)

クラスタ B : テスト文に問題あり

クラスタ B の内容は以下の通りであり，意味は「テスト文に問題あり」とした。

誤り原因 ID	事例数	注目	意味
8	1	○	入力文の情報が少なすぎる
57	1	○	テスト文が短かすぎる
49	2	○	テスト文自体が問題
70	3	○	文が短く，手がかりがない

クラスタ C :【削除】

クラスタ C の内容は以下の通りである．意味付けは困難でありこのクラスタは削除する。

誤り原因 ID	事例数	注目	意味
21	1		素性の不足 (素性の組み合わせ)
9	4		素性も学習データもあるのに解けていない
48	3		表現自体からしか識別できない

クラスタ D :【削除】

クラスタ D の内容は以下の通りである．意味付けは困難でありこのクラスタは削除する。

誤り原因 ID	事例数	注目	意味
34	1		対象語が不適切
69	3		表記のずれ
45	1		複合名詞の認識誤り
50	4		その表現自体が訓練データに必要
30	1		形態素解析の誤り
73	5		MFS に誤分類 (第二語義との差が小さい)

クラスタ E : シソーラスの問題

クラスタ E の内容は以下の通りであり，意味は「シソーラスの問題」とした。

誤り原因 ID	事例数	注目	意味
41	3	○	同義語の情報が分類語彙表にない
4	5	○	シソーラスの不備
60	1		正解語義も出力語義もともに正しい
36	11		訓練データに事例が存在しない
52	11	○	周辺単語に類似単語が必要
68	19		概念情報の不足
31	6	○	シソーラスの不備
10	4		格解析が必要
16	1		素性抽出が不適切 (連体修飾)

クラスタ F : 学習アルゴリズムの問題

クラスタ F の内容は以下の通りであり，意味は「学習アルゴリズムの問題」とした．

誤り原因 ID	事例数	注目	意味
27	5	○	学習アルゴリズムの問題
54	8	○	手法の問題
74	12	○	分からない，自信がない
20	3		素性の不足 (間接的な係り受け)
71	5		再実験では正解した例
1	10		素性の種類の不足
39	14		語義同士の意味が類似しているために判定が難しい

74 (分からない，自信がない) に“○”を付けている．74 は古宮が設定した分類である．古宮の分類観点を見ると，74 が付けられた事例は MFS の観点あるいは素性の様子からでは誤りの原因が特定できないものであることがわかる．これは本分析で利用した SVM による影響と見なせる．そのため 74 も「学習アルゴリズムの問題」と見なした．

クラスタ G : 訓練データの不足

クラスタ G の内容は以下の通りであり，意味は「訓練データの不足」とした．

誤り原因 ID	事例数	注目	意味
28	2		過学習
35	1		単語の語義情報が分類語彙表にない
43	1		慣用句表現の認識
51	9	○	周辺単語に同じ単語が必要
2	19	○	学習データの不足
75	22		テスト事例の素性が訓練事例の素性と等しい
13	13	○	訓練データの不足 (他に手がかりなし)
55	32	○	藤田のシステムで正解
67	26	○	学習データ中の用例パターンの不足

55 (藤田のシステムで正解) に“○”を付けている．藤田のシステムは訓練データを拡張した手法である．そのシステムで正解となったということで，その誤りの原因を「訓練データの不足」と見なした．

クラスタ H : 共起語の多義性

クラスタ H の内容は以下の通りであり，意味は「共起語の多義性」とした．

誤り原因 ID	事例数	注目	意味
22	1	○	素性の不足 (文脈に出現する語の語義)
42	1	○	対象語と共起する語の多義性

クラスタ I : 構文・格・項構造の素性不足

クラスタ I の内容は以下の通りであり、意味は「構文・格・項構造の素性不足」とした。

誤り原因 ID	事例数	注目	意味
15	1	○	素性抽出が不適切 (格の交替)
58	2	○	係り受けの間違い
59	3	○	SRL で解決可能

クラスタ J : データの誤り

クラスタ J の内容は以下の通りであり、意味は「データの誤り」とした。

誤り原因 ID	事例数	注目	意味
32	2	○	訓練データの正解語義が異なる
6	2	○	学習データ, テストデータの誤り
61	1	○	正解の誤り

クラスタ K : 深い意味解析が必要

クラスタ K の内容は以下の通りであり、意味は「深い意味解析が必要」とした。

誤り原因 ID	事例数	注目	意味
29	2		消去法
19	2		素性の不足 (長いコロケーション)
7	5	○	言い換え技術が必要
56	9	○	人間でも判別が困難
46	14	○	辞書の語義が似ていて識別困難 (正解の誤りも含む)
47	9	○	深い意味解析が必要
37	4	○	テスト事例の語義が訓練事例中では低頻度
26	4	○	素性のコーディングが困難 (文脈の解釈)
62	1		正しく素性を構築できていない

46 (辞書の語義が似ていて識別困難 (正解の誤りも含む)) に“○”を付けている。46 は新納が設定した分類である。新納は 46 と 47 (深い意味解析が必要) を区別しているが、そこでの説明にもあるように、これらの違いは微妙である。46 のタイプの誤りのほとんどは、その文脈上で人間は語義を識別できると考え、ここではまとめることにした。また 56 (人間でも判別が困難) にも“○”を付けているが、これは 46 あるいは 47 の意味と考えられるためである。

クラスタ L : 【クラスタ I と合併】

クラスタ L の内容は以下の通りであり、意味は「構文・格・項構造の素性不足」とした。これはクラスタ I の意味と同じであり、クラスタ L はクラスタ I と合併する。

誤り原因 ID	事例数	注目	意味
18	5		素性の不足 (トピック素性)
3	2	○	構文解析が必要
24	5	○	素性の不足 (構文素性)
72	27		MFS に誤分類 (第二語義との差が大きい)
65	15	○	構文情報の不足
66	18		考慮する単語の不足
44	2		省略語の補完
53	4		領域の問題

クラスター M : 素性のコーディングが困難

クラスター M の内容は以下の通りであり、意味は「素性のコーディングが困難」とした。

誤り原因 ID	事例数	注目	意味
12	6		訓練データの不足
25	7	○	素性のコーディングが困難 (文の解釈)
5	13		分析が困難
38	10	○	判定に必要な素性が欠如あるいは抽出自体が難しい
40	2		形態素解析での品詞推定誤り
64	1		テスト文が古文
33	2		テストデータの正解語義が異なる
14	3		素性抽出が不適切 (助詞の扱い)
23	2		素性の不足 (文脈中の自立語)

以上をまとめて表 12 で示される 9 個のベースとなる誤り原因を設定した。以降、上記クラスターの記号をベースとなる誤り原因の記号として用いることにする。

表 12: ベースとなる誤り原因

記号	意味	誤り原因 ID
B	テスト文に問題あり	8,57,49,70
E	シソーラスの問題	41,4,52,31
F	学習アルゴリズムの問題	27,54,74
G	訓練データの不足	51,2,13,55,67
H	共起語の多義性	22,42
I	構文・格・項構造の素性不足	15,58,59,3,24,65
J	データの誤り	32,6,61
K	深い意味解析が必要	7,56,46,47,37,26
M	素性のコーディングが困難	25,38

4.3.2 クラスタリング結果の調整

クラスタリングの対象であった 75 個の誤り原因のうち、ベースとなる誤り原因に置き換えられるものは、35 種類であった。残り 40 種類の誤り原因の中でベースとなる誤り原因に置き換えられるものを調べた。結

果、以下の表 13 に示した 11 個の置き換えができると判断した。

表 13: ベースとなる誤り原因への置き換え

誤り原因 ID	意味	置き換え先記号
11	推論技術が必要	K
9	素性も学習データもあるのに解けていない	F
36	訓練データに事例が存在しない	G
68	概念情報の不足	E
10	格解析が必要	I
16	素性抽出が不適切 (連体修飾)	I
39	語義同士の意味が類似しているために判定が難しい	K
35	単語の語義情報が分類語彙表にない	E
12	訓練データの不足	G
33	テストデータの正解語義が異なる	J
14	素性抽出が不適切 (助詞の扱い)	I

上記の調整を行った後のベースとなる誤り原因は以下の表 14 にまとめられる。本論文ではこれを「統合版誤り原因タイプ分け」と名付けることにする。

表 14: 統合版誤り原因タイプ分け

記号	意味	誤り原因 ID
B	テスト文に問題あり	8,57,49,70
E	シソーラスの問題	41,4,52,31,68,35
F	学習アルゴリズムの問題	27,54,74,9
G	訓練データの不足	51,2,13,55,67,36,12
H	共起語の多義性	22,42
I	構文・格・項構造の素性不足	15,58,59,3,24,65,10,16,14
J	データの誤り	32,6,61,33
K	深い意味解析が必要	7,56,46,47,37,26,11,39
M	素性のコーディングが困難	25,38

4.4 事例への誤り原因のタイプ付与

ここでは分析対象の 50 事例を統合版誤り原因タイプ分けに基づいてタイプ (記号) を付与する。まず対象事例に対する各自の分析結果を表 1 の各記号を統合版誤り原因タイプ分けの記号に置き換える。次に 2 名以上が同じ記号を付けていた場合に、その記号をその事例に対する誤り原因とする。結果を表 15 に示す。「統合タイプ」の列が統合版誤り原因タイプ分けによるタイプ (記号) を表す。

表 15: 50 事例に対する統合誤り原因タイプの付与

事例 ID	村田	白井	福本	新納	藤田	佐々木	古宮	統合タイプ
1	G,I	I	M	K	G	I		I,G
2	G		G	K,G	G	G		G
3	E	E	E	F	G	E	F	E,F
4		M	E		G	I,G		G
5	J,I	I	K	E	G	I,G		I,G
6	K		K	K	K	G		K
7	K		K	K	K	I,G		K
8	G	G	E	K,G	K	G		G,K
9	E,G	G	G		G	I,E		G,E
10	B	K	M	B	B	B		B
11	E	K	K	E	G			K,E
12		G,H	H	G,E	K	E		G,H,E
13		F,I	M	F	G	G	F	F,G
14	G			K	K	G		K,G
15	G		K	F	G		F	F,G
16	G		K	F	K		F	K,F
17		J,F	K		G			
18			K	F	G	G		G
19			K	G	G	G		G
20	G	G	K	K	G	G		G,K
21	G	E,G	M		G			G
22	F	G	K		G	G		G
23	G,I	J,I	K	E	G		F	G,I
24	I	F,I	G	E	I	G,E	F	I,G,E,F
25					G	I		
26	G	G		G	G	G,E	F	G
27	E,I	G,E	K	K		G,E		E,K,G
28	I,G,K	E,I	G	G,E	K	I,G,E		I,G,K,E
29	G,K	F,I	G	F	G	G,E	F	G,F
30	G	G			G	G		G
31		G	E	G	G			G
32		G	G	E	G	I,G,E		G,E
33	F	G	G	G	G	I,G,E	F	G,F
34	G	F,E	G	F	G			G,F
35	J	J	K	K	J	E		J,K
36	G	J,G,M	M	E	G	E	F	G,M,E
37		G,E	G	E	G	I,E	F	G,E
38	F		G		G			G
39	K	K	K	K		G,E		K
40		K	M	K	G	G,B		K,G
41	K			E		I,E		E
42		M				G		
43		I	M	K	I	I		I
44	G	G	K	G	K	I,E		G,K
45	G	G,I	K	E	I	I,E		I,G,E
46	G,E	M	G	K	G	G,E		G,E
47		M	M	K	K	G,E		M,K
48	F	G	K		G	G		G
49		G,M	M	K,B	G	I,B		G,M,B
50		G,M	M	F	G		F	G,M,F

対象事例の 17, 25, 42 にはタイプ（記号）が付与されなかった。これらの事例に対しては、複数の誤り原因が存在し、今回の分析作業では分析者によって見方が異なっていたと言える。

事例 ID	テスト文
17	このため、定期借地権を活用することで、初期投資や地価下落リスクなどを抑制した事業展開もみられるようになってきており、土地利用における多様な需要に応えられる環境を土地 *市場* にもたらすとともに、新たな土地需要を喚起していると考えられる。
25	ただ飲みすぎは神経が完全に麻痺して *立た* なくなったり、射精が出来なくなることがあるので、ほどほどに・・・。
42	千早ぶる神のみ *まへ* のたちばなも もろ木も共においにける哉（倭訓栞 前編十四多）

また表 15 から得られる統合誤り原因タイプの事例数を大きい順に表 16 に示す。累積カバー率はその順位までのタイプを使って分析対象の 50 事例をどの程度カバーしているかを表す。表 16 から誤りの 9 割は上位 3 つの「G: 訓練データの不足」「K: 深い意味解析が必要」「E: シソーラスの問題」のいずれか、あるいはそのいくつかの原因であることがわかる。この表 16 も我々のチームで行った「語義曖昧性解消の誤り分析」における成果の 1 つといえる。

表 16: 統合版誤り原因タイプの事例数とカバー率

記号	意味	事例数	累積カバー率
G	訓練データの不足	36	0.72
K	深い意味解析が必要	14	0.86
E	シソーラスの問題	13	0.90
F	学習アルゴリズムの問題	9	0.90
I	構文・格・項構造の素性不足	7	0.92
M	素性のコーディングが困難	4	0.92
B	テスト文に問題あり	2	0.94
H	共起語の多義性	1	0.94
J	データの誤り	1	0.94

5 考察

5.1 統合版誤り原因タイプ分けの妥当性

ここでは誤り原因のタイプ分け間の類似度を定義し、各人の誤り原因のタイプ分けおよび統合版誤り原因タイプ分け間の類似度を調べることで、統合版誤り原因タイプ分けの妥当性を考察する。統合版誤り原因タイプ分けがどの人の誤り原因のタイプ分けとも類似していれば、多くの人の考えを代表するタイプ分けであり、妥当といえる。

A と B を誤り原因のタイプ分けとし、 A と B の類似度 $Sim(A, B)$ の定義を行う。 A の要素である各誤り原因は、本論文のクラスタリングで利用したように 50 次元のベクトルで表現できる^{*10}。そして A の誤り原因のタイプが m 種類のとき、 A は以下のような集合で表現できる。

$$A = \{a_1, a_2, \dots, a_m\}$$

^{*10} 表 15 を用いれば、統合版誤り原因タイプ分けも同様に、その要素となっている 9 種類の誤り原因が 50 次元のベクトルで表現できる。

同様に、 B の誤り原因のタイプが n 種類るとき、 B は以下のような集合で表現できる。

$$B = \{b_1, b_2, \dots, b_n\}$$

ここで a_i や b_j は 50 次元のベクトルである。

本論文では $Sim(A, B)$ を以下で定義する。

$$Sim(A, B) = \max_Q \sum_Q s(a_i, b_j)$$

ここで $s(a_i, b_j)$ は a_i と b_j の類似度であり、ここでは内積を用いる。また Q は誤り原因のラベルの対応関係を表す。例えば A のラベルが $\{1, 2\}$ であり、 B のラベルが $\{1, 2, 3\}$ である場合、ラベルの対応は以下の 6 通りが存在する。 Q はこの中のいずれかになる。

$$\begin{aligned} & \{ (1,1), (2,2) \}, \quad \{ (1,2), (2,1) \}, \quad \{ (1,2), (2,3) \} \\ & \{ (1,3), (2,2) \}, \quad \{ (1,1), (2,3) \}, \quad \{ (1,3), (2,1) \} \end{aligned}$$

つまり $Sim(A, B)$ はラベル間の対応 Q に基づく誤り原因間の類似度の和を意味する。問題は Q の求め方であるが、一般にこれは組み合わせの数が膨大になるため、求めることが困難である。ここでは単純に以下の擬似コードで示される貪欲法により Q を求め、その Q を用いて $Sim(A, B)$ を算出することにした。

```
Q <- {}; K <- {1,2,...,m}; H <- {1,2,...,n}
while((K != {}) & (H != {})) {
  (i,j) = argmax s(a_i,b_j) with i ∈ K, j ∈ H
  Q <- Q + { (i,j) }
  K <- K - {i}; H <- H - {j}
}
return Q
```

またここではラベルの意味を考慮して Q を設定していないことに注意しておく。つまり分析者 A が誤り原因 i に付与した（主観的な）意味と、分析者 B が誤り原因 j に付与した（主観的な）意味が似ているか似ていないかなどは考慮せずに、 i や j のラベルが付与された事例の分布のみから i と j の類似性を測っている。これは本評価法がクラスタリングの観点からの評価法になっていることを意味する。

またここでの誤り原因のタイプ分けでは、1 つの事例に対して複数の誤り原因を与えることを許している。このため明らかに 1 つの事例に多くの誤り原因を与える方が類似度が高くなる。この問題の解消のために 1 つの事例に k 個の誤り原因を与えている場合、その部分の頻度を $1/k$ に修正した。さらに統合版誤り原因タイプ分けでは、事例 17, 25, 42 にラベルを付与していない、他の分析者は「わからない」「分析していない」などのラベルも許して全ての事例にラベルを付与している。公正な評価のため、統合版誤り原因タイプ分けによる事例 17, 25, 42 にも便宜上「その他」というラベルを付与した。

上記の処理により各誤り原因のタイプ分け間の類似度を求めた結果を表 17 に示す。表中の各人の名前はその人の誤り原因のタイプ分けを示し、【統合】は統合版誤り原因タイプ分けを示す。また類似度の横の括弧内の数値は、その行に注目して類似度の大きい順の順位を表す。

表 17: 誤り原因タイプ間の類似度

	村田	白井	福本	新納	藤田	佐々木	古宮	【統合】
村田		17.75 (2)	15.17 (5)	15.83 (4)	16.67 (3)	13.86 (7)	14.75 (6)	18.25 (1)
白井	17.75 (1)		11.83 (6)	12.08 (5)	14.33 (3)	12.33 (4)	8.94 (7)	15.44 (2)
福本	15.17 (2)	11.83 (7)		15.00 (4)	16.00 (1)	12.95 (6)	15.17 (2)	14.67 (5)
新納	15.83 (2)	12.08 (7)	15.00 (3)		14.33 (4)	14.03 (5)	13.67 (6)	20.92 (1)
藤田	16.67 (2)	14.33 (6)	16.00 (4)	14.33 (6)		14.77 (5)	16.67 (2)	28.33 (1)
佐々木	13.86 (4)	12.33 (7)	12.95 (6)	14.03 (5)	14.77 (3)		15.51 (2)	18.99 (1)
古宮	14.75 (4)	8.94 (7)	15.17 (3)	13.67 (6)	16.67 (1)	15.51 (2)		14.75 (4)

各人の縦の列の順位を足して、要素数で割った結果を表 18 に示す。この値が低いほど全体の分析結果を代表していると考えられる。本論文で導いた統合版誤り原因タイプ分け最も良い値を出しており、妥当なタイプ分けが行えたと考えられる。

表 18: 誤り原因タイプ分けの代表性を表す評価結果

村田	白井	福本	新納	藤田	佐々木	古宮	【統合】
2.50	6.00	4.50	5.00	2.50	4.83	4.16	2.14

5.2 統合版誤り原因タイプ分けと個人のタイプ分けの差

本論文ではクラスタリングを利用して各人の分析結果である誤り原因のタイプ分けを統合した。原理的には多数決と各人の設定した誤り原因の意味を勘案してタイプ分けを行ったことに相当する。しかし当然その過程で各人の分析結果の一部は切り捨てられている。具体的には統合版誤り原因タイプ分けを示す表 14 の誤り原因 ID の項目に属さなかった誤り原因がそれにあたる。

表 19: 切り捨てられた誤り原因

誤り原因 ID	分析者	事例数	意味
1	村田	10	素性の種類の不足
5		13	分析が困難
17	白井	1	素性の不足 (その他)
18		5	素性の不足 (トピック素性)
19		2	素性の不足 (長いコロケーション)
20		3	素性の不足 (間接的な係り受け)
21		1	素性の不足 (素性の組み合わせ)
23		2	素性の不足 (文脈中の自立語)
28		2	過学習
29		2	消去法
30		1	形態素解析の誤り
34		1	対象語が不適切
40		福本	2
43	1		慣用句表現の認識
44	2		省略語の補完
45	1		複合名詞の認識誤り
48	新納	3	表現自体からしか識別できない
50		4	その表現自体が訓練データに必要な
53		4	領域の問題
60	藤田	1	正解語義も出力語義もともに正しい
62		1	正しく素性を構築できていない
63		1	対象語が文頭のため素性が半減
64		1	テスト文が古文
66	佐々木	18	考慮する単語の不足
69		3	表記のずれ
71		5	再実験では正解した例
72	古宮	27	MFS に誤分類 (第二語義との差が大きい)
73		5	MFS に誤分類 (第二語義との差が小さい)
75		22	テスト事例の素性が訓練事例の素性と等しい

誤りの原因のタイプ分けに正解はないので、上記の表は各自の分析の誤りを提示しているものではない。自身が設定した誤り原因のタイプ分けが統合版誤り原因タイプ分けと合致していない点を見ることで、統合版誤り原因タイプ分けを更に改善していけると考える。この点が今後の課題である。

6 おわりに

本論文では Project Next NLP の「語義曖昧性解消」チームの活動として行われた語義曖昧性解消の誤り原因のタイプ分けについて述べた。誤り分析の対象事例を設定し、7名のメンバーが各自誤り分析を行い、誤り原因のタイプ分けを行った。各自の分析結果はかなり異なり、それらを人手で統合することは困難なことから、ここでは各自の設定した誤り原因（計 75 個）を対応する事例を用いてベクトル化し、それらのクラスタリングを行うことで統合処理を行った。

クラスタリングによってベースとなる誤り原因を特定し、クラスタリング結果の微調整によって最終的な誤り原因のタイプ分けが行えた。得られた誤り原因の主要な 3 つにより、語義曖昧性解消の誤りの 9 割が生じていることも判明した。また得られたタイプ分けの結果はタイプ分け間の類似度を定義して考察した結果、妥

当であることも示された。

各自が出したタイプ分けは、誰のものであっても得られたタイプ分けとは差異がある。各自がその差異の部分を検討することで、得られた誤り原因のタイプ分けが改善されていくと考えられる。この点が今後の課題である。

参考文献

- [1] Akinori Fujino, Naonori Ueda, and Masaaki Nagata. A Robust Semi-supervised Classification Method for Transfer Learning. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM'10)*, pp. 379–388, 2010.
- [2] Fujita, Sanae and Fujino, Akinori. Word Sense Disambiguation by Combining Labeled Data Expansion and Semi-Supervised Learning Method. *Transactions on Asian Language Inforamtion Processing, Association for Computinng Machinery (ACM)*, Vol. 12, No. 7, pp. 676–685, 2013.
- [3] Rada Mihalcea and Dan I. Moldovan. An Automatic Method for Generating Sense Tagged Corpora. In *Proceedings of the American Association for Artificial Intelligence (AAAI-1999)*, 1999.
- [4] 村田真樹, 神崎享子, 内元清貴, 馬青, 井佐原均. 意味ソート msort — 意味的並べかえ手法による辞書の構築例とタグつきコーパスの作成例と情報提示システム例 —. 言語処理学会誌, Vol. 7, No. 1, pp. 51–66, 2000.
- [5] 村田真樹, 内山将夫, 内元清貴, 馬青, 井佐原均. SENSEVAL2J 辞書タスクでの CRL の取り組み — 日本語単語の多義性解消における種々の機械学習手法と素性の比較 —. 言語処理学会誌, Vol. 10, No. 3, pp. 115–133, 2003.
- [6] Manabu Okumura, Kiyooki Shirai, Kanako Komiya, and Hikaru Yokono. On SemEval-2010 Japanese WSD Task. *自然言語処理*, Vol. 18, No. 3, pp. 293–307, 2011.
- [7] 白井清昭. SENSEVAL-2 日本語辞書タスク. *自然言語処理*, Vol. 10, No. 3, pp. 3–24, 2003.
- [8] Anders Sogaard. *Semi-Supervised Learning and Domain Adaptation in Natural Language Processing*. Morgan & Claypool, 2013.
- [9] 強田吉紀, 村田真樹, 三浦智, 徳久雅人. 機械学習を用いた同義語の使い分け. 言語処理学会第 19 回年次大会, pp. 585–587, 2013.
- [10] 新納浩幸. R で学ぶクラスタ解析. オーム社, 2007.

付録 1: 誤り分析対象の 50 用例

事例 ID	SemEval ID	テスト文
1	117-46	翌日の新聞は「体重六十六キロの日本人が七百三十二キロを破る」とか「六十六キロが五百七十五秒で *相手* をすべて倒した」と書き立てた。
2	545-11	早く元気な顔を見せて *あげる* 事ですね。
3	545-34	海水は思ったより冷たくて、おとうさんも私も悲鳴を *あげ* ながらずんずん進んだ。
4	755-30	さらにはまた、甲の女には与え得べからざるものを乙の女に、また乙の女には *与え* 得べからざるものを丙の女に、与え得るということもあろう。
5	755-48	村の人らは、お宮さんにおまいりして、「どうぞ、ええ水をお *あたえ* くださいませ」てお願いしてたんやと。
6	2843-10	脂肪を落とすという *意味* なら二の腕のみを細くするのは無理と思いますが、代謝を良くさせむくみを取るということなら何とか・・・？
7	2843-26	相手を尊重する意味でも、自己防衛の *意味* でも。
8	2843-50	エミヤのように無理して平常を装う「やせがまん」も、これらの単語で *意味* が通じるよ。
9	2998-37	十月八日の夜、清瀬の帰りを待ち伏せて詫びを *入れる* 振りをしながら、マニラのバグラスの親分から託かったことがあると持ちかけた。
10	5541-15	*教え* てください。
11	5541-35	あれで 木曜と 木曜の時に 手をぶらぶらさせてる時の 発音を *教え* てください。
12	10703-2	レベリングは結構 *技術* がいるみたいですね？
13	15615-1	入社3年目からずっと間接部門にいて *現場* (ライン) の経験も乏しいです。
14	15615-47	横浜市鶴見区内のマンションで昨年6月、男女4人の遺体が見つかった事件で、鶴見署は二十一日、 *現場* で自殺した同区潮田町、配管工上原三義容疑者(当時二十四歳)を被疑者死亡のまま殺人容疑で横浜地検に書類送検した。
15	17877-24	あとは今少子化で親が *子供* ばかりを監視し、思いどなりにしようとする事が、ある一定の年齢までは我慢できても、小学生くらいになると爆発するといわれます。
16	17877-49	《子供がおかしいと言う前に、大人は *子供* に向き合っているのか》
17	21128-3	このため、定期借地権を活用することで、初期投資や地価下落リスクなどを抑制した事業展開もみられるようになってきており、土地利用における多様な需要に応えられる環境を土地 *市場* にもたらすとともに、新たな土地需要を喚起していると考えられる。
18	21128-28	この結果、かえって医師の処方を経ないで入手できる *市場* が生じている。
19	21128-45	6社と別の1社で *市場* を占有している。
20	24646-6	何か、病院と保険会社間での *情報* の行き来があるのでしょうか？
21	27236-3	したがって、アメリカのビジネスモデルと日本の従来のビジネスモデルの両方に精通していて、アメリカのモデルのアイデアをベースに *し* ながら、日本型のビジネスモデルをつくれる経営者が、日本では最も強いビジネスモデルを創造できるということになる。
22	27236-31	たいいてい場合は、数回に分けてじょじょに色を薄く *し* ていく治療なので、段階的に治していきます。
23	31472-5	二十四歳頃は間接部門(総務部)が嫌でラインへの異動希望も *だし* ていましたが、その部署で6年働いた頃結婚もして子供もうまれました。
24	31472-50	いずれも耐震強度が0.5以下であることが判明し、4棟は退去勧告が *出さ* れている。
25	31640-13	ただ飲みすぎは神経が完全に麻痺して *立た* なくなったり、射精が出来なくなることがあるので、ほどほどに・・・。
26	31640-37	ところが、これまでの半導体生産方式では、ばらつき、雑音が多過ぎて誤動作してしまうため、四端子デバイスの実用化は夢と考えられたが、われわれのラジカル反応ベースの半導体生産技術の完成によってばらつき、雑音が完全に抑制できるようになったため、ようやく実用化のめどが *立っ* た。
27	34522-17	彼らによって今後、 *強い* ベンチャーが続々と誕生してくる可能性が出てきた。
28	35478-23	ダムの場合にはコンクリートの中に冷却水を流すチューブが縦横無尽に走っていて、コンクリートがゆっくり凝固する際に *出る* 熱を冷やしているから、収縮があるレベルに抑えられ、ひびが入らないのだという。
29	35478-43	9日のニューヨーク株式市場は、高値警戒感から利益を確定するための売りが *出* て、ダウ工業株平均は7営業日より下落した。

事例 ID	SemEval ID	テスト文
30	35881-44	二十・三十(十四・三十) I CRCアンマン事務所のマイーン・キッシースさんは、朝日新聞からの国際 *電話* に対して「今は衛星電話も含め、インターネット、無線など、バグダッドとは、すべての連絡手段が断たれている。現地からの連絡もない。医薬品を送る準備をしているが、バグダッドまでの陸路の安全が保証されれば、すぐにでも向かう予定だ」と話した。
31	37713-8	どうすればくもりを *取る* ことができますか？
32	37713-22	そこで、皆様に質問ですが、ヤフオクでは出品するだけで1品ごとに手数料を *とら* れると今日友人から聞きました。
33	37713-37	もちろん白川氏が実際に経営の指揮を *とる* わけではない。
34	40289-27	待ち時間がほとんどなく、5時間の滞在で7～8つのアトラクションに *乗れ* ました。
35	40333-17	しかも、その *場合*、講習後に大変難しい筆記試験があり、合格しなければ、免許取り消しになると交通課の方に脅かされました。
36	40699-20	側から *入っ* て、いちばん奥の、上座に当たる位置に、左から吉田松蔭、頼三樹三郎(鴨崖)…と居並び、更に西側にかけて、安政大獄で処刑された志士達、合わせて十五人、東側から南側にかけて、桜田門外で井伊直弼を襲撃した水戸藩士ら(うち一人は薩摩藩士)十八人、松蔭の墓だけ少し大きめの他、すべて同じ形、大きさの墓が整然と居流れています。
37	40699-40	、四月に *入り* 芝の根が勢いよく伸びてきたことや、二月以降、同スタジアムを管理する埼玉県が芝の養生に努めたことが、改善につながったとの見方を示した。
38	41135-31	夜は栗津歓迎の柔道大会が開かれ、ブーシュ・デュ・ローヌ県の県知事やマルセイユ市長、民間及び軍隊関係のお歴々を *初め* 柔道家、そして一般市民達が観戦し盛会だった。
39	41150-32	ここは、かけがえのない私の *場所* だ！
40	41912-26	知恵袋の中の回答を見ていると、「*早* 過ぎる！」という方がちらほら…。
41	44126-6	*開い* たときに「請求書ご案内」が上に来るように入れます…。
42	48488-8	千早ぶる神のみ *まへ* のたちばなも もろ木も共においにける哉 (倭訓栞 前編十四多)
43	49355-13	島がびっしょり濡れているようにさえ *見え* た。
44	49812-15	また、テレワークを導入した企業の二十三. 一%が、テレワークは「非常に効果があった」と答え、七十二. 七%の企業が「ある程度効果があった」と答えており、テレワークを導入した大半の企業が積極的な効果を *認め* ている(図表4)。
45	50038-16	けど難病患者や理解の少ない病気の患者はどこの病院でも *診* て貰えないのでしょうか？
46	51332-36	女は両手に皿を *持つ* てキッチンから出てきて、ひとつをぼくの前に、ひとつを自分の席に置く。
47	51409-24	この創造的な知識の活用能力としての「コンピテンス」をどう定義し、どう内容を定めていくかは、まだまだ議論と研究の最中で明確ではありませんが、二十一世紀の教育が「コンピテンス」と呼ばれる一般的な知的能力を *求め* て展開することは確実です。
48	52310-21	送られてきた封筒には出品者の住所氏名が書いてあるので根に持って意味もなく保管して *やり* ます。
49	52935-25	見なければ *よかっ* たです。
50	52935-41	落札する前に聞いた方が *いい* ですか？

付録 2: 誤り原因の ID 番号

誤り原因 ID	分析者	記号	意味
1	村田	f	素性の種類の不足
2	"	d	学習データの不足
3	"	s	構文解析が必要
4	"	t	シソーラスの不備
5	"	n	分析が困難
6	"	w	学習データ, テストデータの誤り
7	"	p	言い換え技術が必要
8	"	i	入力文の情報が少なすぎる
9	"	u	素性も学習データもあるのに解けていない
10	"	c	格解析が必要
11	"	r	推論技術が必要
12	白井	12	訓練データの不足
13	"	13	訓練データの不足 (他に手がかりなし)
14	"	14	素性抽出が不適切 (助詞の扱い)
15	"	15	素性抽出が不適切 (格の交替)
16	"	16	素性抽出が不適切 (連体修飾)
17	"	17	素性の不足 (その他)
18	"	18	素性の不足 (トピック素性)
19	"	19	素性の不足 (長いコロケーション)
20	"	20	素性の不足 (間接的な係り受け)
21	"	21	素性の不足 (素性の組み合わせ)
22	"	22	素性の不足 (文脈に出現する語の語義)
23	"	23	素性の不足 (文脈中の自立語)
24	"	24	素性の不足 (構文素性)
25	"	25	素性のコーディングが困難 (文の解釈)
26	"	26	素性のコーディングが困難 (文脈の解釈)
27	"	27	学習アルゴリズムの問題
28	"	28	過学習
29	"	29	消去法
30	"	30	形態素解析の誤り
31	"	31	シソーラスの不備
32	"	32	訓練データの正解語義が異なる
33	"	33	テストデータの正解語義が異なる
34	"	34	対象語が不適切
35	福本	1-a	単語の語義情報が分類語彙表にない
36	"	1-b	訓練データに事例が存在しない
37	"	1-c	テスト事例の語義が訓練事例中では低頻度
38	"	1-d	判定に必要な素性が欠如あるいは抽出自体が難しい
39	"	1-e	語義同士の意味が類似しているために判定が難しい
40	"	2-a-i	形態素解析での品詞推定誤り
41	"	2-a-ii	同義語の情報が分類語彙表にない
42	"	2-a-iii	対象語と共起する語の多義性
43	"	2-a-iv	慣用句表現の認識
44	"	2-c-i	省略語の補完
45	"	2-b-i	複合名詞の認識誤り
46	新納	1-a	辞書の語義が似ていて識別困難 (正解の誤りも含む)
47	"	1-b	深い意味解析が必要
48	"	1-c	表現自体からしか識別できない
49	"	1-d	テスト文自体が問題
50	"	2-a	その表現自体が訓練データに必要

51	新納	2-b	周辺単語に同じ単語が必要
52	"	2-c	周辺単語に類似単語が必要
53	"	3-a	領域の問題
54	"	4	手法の問題
55	藤田	*	藤田のシステムで正解
56	"	Difficult	人間でも判別が困難
57	"	TooShort	テスト文が短かすぎる
58	"	Kakari	係り受けの間違い
59	"	SRL	SRL で解決可能
60	"	BothAreOK	正解語義も出力語義もともに正しい
61	"	GuessIsCorrect	正解の誤り
62	"	FeaMakingError	正しく素性を構築できていない
63	"	FewFea	対象語が文頭のため素性が半減
64	"	Ancient	テスト文が古文
65	佐々木	a	構文情報の不足
66	"	b	考慮する単語の不足
67	"	c	学習データ中の用例パターンの不足
68	"	d	概念情報の不足
69	"	e	表記のずれ
70	"	f	文が短く、手がかりがない
71	"	z	再実験では正解した例
72	古宮	M	MFS に誤分類 (第二語義との差が大きい)
73	"	(M)	MFS に誤分類 (第二語義との差が小さい)
74	"	?	分からない, 自信がない
75	"	F	テスト事例の素性が訓練事例の素性と等しい