類語集合対応の推定と英語を介した辞書合成への応用

田中, 省作
九州大学情報基盤センター外国語情報メディア研究部門

丸林, 哲也
九州大学工学部電気情報工学科 | 株式会社アイ・ティ・フロンティア

冨浦, 洋一
九州大学大学院システム情報科学研究院知能システム学部門

https://doi.org/10.15017/1516050

出版情報：九州大学大学院システム情報科学紀要. 9 (2), pp.73-78, 2004-09-24. 九州大学大学院システム情報科学研究院
バージョン：
権利関係：

九州大学大学院
システム情報科学紀要
第9巻 第2号 平成16年9月

Research Reports on Information Science and
Electrical Engineering of Kyushu University
Vol.9, No.2, September 2004

類語集合対応の推定と英語を介した辞書合成への応用

田中省作* ・ 丸林哲也**・ 冨浦洋一***

Estimation of Synonym's Mappings and its Application to Compiling Translational Dictionary Using English

Shosaku TANAKA, Tetsuya MARUBAYASHI and Yoichi TOMIURA

(Received June 8, 2004)

Abstract: A translation between languages using English as an intermediary cannot be always valid because of polysemy of an English expression. If there are word's sense mappings from one language to another language, we can get valid α-β translations using them from α to English and from English to β transitively. Though this mapping is a piece of basic knowledge for bilingual information such as translation dictionaries and synonym's mappings, it has not been constructed yet. Then this paper uses synonyms in each language for subdividing a word into the word's senses, and estimates a synonym's mapping just from translation dictionaries and the synonyms in each language. Furthermore, the word's sense mappings are reversely derived from translation dictionaries and the estimated synonym's mappings. Finally, this paper shows the experimental result of the construction of German-Japanese translations by the proposed method.

Keywords: Translation dictionary, Translation via English, Synonym, Word's sense mapping, Synonym's mapping

1. はじめに

現在，英語は事実上リンガフランカ（国際語）としての地位を築いており，その他の言語との間の言語資源も充実している．そこで，英語以外の言語間の対訳辞書を編纂する際に，英語を介すことでその作業を効率化しようという試みがある．しかし，英語を介して得られる対訳には，中間表現である英訳の多義性によって，不適当な対訳が含まれる可能性がある．したがって，多義な英訳を介して得られる対訳から妥当なものに絞り込むことは，英語を介した対訳辞書の合成における最も重要な課題の一つである．

本論文では英語以外の言語 α から言語 β への $\alpha\beta$ 辞書[†1]の合成に際して，(1) 二言語間の知識はできるだけ用いない（英語を軸とした対訳辞書のみ），(2) 単言語知識は積極的に導入する，という立場をとり，α 英辞書，英 β 辞書に加え，新たに α 上の類語集合（以後，α 類語集合と記す），英語類語集合，β 類語集合を導入した合成手法を提案する．類語集合はある意味的観点で類似した語の集合である．よって，類語集合は1つの概念に相当し，語 w とそれが属す類語集合の1つ S の組 $\langle w, S\rangle$ は，w の S に対応した語義と考えることができる．ここで，各言語における語義の意味的対応（語義対応）があれば，語義対応を推移的にたどることによって，妥当な対訳のみを合成することができるが，このような語義対応は英語を軸としたとしても現時点では十分に整備されてはいない．そこで本論文では，本来，語義対応から導出される各言語における類語集合の意味的対応（類語集合対応）を，α 英/英 β 辞書と各言語の類語集合を用いて，α 類語集合から英語類語集合，英語類語集合から β 類語集合への類語集合対応と求め，それらの類語集合対応から逆に近似的な語義対応を得，対訳を合成する．

本論文はまず，2節で二言語間の類語集合対応に基づく英語を介した対訳の合成手法を示す．3節で類語集合対応を類語集合と対訳辞書から推定する手法について述べ，4節で関連研究について紹介する．5節では具体的な合成対象として，ドイツ語から日本語への独日辞書を考え，ドイツ語基本単語に対する対訳合成の比較実験について述べる．

2. 英語を介した対訳合成

2.1 英語を介した対訳合成とその問題点

例として，英語を介したドイツ語から日本語への独日辞書の合成を考える．ドイツ語はサンセリフ体（ABC）で，英語はローマン体（ABC）で記すことにする．

平成16年6月8日受付

* 情報基盤センター外国語情報メディア研究部門
** 工学部電気情報工学科（現在，株式会社アイ・ティ・フロンティア）
*** 知能システム学部門

†1 一般に「二言語辞書」には，対訳以外に語法に関する情報や例文も記載されている．本論文では，対訳の集合のみが記載された辞書を考え，「対訳辞書」または単に「辞書」とよぶ．

ドイツ語単語 Wörterbuch を独英辞書で引くと，dictionary という対訳が得られ，さらに dictionary を英日辞書で引くと，「辞書，字引」といった対訳が得られる．dictionary は一義であるので，Wörterbuch の対訳として「辞書，字引」を考えてもよい．

一方，Bank に対して独英辞書を引き，さらに英日辞書を引くと次のような対訳が得られる（括弧内は介した英訳）．

1. 銀行，貯金箱，*堤防，*土手，浅瀬，海嶺（bank）
2. ベンチ，裁判官席，*法廷（bench）

これらの対訳のうち，* を付した「堤防，土手，法廷」は，Bank の対訳としては不適当なものである．上記の独英辞書における対訳 Bank に対する bank は，“金融機関としての bank”，“海の地形としての bank”を意図したに過ぎず，英訳の bank や bench が多義であるために，* を付したような不適当な対訳が導出されてしまう．したがって，多義な英訳を介して得られる対訳の中から，妥当なもののみに絞り込むことは，英語を介した対訳合成における最も重要な課題といえる．

2.2 語義対応に基づく対訳合成

独英/英独辞書や英日/日英辞書といった対訳辞書のみでは，高精度の絞り込みは難しい．そこで，本論文では各言語で類語集合を導入することにする†2．類語集合とはある意味的観点で類似した語の集合である．例えば，さきの例で挙げた Bank には，次のように 3 つの類語集合が存在する[6] †3．

$G_1 = \{\text{Bank, Sandbank, Riff, } \cdots\}$
$G_2 = \{\text{Bank, Sitzbank, Sessl, } \cdots\}$
$G_3 = \{\text{Bank, Bankhaus, Staatssäckel, } \cdots\}$

当然ではあるが，ドイツ語類語集合において，Bank は「土手」といった不適当な対訳に関する類語集合には分類されていない†4．このように，類語集合は 1 つの概念に相当し，類語の語義を特化しているとみなすことができる．したがって本論文では，G が g の類語集合（$g \in G$）の 1 つであるとき，$\langle g, G\rangle$ を G に対応する g の語義と考えることにする．以降，$\langle g, G\rangle$ と記したときは，暗に $g \in G$ も成り立つものとする．さきの例でいえば，$\langle \text{Bank}, G_1\rangle$ は Bank の複数の語義のうち「浅瀬，海嶺」に相当する語義といえる．また，英語/日本語の類語集合についても同様の議論ができる．

つづいて，各言語間の語義の対応関係（語義対応）を考えてみる．各言語によって，意味体系には若干の相違，類語集合の構成にも相違が認められることから，各言語間の語義対応は多対多と考えるのが自然であろう．ここで，言語 α 上の語義から言語 β 上の語義への対応関係を考え，$\langle a, A\rangle$ に対する β 上の語義対応，すなわち β 上の語義の集合 $\{\langle b_1, B_1\rangle, \langle b_2, B_2\rangle, \cdots, \langle b_n, B_n\rangle\}$ を $\varphi_{\alpha\to\beta}(a, A)$ と表すことにする．対訳辞書に記載される対訳集合は語義対応 $\varphi_{\alpha\to\beta}$ から導かれるもので，α 上の単語 a に対する $\alpha\beta$ 辞書に記載されている対訳集合は，

$$t_{\alpha\to\beta}(a) = \{b \mid \exists AB\ \langle b, B\rangle \in \varphi_{\alpha\to\beta}(a, A)\}$$

となる．

α から英語，英語から β への語義対応があれば，α, β 間に直接の語義対応が無くとも，英語を介して β 上の語義（対訳）を推移的に得ることができる．α 上の語義 $\langle a, A\rangle$ に対応する β 上の語義の集合は，

$$\varphi_{\alpha\to\beta}(a, A) = \bigcup_{\langle e, E\rangle \in \varphi_{\alpha\to E}(a, A)} \varphi_{E\to\beta}(e, E) \qquad (1)$$

と求まる．たとえ英訳が多義であっても語義対応を利用する限りにおいては，不適当な対訳が混在する余地は無い．

しかし，このような二言語知識である語義対応は，英語を軸としたとしても，現時点では十分に整備されているとはいえない．そこで本論文では，次節で述べるように，本来，語義対応から導かれる類語集合間の意味的対応を対訳辞書と各言語の類語集合から推定し，対訳辞書と推定した類語集合間の意味的対応から逆に語義対応を近似し，(1) を適用する．

2.3 類語集合対応による語義対応の近似

語義対応から導かれる知識として，各言語間の類語集合間の意味的対応（類語集合対応）を考える．語義対応 $\varphi_{\alpha\to\beta}$ によって規定される言語 α 上の類語集合 A に対する β 上の類語集合の集合は $\{B \mid \exists ab\ a \in A, \langle b, B\rangle \in \varphi_{\alpha\to\beta}(a, A)\}$ と考えることができ，これを $\Phi_{\alpha\to\beta}(A)$ と表す．これらの類語集合間の意味的対応を類語集合対応とよぶ．

---

†2 対訳辞書といった多言語にまたがる知識の整備には，組み合わせ的な仕事量を要し，各言語間に十分通じた人材を確保することも大きな問題となる．それに対し，類語集合は単言語で閉じており，独立に構築しうるもので，仕事量や人的問題も多言語の場合に比べれば現実的なものとなる．一度，各言語で類語集合が整備され，英語を介した対訳合成手法が確立されれば，仕事量など総体的には大きな軽減となることが期待できる．本論文では，このような理由から，単言語知識である類語集合を積極的に導入する．

†3 Sandbank は「砂州，浅瀬」，Riff は「礁」，Sitzbank は「ベンチ」，Sessel は「アームチェアー，地位，…」，Bankhaus は「銀行」，Staatssäckel は「国庫」といった意味である．

†4 「土手」に対応する類語集合は，{Damm, Deich, Hindernis, …} とあり，Bank は含まれない．Damm は「堤防，ダム，車道」，Deich は「堤防，防波堤」，Hindernis は「障害物，バリケード」といった意味である．

この類語集合対応も語義対応同様，二言語知識ではあるが，非常に粒度が細かい語義対応に比べれば既存の各言語の類語集合と対訳辞書からある程度推定されることが期待される．さらに，次の仮定をおくことによって，対訳辞書と類語集合対応を用いて，語義対応を表現することが可能となる．

$$\begin{aligned}&\forall\langle a,A\rangle\langle b,B\rangle a'b'\\&\quad[\langle b,B\rangle\in\varphi_{\alpha\to\beta}(a,A)\wedge b'\in B\wedge a'\in A\\&\qquad\wedge b'\in t_{\alpha\to\beta}(a')\supset\langle b',B\rangle\in\varphi_{\alpha\to\beta}(a',A)]\end{aligned}\tag{2}$$

(2) は，$\langle a,A\rangle$ と $\langle b,B\rangle$ が語義対応にあるとき，a が属す A の類語 a' と b が属す B の類語 b' との間に対訳関係にあれば，$\langle a',A\rangle$ と $\langle b',B\rangle$ も語義対応があると見なす，という言明である．

これを仮定すると，$\langle a,A\rangle$ の語義対応は，対訳辞書と類語集合対応から，

$$\begin{aligned}&\varphi_{\alpha\to\beta}(a,A)\\&\quad=\{\langle b,B\rangle\mid b\in t_{\alpha\to\beta}(a),B\in\Phi_{\alpha\to\beta}(A)\}\end{aligned}\tag{3}$$

となる（証明は付録A）．よって，語義対応を求める問題は α 英/英 β 間の類語集合対応の推定問題に帰着される．

なお，以降，φ,t,Φ に下添えしている言語間の方向 $\alpha\to\beta$ は，簡単化のため省略する．

3. 類語集合対応の推定

3.1 重複度

類語集合対応において，α 上の類語集合 A と β 上の類語集合 B が完全に対応しているとは限らない．そこで，A に対して B が対応している度合いを表す指標として，次のような重複度を考える．

定義 (重複度)
α 上の類語集合 A に対する β 上の類語集合 B の重複度は，

$$\delta(A,B)=\frac{\#\{a\in A\mid\exists b\,\langle b,B\rangle\in\varphi(a,A)\}}{\#A}$$

とする． □

重複度 $\delta(A,B)$ は，B 上の類語と語義対応がある $a\in A$ の割合である．よって，重複度の最大値は 1 で，全ての $a\in A$ が B 上の類語と語義対応がある場合である．

3.2 類語集合対応の推定

本論文では，A に対する B の重複度 $\delta(A,B)$ を，

$$\widehat{\delta}(A,B)=\frac{\#\{a\in A\mid\exists b\,b\in t(a)\}}{\#A}$$

と算出し，類語集合対応を推定する．まず，この $\widehat{\delta}$ と δ の関係について議論する．

A の各類語 a は，必ず $\Phi(A)$ に含まれるいずれかの B と対訳関係にあり，$b\in B$ かつ $b\in t(a)$ となる b が存在する．また，a が多義であれば，$t(a)$ には $\langle a,A\rangle$ 以外の語義に対応する β 上の対訳も含まれるので，$B'\notin\Phi(A)$ という B' に，$b'\in B'$ かつ $b'\in t(a)$ という対訳 b' が存在する可能性がある．この場合，本来 δ では加算対象とならない b' が $\widehat{\delta}$ では加算対象となり，その分，δ に比べ大きくなる．したがって，値そのものについては $\widehat{\delta}(A,B)\geq\delta(A,B)$ である．

しかし，類語集合はある意味的観点での類似した語が集められているだけであり，同じ類語集合に同じ多義性を有した語が含まれること，つまり，A 中の類語 a_1,a_2 が，$B'\notin\Phi(A)$ の $b',b''\in B'$ において，$b'\in t(a_1)$ かつ $b''\in t(a_2)$ となることは少ないことが予想される．したがって，$\widehat{\delta}(A,B)$ が比較的大きい場合は，$\delta(A,B)$ は $\widehat{\delta}(A,B)$ で近似することができる．ただし，$\widehat{\delta}(A,B)$ が比較的小さな場合は，$\delta(A,B)=0$ つまり $B\notin\Phi(A)$ であることも考えられ，近似の信頼性は低い．

そこで，$\widehat{\delta}(A,B)$ の高い B から優先的にA の類語集合対応の要素とし，$\widehat{\delta}(A,B)$ の変化率が大きく下がる B までとした．具体的には次のようになる．

適当な閾値 $\mu(\in[0,1])$ を設定，言語 β 上の類語集合で $\widehat{\delta}(A,B)>0$ なる B を $\widehat{\delta}(A,B)$ の降順で並び替えたものを，$B_1,B_2,\cdots,B_N$ とすると，

$$\frac{\widehat{\delta}(A,B_{i+1})}{\widehat{\delta}(A,B_i)}\leq\mu$$

が成立する最小の i が K のとき，A に対する類語集合対応 $\Phi(A)$ を $\{B_1,B_2,\cdots,B_K\}$ と推定する．

4. 関 連 研 究

類似研究として Sanfilippo らの手法がある[8]．残念ながら手法の提案に留まり，実験および評価が全く行われていない．Sanfilippo らは，英単語 e と英語類語集合 $E(\ni e)$ が与えられたとき，$\langle e,E\rangle$ の対訳に優先順位を与える手法を提案し，その応用の 1 つとして英語を介した多言語辞書の合成を挙げている．基本的なアイディアは，英語類語集合を各言語における普遍的な意味ラベルとし，各言語の単語に英語類語集合を対応付け，共通の意味ラベルをたどることによって対訳を合成する，というものである．

独日対訳の合成を例に説明する．例えば，英単語 bad

は“身体的に悪い”と“倫理的に悪い”という両方の意味をもつため，badから得られるドイツ語対訳にはkrank（身体的に悪い）やböse（倫理的に悪い，状態・性質が悪い），日本語対訳には「具合がよくない」「不道徳な」といった対訳が含まれ，素直に独日対訳を合成することができない．一方，WordNet[10]におけるbadの類語集合は，

$E_1 = \{\text{bad}, \text{tough}\}$
$E_2 = \{\text{bad}, \text{immoral}\}$

などがあり，類語集合ごとに各類語のドイツ語/日本語対訳を得，2回以上重複して得られた対訳には，その英語類語集合を意味ラベルとして付与する．その結果，ドイツ語/日本語対訳には事前にkrank：E_1，böse：E_1, E_2，「具合がよくない」：E_1，「不道徳な」：E_2と意味ラベルが付与され，多義性を有するbadを介しても，

krank：「具合が良くない」
böse：「具合が良くない，不道徳な」

と正しく対訳が合成される．

5. 実　　験

対象を独日対訳とし，提案手法とSanfilippoらの手法による合成実験を行った．

5.1 データ

独英/英独辞書は，Chemnitz工科大学のFrank Richter氏が作成したGerman-English wordlistsを用いた．ドイツ語単語は単数形/複数形共々登録されており，ドイツ語側から見た場合，延べの登録表現数は95,958である．英日辞書は，Electronic Dictionary Projectの英辞郎 Ver. 3.9[2]を用いた．登録表現数は942,628である．

ドイツ語類語集合はドイツ語シソーラスDUDEN Die sinn- und sachverwandten Wörterの最小項目で構成した．約82,000語のドイツ語単語が登録されている．英語類語集合はPrinceton大学Geroge A. Miller博士らによるWordNet Ver. 1.7.1を用いた．日本語類語集合は，日本語電子化辞書の概念体系 Ver. 2.01[3]を用いて，概念体系の最下層の概念とした．

実験対象のドイツ語単語は，ドイツ語基本単語1000[5]のうち，以下の条件を充たすものとした．

1. 名詞，動詞，形容詞，副詞のいずれか．
2. 中間の英訳が多義．

この2条件を充たすドイツ語基本単語の総数は508語で，内訳は名詞233語，動詞152語，形容詞97語，副詞26語である．

これらの単語から英語を介して得られる対訳に対して，既存の独和辞典[1]を参考に，2人の人間で妥当かどうかを判定した．実験対象となったドイツ語単語から英語を介して得られる対訳の数を**Table 1**（括弧内は妥当なものの数）に示す．

Table 1 The number of translations (the number of valid translations).

	Num. of translations	
All	9,897	(3,592)
Noun	4,159	(1,278)
Verb	2,705	(999)
Adjective	2,715	(1,190)
Adverb	298	(117)

Table 2 The result of the experiment by the proposed method.

a		.1	.3	.5	.7	.9
F		.863	.690	.633	.622	.680
All	P	.391	.497	.606	.606	.735
	R	.997	.829	.664	.664	.406
Noun	P	.327	.451	.599	.599	.708
	R	1.00	.832	.628	.628	.456
Verb	P	.372	.441	.523	.523	.719
	R	.998	.915	.810	.810	.433
Adjective	P	.481	.614	.729	.729	.783
	R	.993	.769	.595	.595	.338
Adverb	P	.510	.610	.595	.595	.758
	R	1.000	.720	.500	.500	.500

5.2 結果

提案手法については，ドイツ語から英語への類語集合対応に対する閾値，英語から日本語への類語集合対応に対する閾値をそれぞれ0～1まで0.01刻みで動かし，妥当な対訳に対する精度（P）と再現率（R），精度と再現率の要約指標であるF-measureを計算した．精度は合成された対訳のうち妥当であったものの割合，再現率は実験対象の妥当な対訳のうち合成されたものの割合で，F-measureは次式で計算される．

$$F = \frac{PR}{(1-a)P + aR}$$

ただし，$a(\in [0,1])$は精度，再現率に対する重みで，0に近ければ再現率を重視し，1に近ければ精度を重視してF-measureを算出する．

提案手法については，$a \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$それぞれについて，$F$-measureが最も高かった結果を**Table 2**に，Sanfilippoらの手法による結果を**Table 3**に示す．

5.3 考察

提案手法とSanfilippoらの手法の結果を比較すると，精度はSanfilippoらの手法の方が全般的に優れているも

Table 3 The result of the experiment by the Sanfilippo's method.

a		.1	.3	.5	.7	.9
F		.159	.193	.247	.341	.552
All	P	.798				
	R	.146				
Noun	P	.752				
	R	.138				
Verb	P	.804				
	R	.144				
Adjective	P	.854				
	R	.153				
Adverb	P	.724				
	R	.179				

のの，再現率は大幅に提案手法の方が優れており，a がいずれの場合も，F-measure の観点では優位な結果が得られている．Sanfilippo らの手法では，合成元となるドイツ語単語に意味ラベルが全く付与されない場合が多く，これは非常に深刻な問題といえる．なぜなら，一般の単語を対象とした場合には，このような意味ラベルが割り当てられる割合は，今回の基本単語の場合よりも大幅に低下することが予想され，加えて日本語側でも同様の問題が生じるからである．

提案手法の精度が劣る原因としては主に次のようなものが考えられる．今回利用したドイツ語類語集合では異なる品詞の単語が混在しており，機械処理まで念頭とした WordNet などに比べると非常にアンバランスであった．また，提案手法は独英/英日と 2 回対訳を引くこととなるが，動詞や形容詞の辞書記述における表現の多様性や補助的記述によって，対訳関係がうまく得られない場合があったこと，同様の理由で日本語対訳と日本語類語集合の対応がうまく取れない場合があったことなどが挙げられる．十分な前処理を施せば，若干の精度向上が期待される．他方，Sanfilippo らの手法では英語から各言語へ 1 回対訳を引けばよく，このような問題は起こりにくい．

6. おわりに

本論文では，類語集合対応を利用した英語を介した対訳の合成手法を示した．また，英訳が多義なドイツ語基本単語に対して比較実験を行い，提案手法および提案のみに留まっていた Sanifilippo らの手法，両手法の性能と問題点を指摘した．英日間における類語集合対応推定に関する予備調査については付録に添える．なお，本手法で合成した独日辞書は，情報基盤センターで開発したドイツ語多読支援システムで利用している[9]．

類語集合対応は，辞書合成のための補助的な役割を果たすものであったが，語義的曖昧さ解消などにも利用でき，言語知識の 1 つとしても興味深い．本論文では，辞書合成を念頭としていたため，類語集合対応そのものの評価は十分とはいえないが，今後，このような点についても検討を進めていく予定である．

謝辞

本研究について示唆に富む御助言を頂きました九州大学 日高 達名誉教授，ドイツ語に関する御助言と言語資源を御提供頂きました九州大学 田畑義之教授に深く感謝致します．

参考文献

1) クラウン独和辞典 初版 CD-ROM 版, 三省堂 (1999).
2) EDP: 英辞郎, http://www.niftyserve.ne.jp/eijiro/.
3) 日本電子化辞書研究所: EDR 概念体系辞書.
4) JACET 4000 Word List, http://members.tripod.co.jp/jacetvoc-/4000/4000.htm.
5) 大岩信太郎編: ドイツ語基本単語 4000, 郁文堂 (1984).
6) Wehrle, H. and Eggers, H.: Deutscher Wortschatz, Klettbuch (1993).
7) Richter, F.: German-English wordlists, http://dict.tu-chemnitz.de/.
8) Sanfilippo, A. and Steinberger, R.: Automatic Selection and Ranking of Translation Candidates, *TMI-97*, pp. 200-207 (1997).
9) 田中省作, 田畑義之: Web を活用したドイツ語多読支援システムの構築, 情報処理学会人文科学とコンピュータシンポジウム, pp. 103-110 (2002).
10) WordNet, http://www.cogsci.princeton.edu/~wn/.

付録

A 対訳辞書と類語集合対応の語義対応の等価性の証明

言語 α から言語 β への語義対応を考える．(3) による語義対応を φ' で表し，$\varphi = \varphi'$ を証明する．

(i) $\langle b, B\rangle \in \varphi(a, A) \supset \langle b, B\rangle \in \varphi'(a, A)$

前提の語義対応より $b \in t(a), B \in \Phi(A)$ で，(3) より，$\langle b, B\rangle \in \varphi'(a, A)$ が得られる．

(ii) $\langle b, B\rangle \in \varphi'(a, A) \supset \langle b, B\rangle \in \varphi(a, A)$

前提と (3) の定義より，$B \in \Phi(A)$ かつ $b \in t(a)$ である．Φ の定義より，$\langle b', B\rangle \in \varphi(a', A)$ という $a' \in A, b' \in B$ が存在する．また，語義対応より $b' \in t(a')$ であり，仮定 (2) から，$\langle b, B\rangle \in \varphi(a, A)$ が得られる．

B 類語集合対応の予備調査

類語集合対応の推定について，予備調査を行った．対象は，英語類語集合から日本語類語集合への対応である．

B1 データと方法

英日辞書，英語類語集合，日本語類語集合は 5.1 節で示したものと同様のものを用いた．ただし，各英語類語

E_1 = {competition}
Description of E_1: a business relation in which two parties compete to gain customers; "business competition can be finished at times".

E_2 = {contest, competition}
Description of E_2: an occasion on which a winner is selected from among two or more contestants.

E_3 = {competition, contention, rivalry}
Description of E_3: the act of competing as for profit or a prize; "the teams were in fierce contention for first place".

E_4 = {rival, challenger, competitor, competition, contender}
Description of E_4: the contestant you hope to defeat; "he had respect for his rivals"; "he wanted to know what the competition was doing".

Fig. 1　Synonyms of "competition" (in WordNet).

集合は大学英語教育学会が選定した基本英単語集 JACET 4000 Word List[4] 中の単語を少なくとも 1 語以上含み，かつ英日辞書で対訳が得られる表現を 3 つ以上含むものとした．総数は 86 である．86 の類語集合のうち，34 が名詞，36 が動詞，12 が形容詞，4 が副詞に関するものであった．

WordNet では類語集合ごとに，その類語集合を規定した定義文が与えられている．対訳の妥当性は，この定義文と照らし合わせて判定すればよく，揺れなく決定できる．例えば，competition は 4 つの英語類語集合に含まれており，類語集合および定義文は **Fig. 1** のとおりである．competition の対訳は，「競争，競争者，競合，競合者，競争相手，コンペ，ライバル，争い，争奪，試合，競技会」がある．このうち $\langle$competition, $E_4\rangle$ の対訳として妥当なものは「競争者，競合者，競争相手，ライバル」となる．

また日本語対訳のうち，いずれの日本語類語集合にも含まれないものについては，その対訳唯一の日本語類語集合として扱い，$\widehat{\delta}$ を計算した．

B2　結果

71 (82.6%) の英語類語集合では，妥当と判定された対訳を含む日本語類語集合が最上位となった．ただし，$\widehat{\delta}$ の分子の最大値が 1 となったものは上記 71 には含めていない．**Fig. 1** の E_4 に対する上位の日本語類語集合を **Table 4** に示す．

うまく対応できなかった類語集合の数と原因を以下に挙げる．ただし，複数の原因を含むものについては，重複して数えている．

1. **英日辞書の対訳の記述洩れ**: 4
   類語集合 {command, bid, bidding, dictation}，定義文：an authoritative direction or instruction to do something.
   これは，「命令，指図」といった日本語類語集合に対応されるべきである．bidding や dictation に「命令」といった対訳が洩れていたため，「入札」に関する類語集合が最上位となった．
2. **概念体系の不一致**: 8
   類語集合 {subroutine, function, routine, procedure}，定義文: a set of sequence of steps, part of larger computer program.
   これは，「サブルーチン，関数，手続き」といった「プログラムの一部」に関する類語集合である．しかし，これらをちょうど包含するような日本語類語集合が存在せず，結果，$\widehat{\delta}$ の分子の最大が 1 となった．
3. **同じ多義性を含む類語が存在**: 5
   類語集合 {invest, clothe, adorn}，定義文: furnish with power or authority; of kings or emperors.
   これは，「(権力・権威などを) 与える，授ける」といった意味の類語集合である．clothe と invest は「(権力・権威を) 授ける」とは別に「(衣服を) 着せる」という同じ多義性を有している．その結果，「着せる」に相当する語義も $\widehat{\delta}$ が最大となった．類語集合で同じ多義性をもつ類語が存在した例であり，3.2 節で述べた予想に反するものである．

Table 4 The $\widehat{\delta}$ of E_4 and J.

Rank	$\widehat{\delta}$	Translations of E in J
1	0.8	ライバル，競争者
1	0.8	ライバル，好敵手，競争者，商売敵
3	0.6	競争相手
4	0.4	ライバル
4	0.4	ライバル
⋮	⋮	⋮

"ライバル", "競争者", "好敵手" and "競争相手" roughly mean "rival", "competitor" and so on. "商売敵" means "business competitor" or "bunsiness rival".