ID	KK290702
種別	調査報告
タイトル	テキストにおける多義語の語義の分布 —『現代日本語書き言葉均衡コーパス』を利用して—
Title	Distribution of Senses of a Polysemous Word in Japanese Text: Using “Balanced Corpus of Contemporary Written Japanese”
著者	山崎　誠
Author	YAMAZAKI Makoto
掲載号	29巻7号
発行日	2014年12月20日
開始ページ	251
終了ページ	262
著作権者	計量国語学会

計量国語学 29 巻 7 号（2014 年 12 月）pp.251-262.

調査報告

テキストにおける多義語の語義の分布

―『現代日本語書き言葉均衡コーパス』を利用して―

山崎　誠（国立国語研究所）

要旨

本稿は，テキストにおいて多義語の語義がどのような出現傾向を示すかを『現代日本語書き言葉均衡コーパス』をデータとして調査し，次の結果を得た．（1）多義語がテキスト中で 2 回以上使われる際，同じ語義で使われることが多い．ただし，例外もあり，必ずしも強い制約ではない．（2）同じ語義で使われる多義語の間の出現間隔は異なる語義で使われる同一の多義語の出現間隔よりも短い．（3）多義語のうちのある語義に対する類義・対義関係を作る語のうち出現間隔が短いものが認められた．これらの現象は語彙的結束性がひとまとまりのテキストに対して働いていることの現れであろうと推測される．

キーワード：多義語，出現間隔，語彙的結束性，『現代日本語書き言葉均衡コーパス』

1．はじめに

本研究は多義語の使用実態の一面をテキストにおける語彙的結束性により説明しようとするものである．具体的には，多義語がテキストの中で複数回使用される場合，特定の 1 つの語義が繰り返し使われる可能性が高いことを『現代日本語書き言葉均衡コーパス』をデータとして検証する．

多義語の研究は認知意味論の発展もあり，近年盛んになっている．日本語でも国広 (2006) により具体的かつ詳細な記述が行われているほか，柏野 (2007) のような辞書を比較した研究もある．多義性の解消が重要なテーマとなっている自然言語処理では，Gale et als.(1992) により，"one sense per discourse" という 1 つのテキストにおける多義語の語義の偏りに関する原則が示されている．ただし，これは英語における傾向であり，日本語での具体的な研究は管見の限りまだないようである．

2．方法

本研究では，テキストにおいて多義語の語義が実現する際には Halliday & Hasan(1976) の提案する語彙的結束性が働いていると仮定する．そのために，以下の 3 つの方法で多義語の語義の分布がどのようになっているか調査を行う．

（1）多義語を構成する複数の語義の種類の計量調査

例えば語Wが語義 1，語義 2，語義 3 から成る多義語の場合，調査対象であるテキストに何種類の語義が現れるかということである．語彙的結束性により，実現する語義は特定

の1つに偏ることが予想される．

（2）同一の多義語の出現間隔の計量調査

多義語が同一テキストに複数回出現する場合，「語と語の距離」と「語義の一致・不一致」との関係を調査し，距離と語彙的結束性の強さとの関係を調査する．同じ語義どうしなら出現間隔が小さく，異なる語義であれば出現間隔が大きいと予想される．

（3）類義語・対義語を含めた出現間隔の計量調査

上記（2）の調査を類義語・対義語にも拡大して調査する．例えば，「起きる（「発生する」の意味）」と「起こる」，「起きる（「目覚める」の意味）」と「寝る」のように，多義語の意味に類似した語や反対の意味を表す語は，語彙的結束性により語義の遠い語と比べて相対的に近い位置に出現していることが予想される．

3．データ

本調査で使用するデータは，国立国語研究所で開発されている『現代日本語書き言葉均衡コーパス』（以下 BCCWJ と略す）の可変長サンプルである．可変長サンプルとは，1つの章や節などのまとまりを単位とするテキストで，平均 4,000 字程度の長さを持つデータである．語彙的結束性が 1 つの話題の中で閉じていると仮定すればそれを観察するのに適したデータであると言える．使用したデータは，BCCWJ のサブコーパスである出版サブコーパスの書籍部分と図書館サブコーパス全体である．これら 2 つの対象を併せた延べ語数は約 4800 万語（長単位）であり，BCCWJ の過半を占めるデータである．

調査対象とした語は，「甘い」「起きる」「起こす」「下りる」「式」「高い」「電話」「寝る」「乗る・載る [1]」「低い」「招く」「呼ぶ」の 12 語である．また，比較のための関連語として「起こる」「眠る」「安い」も併せて調査した [2]．BCCWJ からの抽出は『中納言』を利用し，検索は長単位で行った．長単位を利用した理由は，以下のとおりである．仮に短単位を採用した場合，複合語を構成する個々の要素が 1 短単位として切り出されることが多くなる．例えば，「延喜式」の「式」が切り出されたりするような場合である．この「式」は，古代の法律の意味であるが，この意味での「式」は現代では単独では用いられないだろう．短単位を用いると，このような化石的な語義も扱うことになるため，「式」が他の語と複合せず，単独で用いられた場合を対象とすることにした．そのための条件として，長単位で「式」を検索し，複合語の中に埋め込まれていない，単独の「式」を抽出したものである．なお，長単位を採用することにより，「甘い」の派生形である「甘さ」「甘すぎる」などが調査対象に含まれなくなる問題もあるが，今回の調査は全ての品詞を通じて単独での使用に限るという条件で統一した．

4．多義の認め方

調査に当たって各語について多義をどのように認めるかが問題となる．どの程度語義が異なれば異なる語義と認めるべきか，客観的に示すのは難しい．国語辞書の語義の利用は

1　BCCWJ では「乗る」「載る」を 1 語として扱っているため，これらを同じ語の範囲内とした．ちなみに『明鏡国語辞典第 2 版』でも 1 語扱いである．以降，見出し語形は「乗る」で統一する．

2　「起こる」は本調査で典拠とした『明鏡国語辞典第 2 版』では多義語ではなかった．

3　用例は省略した．

1つの解決法であるが，多義を構成する語義の数が非常に多い場合，調査する上で現実的でないと考えられる．例えば，『明鏡国語辞典第2版』（以下，『明鏡』と略す）では「甘い」の語義として次の11個[3]を立てている．

①食物に砂糖や蜜のような味を感じる．
②食物に（砂糖っけが多く）塩けや辛みの刺激が少ないと感じる．からくない．
③においが糖分を思わせるようで快い．甘美だ．
④〔多く連体形で〕心がとろけるように快い．また，愛情こまやかにうちとけている．甘美だ．
⑤採点や規制がきびしさに欠けるさま．
⑥物事に対する判断や見通しが安易で，厳しさに欠けるさま．
⑦《「－・くない」の形で》あなどりがたく厳しい．
⑧しっかりとひきしまらず，十分に機能しない．
⑨刃物の切れ味が鈍い．
⑩守りがゆるやかですきがある．また，攻撃が厳しさを欠いて手ぬるい．
⑪株価の動きが鈍く低落ぎみだ．

実際の用例に当たってみると，①②③の区別，⑤⑥⑦の区別は往々にして付けがたい．そのような例が多くなると結果の精度に影響が出かねないため，今回は，語義の認定でなるべくゆれが少なくなるよう，適宜語義をまとめる方法を採ることにした．「甘い」の例で言えば，①②③を1グループ（味・香りの類）[4]，④を1グループ（甘美の類），⑤〜⑪を1グループ（厳しくないの類）とした．

今回調査したのは表1に掲げた12語である[5]．これらの語は，多義であり，一定以上の頻度があり，かつ，語義の区別が明瞭に付けられるものから適宜選んだ．品詞は名詞，動詞，形容詞である．

4　語義のまとめ方は筆者の恣意性に基づく部分がある．例えば，『岩波国語辞典 第七版新版』では，「味」と「香り・甘美」はそれぞれ別ブランチであるが，本稿では上記のように分けた．これは，「味」と「香り」は味覚と嗅覚という，近い感覚を表すものであること，「甘美」のグループは視覚に基づいた感覚（「甘いマスク」）や聴覚に基づいた感覚（「甘い言葉」）であり，存在しない対象について比喩的に用いられていることから，「味」「香り」のグループとは距離があると判断した．

5　語義分類は，『明鏡国語辞典第2版』を参考にそれらを適宜まとめたものである．同辞典を採用した理由は，比較的語義が細かく分類されており，個々の語義に用例が付いていることから，それらを適宜比較して筆者自身が語義をまとめていく際の情報量が多く，作業がしやすかったためである．また，最終的には筆者の判断で語義をまとめているので，使用した辞書にはそれほど大きく依存していないものと思われる．なお，分類に当てはまらないものは「その他」とし，分析から外した．「その他」は，ほとんどの調査語において1%未満である．

表1　各調査語の多義の認定

調査語	語義番号	語義の概要	対応する『明鏡』の語義
式	1	方法・方式	接尾
	2	行事・儀式	①
	3	数式・化学式等	②
電話	1	通話	
	2	電話機	②
起きる	1	立った状態になる	①
	2	目を覚ます	②③
	3	出来事が発生する	④
起こす	1	立った状態にする	①②④⑤⑥⑦⑧
	2	目を覚まさせる	②
	3	出来事を発生させる	⑨⑩⑪⑫⑬
下りる	1	低い所へ移動する	①③⑦⑧
	2	乗り物から出る	①
	3	退く	②
	4	許可が出る	⑥
寝る	1	横になる・倒れる	①③④
	2	睡眠する	②
	3	共寝する	⑤
	4	不活用	⑥
乗る	1	乗り物の中に移動する	②③⑪⑫
	2	高い物の上に移動する	①
	3	合致する	④⑤⑥⑦⑧
	4	応じる	⑨⑩
	5	印刷物等に出る	⑬
招く	1	呼び寄せる	①
	2	招待する	②③
	3	引き起こす	④
呼ぶ	1	声を掛ける	①②⑤
	2	招待する	③
	3	称する	④
	4	引き起こす	⑥
甘い	1	味・香りがよい	①②③
	2	甘美な	④
	3	厳しくない	⑤⑥⑦⑧⑨⑩⑪
高い	1	物理的な距離が大きい	①
	2	程度が大きい	②③⑤⑥⑦
	3	金額が大きい	④
低い	1	物理的な距離が小さい	①
	2	程度が劣っている	②③
	3	金額が小さい	③の一部[6]

6　金額に関するものをここに分類した.

5．調査結果

多義語が同じ語義で使用されているか異なる語義で使用されているかを調べるためには，同一サンプル中に複数使用されていなければならない．そこで調査対象とした 14 語が同一サンプルに 2 回以上出現するものを選び出した．その結果，対象となったサンプルの数を表 2 に示した．これらのサンプルにおいて出現した調査語が，表 1 で設定した語義のどの語義で使われているか，その情報を人手で付与した．表 3 は，表 2 で示した複数回出現したサンプルにおける語義ごとの出現頻度である．

表 2　調査語が複数回出現したサンプルの数

調査語	サンプル数
式	194
電話	1,377
起きる	1,049
起こす	1,129
下りる	977
寝る	789

調査語	サンプル数
乗る	1,847
招く	293
呼ぶ	4,525
甘い	320
高い	3,740
低い	1,228

表 3　語義区分ごとの出頻頻度

調査語	語義	頻度
式	1	43
	2	124
	3	1,183
電話	1	4,793
	2	1,058
起きる	1	84
	2	739
	3	2,353
起こす	1	444
	2	221
	3	2,509
下りる	1	1,871
	2	943
	3	21
	4	45

調査語	語義	頻度
寝る	1	411
	2	1,966
	3	114
	4	5
乗る	1	3,721
	2	672
	3	406
	4	244
	5	6,655
招く	1	14
	2	414
	3	315
呼ぶ	1	2,839
	2	157
	3	11,281

調査語	語義	頻度
	4	361
甘い	1	423
	2	296
	3	172
高い	1	2,491
	2	9,143
	3	1,545
低い	1	844
	2	2,769
	3	74

5.1　出現した語義の種類

図 1 は個々のサンプルについて，1 つの語義で使われたサンプルと複数の語義で使われたサンプルの割合を示したものである．図 1 からは，1 つの語義で使用されるサンプルの割合が「式」では 90 % 以上と高くなっているが，それは表 3 の数値からわかるように，もともと特定の語義での使用が多かったためと考えられる．それ以外の語は，およそ 1 つ

の語義で使用されるサンプルの割合は約 60% ～80% の間に分布している．Gale ら (1992) の調査では Brown Corpus において campaign, deposit, interior, landscape, marine の 5 語で同一サンプルから抜き出した 106 組の多義語のペアのうち 102 組が同じ語義で使われていたと報告している．一方，Krovetz(1996) では，語義タグのついたコーパスを調査した結果，1 テキストあたり複数の語義が認められたものが 33% あったと報告している（すなわち，同一の語義での使用は 67%）．本調査での結果は Krovetz の報告に近いものとなっており，1 つのテキストに 1 つの語義という傾向がゆるやかに認められるということが言える．

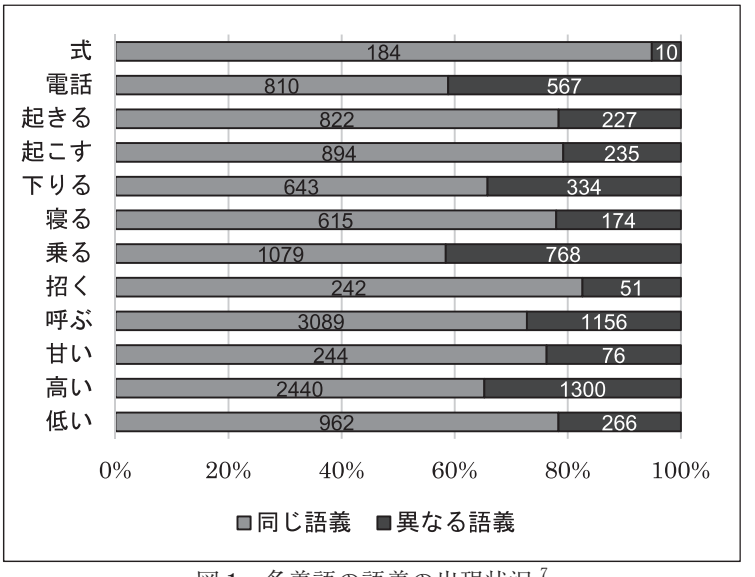

図 1　多義語の語義の出現状況 [7]

5.2　多義語の出現間隔

同一の多義語がテキスト中に複数出現する場合，ある出現 a とその次の出現 b との間の距離が，a と b の語義の異同にどのように影響するかに着目する．出現間隔の計測法は，2 つの方法を採用した．1 つは長単位で数えた [8] 際の連番の差である．例えば，あるテキストで特定の多義語 p が n 回出現したとして，それぞれの出現を出現順に p(1)，p(2)，p(3)，…，p(n) と表したとき，p(1) と p(2)，p(2) と p(3)，…，p(n-1) と p(n) の n-1 個のペアについて出現位置（先頭からの連番）の差で表す．

もう 1 つの計測法は調査語が出現する文に付与した連番の差である．文の認定は BCCWJ の <sentence> のタグに従っている．同一の文に調査語となる多義語が複数出現していれば，それらの出現間隔は 0 となる．

7　グラフ中の数字はサンプル数を表す．

8　出現間隔の差を計る場合，句読点などの記号は無視した．具体的には，長単位の持つ形態論情報の 1 つである品詞の値が「空白」「補助記号」「記号」となっているものを除外した．2 つめの計測方法である，文の場合も同様である．

表 4 は，出現間隔を語（長単位）で計測した場合の，表 5 は文で計測した場合の，それぞれの調査語が同じ語義のペアであった場合と異なる語義のペアであった場合の出現間隔の中央値を示した．出現間隔の分布を見ると，調査したほとんどの例で正規性，等分散性が認められなかったため[9]，マン・ホイットニーの U 検定を行った．その結果，「電話」を除くすべての例で 5％ 水準で有意差が認められた．すなわち，ある多義語の次に現れる同一の多義語が同じ語義である場合は出現間隔が相対的に狭く，異なる語義である場合は出現間隔が相対的に広いことが分かった．この現象の解釈としては語彙的結束性により，類縁性のある語義が選ばれやすいということが想定できる．表 4 では，「電話」が例外となっているが，語義区分とした「通話」と「電話機」の語義はメトニミー的に近い関係にあり，他の調査語における，同じ語義・異なる語義との関係性が違うことを指摘したい．

表 4　多義語の語義の出現間隔（語を単位として計測した場合の中央値）

調査語	同語義	異語義	MW 検定	調査語	同語義	異語義	MW 検定
式	48.5	427.5	*	乗る	189.0	715	**
電話	99	114.0	n.s.	招く	293.0	544.5	**
起きる	214.0	712.0	**	呼ぶ	252.0	617.5	**
起こす	265.5	824	**	甘い	178.0	570.5	**
下りる	242.0	515.5	**	高い	201.0	535.0	**
寝る	111.0	289.5	**	低い	219.0	703.0	**

(*：5％ 水準で有意差あり，**：1% 水準で有意差あり)

表 5　多義語の語義の出現間隔（文を単位として計測した場合の中央値）

調査語	同語義	異語義	MW 検定	調査語	同語義	異語義	MW 検定
式	3.0	28.0	*	乗る	13.0	48.0	**
電話	8.0	9.0	n.s.	招く	14.0	29.0	**
起きる	13.0	48.0	**	呼ぶ	14.0	39.0	**
起こす	15.5	57.0	**	甘い	13.0	39.0	**
下りる	17.0	38.0	**	高い	11.0	32.0	**
寝る	8.0	19.0	**	低い	12.0	46.0	**

(*：5％ 水準で有意差あり，**：1% 水準で有意差あり)

表 6 は多義語を構成する個々の語義について同じ語義か異なる語義かで出現間隔を長単位で計測したものである[10]．こられの例も正規性・等分散性がないものが多かったため，マン・ホイットニーの U 検定を行った結果，40 の事例のうち 33 事例で 5% 水準で有意差が見られた．表は省略するが出現間隔を文で測定した場合も同様の検定結果が得られた．

9　Shapiro-Wilk 検定の結果，5% の有意水準で正規分布に従うと言えないもののうち，例外となる例は「式」が異なる語義で使われた場合（文で計測）のみであった（p=0.091）．また F 検定の結果，「起きる」「式」「招く」「寝る」の 4 語に，5％ の有意水準で等分散性が認められないとは言えないものがあった．なお，本稿の統計分析はフリーソフト R を用いた．

10　「異なる語義」の距離の計測のしかたは次の通り．例えば「甘い①（味・香り）」の場合，「甘い①」が出現した場合，それ以外の語義での「甘い」が直近に来る場合の 2 語間の距離と，「甘い①」が後ろにあり，それ以外の語義での「甘い」が前（直近）にある場合の 2 語間の距離の算術平均．

このことから，多義語全体でみても，多義語を構成する個々の語義でみても語彙的結束性が関与しているのではないかということが想定できる．

表 6　多義語の語義の出現間隔（個々の語義）

調査語	同語義	異語義	MW 検定	調査語	同語義	異語義	MW 検定
式①	413.5	45.50	n.s.	乗る②	210.0	569.5	**
式②	122.5	345.0	n.s.	乗る③	227.5	595.0	**
式③	43.5	105.5	n.s.	乗る④	108.0	796	**
電話①	102.0	113.0	n.s.	乗る⑤	216.0	765.5	**
電話②	40.0	110.0	**	招く①	18.0	1236.0	n.s.
起きる①	64.0	530.0	**	招く②	210.0	750.0	**
起きる②	89.0	679.0	**	招く③	384.0	670.5	*
起きる③	246	712.0	**	呼ぶ①	224.5	591.0	**
起こす①	312.0	684.0	**	呼ぶ②	85.0	491.0	**
起こす②	104.0	794	**	呼ぶ③	244.0	588.0	**
起こす③	264.0	783.5	**	呼ぶ④	164.5	530.0	**
下りる①	256.0	460.5	**	甘い①	106.0	529.5	**
下りる②	131.0	456.0	**	甘い②	308.0	465.0	n.s.
下りる③	100.0	1038.0	**	甘い③	197.0	830.0	**
下りる④	71.5	906	**	高い①	214.0	589.5	**
寝る①	101.0	265.0	**	高い②	195.0	429.0	**
寝る②	116.0	222.0	**	高い③	79.0	380.0	**
寝る③	66.0	190.0	**	低い①	226.0	676.5	**
寝る④	33.0	-[11]	-	低い②	208.0	637.0	**
乗る①	159.0	610.5	**	低い③	118.0	259.5	*

(*：5％ 水準で有意差あり，**：1% 水準で有意差あり)

表 6 の中で唯一，「式①（方法・方式）」は同じ語義の場合よりも異なる語義のほうが出現間隔が短い．これは「式①」が出現頻度が低く，一方，「式③（数式・化学式等）」のような多数出現する例があることがその原因と考えられるが，似たような事情である「下りる③」「下りる④」「招く①」ではそのような現象は起きていないことから，別の理由が求められる．

表 6 から同じ語義の場合は出現間隔が短いことが分かったが，逆に，出現間隔が短くなれば，同じ語義になりやすいのだろうか．それを調査したものが図 2 である．図 2 は，出現間隔が長単位で 100 以下の「間隔小」の場合と 1,000 以上の「間隔大」の場合とで，その 2 語が同じ語義か異なる語義かの割合を求めたものである．「式」の例を除き，いずれの語についても間隔小のときのほうが間隔大のときよりも同じ語義の割合が大きくなっている．

11 「寝る④」の意味で直近に異なる語義の「寝る」が来る例はなかった．

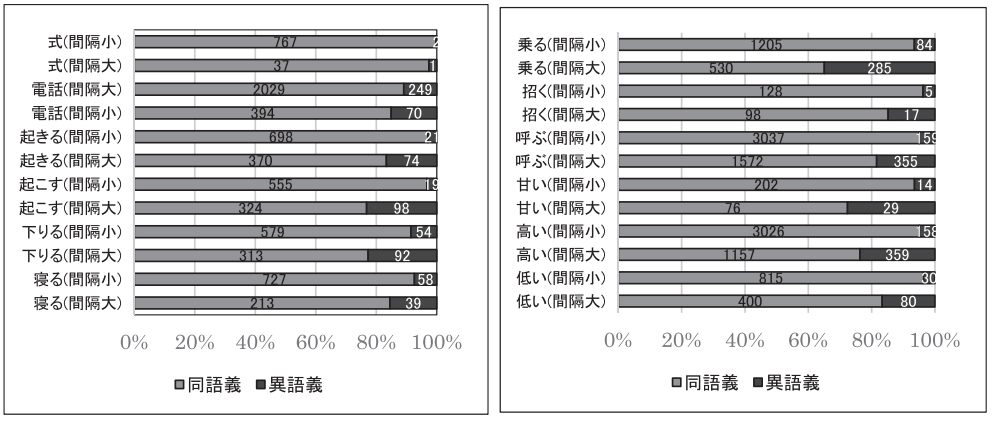

図 2　出現間隔と語義の一致傾向

5.3　類義語・対義語を含めた出現間隔

「起きる」「乗る・載る」「呼ぶ」「高い」の 4 語について，関連語（類義語・対義語）を含めた出現間隔の調査を行った．関連語を含めた出現間隔の測定のしかたは 5.2 で示したものと少し異なる．例えば「起きる②（目覚める）」の類義語の関連語「起こす②（目を覚まさせる）」との距離を調べる場合，他の語義も含めた「起きる」全体に「起こす②」を加えた語を測定の対象とする．あるテキストにおけるこれらの語の数を n とすると p(1)，p(2)，p(3)，…，p(n) という系列ができ，これを対象として出現間隔を測定する．「起こす②」が追加されていることから，同義語の場合の距離は，表 6 とはいくぶん異なる値になっている．また，どの関連語を追加するかで，同義語の場合の距離が少し変わってくる．

結果を表 7 に示す．調査語どうしの場合（同義語の場合）の距離と関連語との距離についてマン・ホイットニーの U 検定を行うと，調査した 14 組のうち，5 % 水準で有意差があったのは，8 例であった．有意差が見られなかった組は同語義の場合とおなじ程度の出現間隔ということになり，語彙的結束性が強い組み合わせと考えられる．

表7　関連語との出現間隔

調査語	関連語	調査語と関連語の距離	同語義の場合の距離	MW検定
起きる①	起こす①	259.5	61.0	*
起きる②	起こす②	81.0	84.0	n.s.
起きる②	眠る（睡眠）	151.0	62.0	**
起きる②	寝る②	93.0	89.0	n.s.
起きる③	起こす③	220.0	207.0	n.s.
起きる③	起こる	259.0	207.5	n.s.
高い①	低い①	141.0	197.0	n.s.
高い②	低い②	118.5	172.0	**
高い③	低い③	113.5	79.0	n.s.
高い③	安い（金額）	101.0	56.0	**
乗る①	下りる②	211.0	135.0	**
乗る②	下りる①	67.5	196.0	**
呼ぶ②	招く②	148.0	81.0	*
呼ぶ④	招く③	336.0	164.5	*

(*：5％水準で有意差あり，**：1% 水準で有意差あり)

6　まとめと今後の課題

本稿では，テキストにおける多義語の語義が1つの語義に偏りやすいが，必ずしも強い制約ではないことをコーパスを利用して明らかにした．また，その偏りは出現間隔が短いほど起こりやすいことを指摘した．これらの現象は語彙的結束性がひとまとまりのテキストに対して働いていることの現れであろうと推測される．

今後の課題としては，例外的な振る舞いをする語の説明，多義語における語義の違い方の程度と出現状況の関係，テキスト全体の語彙的結束性を計測する方法などについて研究を進めたい．

謝辞

本研究は，2009年～2012年にかけて行われた国立国語研究所萌芽発掘型共同研究プロジェクト「テキストにおける語彙の分布と文章構造」及び文部科学省科学研究費特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築」（領域代表者：前川喜久雄）の支援を受けた．なお，本発表は山崎 (2010) の発表を元にして発展させたものである．

参考文献・使用データ

Gale, W., K. Church, and D. Yarowsky. (1992) One sense per discourse. Proceedings of the 4th DARPA workshop on Speech and Natural Language, pp.233-237, Harriman, NY.

Halliday, M.A.K. and Hasan, R. (1976) Cohesion in English. London:Longman（邦訳『テクストはどのように構成されるか』，大修館書店，1997）

Krovetz,Robert. (1998) More than One Sense Per Discourse. Proceedings of the ACL-SIGLEX Senseval Workshop, ACL

柏野和佳子 (2007)「国語辞典における多義語の意味記述の比較」，言語処理学会第 13 回年次大会
北原保雄 (2010)『明鏡国語辞典第 2 版』，大修館書店
国広哲弥 (2006)『日本語の多義動詞（理想の国語辞典２）』，大修館書店
国立国語研究所 (2011)『現代日本語書き言葉均衡コーパス』
山崎誠 (2010)「テキストにおける多義語の意味実現の傾向」，計量国語学会第 54 回大会 2010 年 9 月，同予稿集 pp.25-30.

（2014 年 9 月 23 日受付）

Mathematical Linguistics, Vol.29 No.7 (December 2014) pp.251-262.

Report

Distribution of Senses of a Polysemous Word in Japanese Text:

Using "Balanced Corpus of Contemporary Written Japanese"

Makoto Yamazaki (National Institute for Japanese Language and Linguistics)

Abstract:
In this paper, we report how the senses of a polysemous word are distributed in a given text. Using the "Balanced Corpus of Contemporary Written Japanese," (BCCWJ) we received the following results. (1) Polysemous words are likely to share the same sense when occurring more than once in a given text. The tendency is not so strong as to admit no exceptions. (2) The gaps between instances of a polysemous word with the same sense are shorter than gaps with different senses and vice versa. (3) This phenomenon is also observed between a polysemous word and its synonyms and antonyms to some extent. These results suggest that the lexical cohesion positively affects the distribution of the senses of polysemous words as well as related terms to a lesser extent.

Keywords: polysemy, distance, lexical cohesion, Balanced Corpus of Contemporary Written Japanese