

『計量国語学』アーカイブ

ID	KK290401
種別	論文
タイトル	形態素の結合用法(語構成)と自立用法(句構成) —外来語系語基にみる量的関係—
Title	Bound-form Usage (Word Formation) and Free-form Usage (Phrase Formation) of Morphemes: The Quantitative Relation between Loanword Stems
著者	金 姝伶
Author	KIM Juyoung
掲載号	29巻4号
発行日	2014年3月20日
開始ページ	111
終了ページ	127
著作権者	計量国語学会

論文

形態素の結合用法（語構成）と自立用法（句構成）

—外来語系語基にみる量的関係—

金 姝伶（桃山学院大学非常勤講師）

要旨

形態素（語基）の結合用法と自立用法との関係を、新聞に現れた外来語系語基上位 50 種を対象として、主に量的な側面から調査・考察した。調査は、外来語系語基が漢語・外来語の語基と共起してつくる語構成（M結合）・句構成（W結合）を異なりでカウントし、非主要部・主要部別に、相関係数・共通度・差異係数を求めた。その結果、非主要部用法ではM結合とW結合の間に強い正の相関関係が認められたが、主要部用法では認められなかった。結合相手の共通度も、非主要部用法の方が主要部用法より大きかった。また、非主要部用法ではW結合が、主要部用法ではM結合が多かった。これらのことから、対象とした外来語系語基は、非主要部になる場合にはM結合とW結合との間に構文的な語形成の傾向、主要部になる場合には命名的な語形成の傾向をもつことが推測された。今後、個別の語基についてM結合とW結合との関係进行分析し、この推測を検証していく必要がある。

キーワード：M結合，W結合，連接率，相関係数，語彙の共通度，差異係数

1. 目的

派生形態論（語構成論）では、形態素は語基と接辞とに分けられるが、このうち語基は、原則として、他の語基や接辞と結合して合成語の要素となる「結合用法」と、そのまま単語として使われ文の要素となる「自立用法」とを兼ね備えているとされる（接辞や一部の語基には、結合用法しかない）。本稿は、この、語基の結合用法と自立用法との関係に、外来語系語基¹を対象として、主に量的な側面から接近することを目的とする。従来の派生形態論では、語基と接辞の区別の問題は重要視されたが、語基における結合用法と自立用法との関係については、語基の一般的な性質としても、個別の語基の特徴としても、詳しく検討されたことがない。本稿では、そうした検討の理論的な必要性・意義を明確にした上で、まずは両者の量的な関係の把握を試み、今後の質的な検討の足がかりを築くことをめざす。

2. 結合力，連接率

上述したようにこの方面の先行研究は少ないが、語基と接辞とを特に区別せず、形態素

1 「外来語系語基」という用語は、森岡(1985)にしたがう。

の結合用法と自立用法との関係を量的に追究した調査・研究に、斎賀(1957)と国立国語研究所(1972)とがある。

斎賀(1957)は、国立国語研究所の「総合雑誌の調査における標本度数十以上の語(延べ八万三千語)」について、それが単独に用いられた場合と、結合の部分として用いられた場合との、それぞれの延べ語数²から、結合の部分として用いられた場合の比率を求め、それを「語の結合力」と呼んでいる。ここで斎賀の言う「語」とは「 β 単位」すなわち基本的には形態素であり²、「それが単独に用いられた場合」とはその自立用法、「結合の部分として用いられた場合」とは結合用法に相当し、したがって、結合力とは、自立・結合全体の度数(延べ語数)に対する結合用法の度数の割合である。斎賀は、この結合力について、「現代の日本語では複合・派生等の語結合が実際にどのくらいの勢力で用いられているのか、またどのような種類の語が他の語と結合しやすく、どのような語が結合しにくいのか」を示す指標であるとし、「現代の日本語において、結合の要素として最も生産的な力を持っているのは漢語である。それに比べて和語の結合力は、実際にはかなり貧弱である」と結論づけている。

国立国語研究所(1972)は、「新聞の語彙調査」の一紙一年分のデータについて、短単位が長単位の構成要素になる割合を「連接率」として求めている。「短単位」とは β 単位と同じく基本的には形態素、「長単位」とは単語とおおよそ考えてよい。連接率の規定およびその計算法は、以下のように説明されている。

- (1)ここで、連接率とは次のようなものである。報告37は本報告と同じ資料によるものだが、その中の短単位度数順語彙表で「区」は度数1591である。ところが本報告位置別集計表の「区」は度数1583となっている。これは、1583回は「～区」などの形で接続して使われ、8回は「区では～」などの形で単独で使われたことを示している。ここで「区」が接続して使われた割合は

$$\text{“ 接続して使われた度数 ”} \div \text{“ 全体度数 ”} \times 100$$

の式で計算され、この場合、 $(1583 \div 1591 \times 100 \div 99.5)$ である。この数値を連接率という。

ここでいう「接続して使われた度数」とは短単位の結合用法の度数であり、連接率が斎賀(1957)の「結合力」とまったく同じ計算式による数値であることがわかる。国立国語研究所(1972)も、「連接率の高い語は、語構成要素になったことが多い語であり、連接率の高い語の中には、造語能力の高い語のほとんどが含まれている可能性が高い」とし、連接率を形態素の「生産的な力」(造語力)を示す尺度ととらえていることがわかる。

このように、結合力＝連接率は、形態素の「結合のしやすさ／しにくさ」を、結合用法の使用度数の、自立用法も含めた全体の使用度数に対する割合で表したものである。しかし、この考え方・方法はその使用度数に主眼をおいているため、結合の種類の多さを計るには別の指標が必要である。なぜなら、形態素の「結合のしやすさ」は、その使用度数では測れないからである。たとえば、上の「区」の例でいえば、確かに、結合用法での使用

2 「 β 単位」は、正確には、「形態素の1回結合」および「それに前後から結合する形態素」をいう。

がほとんどで、それに比べれば自立用法での使用は非常に少ないことがわかるが、これは、単なる使用度数をカウントしただけなので、極端な場合、たとえば「選挙区」という一つの単語が1583回使われたとしても、「区」の結合用法の使用度数は1583になってしまうのである。この数値は、直接には、「選挙区」という単語の「使われやすさ」を示すものであって、その要素である「区」の「結合のしやすさ」を示すものではない。この場合、「区」は「選挙」としか結びついていないのだから、その結合用法は、むしろ、きわめて制限されたものととらえなければならない。

形態素の「結合のしやすさ」については、石井（1992）がその形態素がいくつの異なる形態素と結合したかを指標としているが、本稿でもこの考え方を踏襲する。つまり、結合力＝連接率のように使用度数（延べ語数）で測るのではなく、結合相手となる形態素の異なり語数をカウントするのである。その際、接辞のように結合用法しかない形態素の場合はその数値だけでよいが、語基のように自立用法も備えている形態素の場合は、その「結合のしやすさ」すなわち結合用法の働き具合が、「自立のしやすさ」すなわち自立用法の働き具合とどのような関係にあるかをみることも重要になる。その意味で、結合用法と自立用法とを比較する結合力＝連接率の考え方は、基本的には間違っていない。

しかし、ここで大きな問題にぶつかる。結合用法は異なりでカウントするとして、では、自立用法はどのようにカウントすればよいのだろうか。

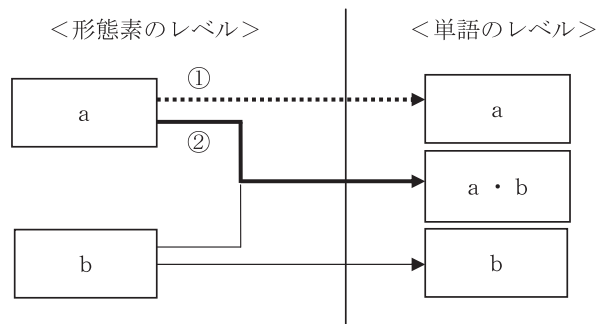


図1：形態素の自立と結合

派生形態論の図式では、形態素の自立と結合とは、図1のようにとらえられる。形態素「a」に注目すれば、①が自立用法であり、②が結合用法である。ここで①と②とを異なり語数で比較しようとすれば、②は結合相手となる語基の異なり語数を使えばよいが、①は結合相手というものがないから、異なり語数は意味をもたない。①が意味をもつのは、それが一定の文章や談話の中で何回行われたかという使用度数である。とすれば、①と②との比較は、結局、結合力＝連接率が行ったように、使用度数（延べ）でしかできないということになる。

3. M結合とW結合

従来の派生形態論では、自立とは、形態素がそのまま単語になること（図1の①）であ

るが、本稿では、自立もまた（単語としての）結合ととらえ³、その結合力を計るために、句を計量の対象に新たに導入して、比較の論点を、従来からの結合用法と、自立した状態で他の単語と結合する用法とに展開する。これを図示すれば、図2のようになる。形態素「a」に注目すれば、それは①で単語となり、同時に、③で他の単語と結合して、文の直接的な要素としての「句」を構成する。形態素の自立とは、①だけではなく、①と③とを合わせた過程ととらえることができる。そう考えれば、自立用法もまた、それが句を構成するときの結合相手を異なりで数えることが可能になる。

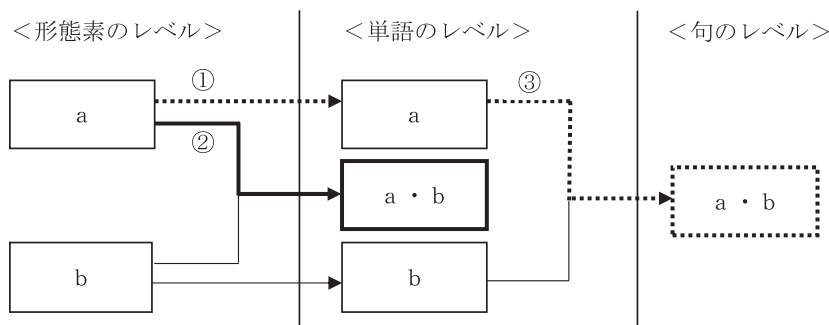


図2：M結合（→）とW結合（→）

このように、本稿では、形態素の自立用法を、従来の自立（①）に句構成（③）を加えた過程とし、それを（wordの頭文字をとって）「W結合」と呼ぶことにする。これに合わせて、結合用法（②）も（morphemeの頭文字をとって）「M結合」と呼び換えることにする。そして、両者の関係を、ともに、その（M結合の場合は合成語の、W結合の場合は句の）異なり語数でカウントすることによって、量的に把握する。

これまで、形態素の自立用法は①に限定されていたため、結合用法との関係は延べでしか比較できなかった。また、同じく従来、③は、構文論（句論、連語論）の問題として、①や②とは異なるレベルにあるものとされてきたが、それを①とともに形態素の自立用法ととらえることにより、語基の結合用法（M結合）の働き具合（結合のしやすさ）と自立用法（W結合）の働き具合（自立のしやすさ）とをともに、それぞれの結合（合成語と句）の異なり語数としてカウントすることが可能になり、両者の関係を量的に追究する方がひらけることになるのである。

ところで、結合力＝連接率のもう一つの問題は、その計算式において、注目する形態素が合成語の主要部（後要素）になる場合と非主要部（前要素）になる場合とを区別していないことである⁴。先の「区」の結合用法でも、主要部（「～区」）の度数は1502、非主要部（「区～」）の度数は16（「中部分」65を除く）で、これらを別に計算すれば、それぞれの連接率は大きく異なるはずである。ただし、自立用法については主要部・非主要部

3 この点については、斎藤（2004）の「単語化」の理論を考慮する必要があるが、「単語化」の概念は複雑で、ここでいう「自立」との異同は単純にはとらえられない、今後の課題としたい。

4 本稿では、影山（1993）に従い、句結合と語結合の後要素を主要部（head）、前要素を非主要部（non-head）と呼ぶことにする。

を区別できないので、結局、連接率を主要部・非主要部別に求めることはできない。しかし、結合のしやすさ、自立のしやすさを考えようとするとき、主要部の場合と非主要部の場合とを区別することは非常に重要である。その点、M結合・W結合では、M結合（合成語）だけでなく、W結合（句）についても主要部・非主要部の区別が可能であり、それぞれにおける比較ができる。図3・4は、「メール」という語基を例に、主要部・非主要部それぞれのM結合・W結合を示したものである。語基「メール」のM結合とW結合との関係は、それぞれに計測・検討されることになる。

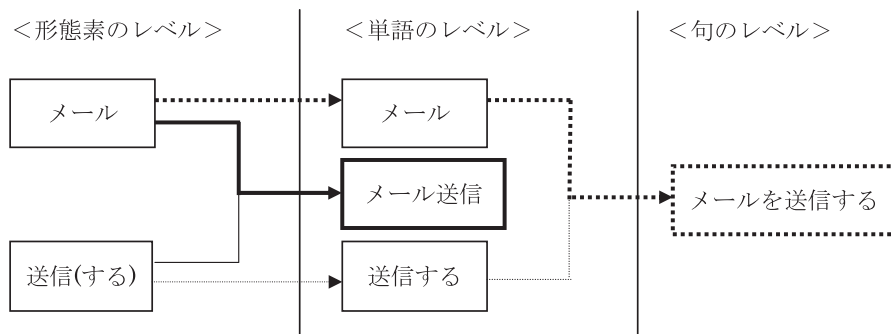


図3：非主要部のM結合（→）とW結合（……）

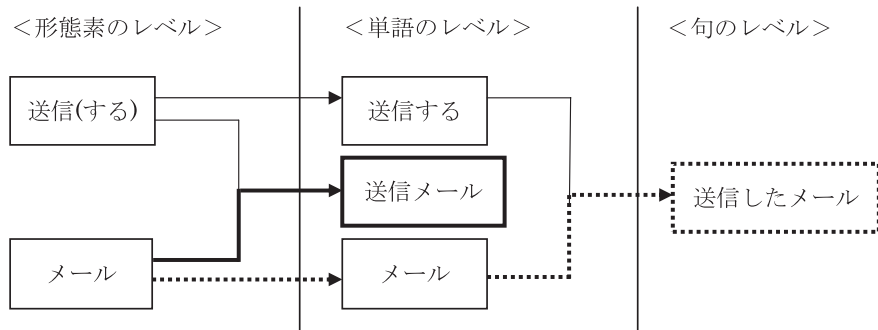


図4：主要部のM結合（→）とW結合（……）

このように整理することによって、M結合とW結合の関係を正しく計ることができるものと考えられる。そして、このことは、主要部・非主要部における語形成のありかたの異同を把握することの前提となる。そうした把握が成功すれば、その結果は、自動抄録や日本語教育への応用につなげることも可能になる（句から自動的に字数の少ない複合語に変換できるかどうかの判定材料になる、複合語に交代可能な句を先に教えることで学習の効率化が期待される、など）。

4. 調査：新聞における外来語系語基のM結合とW結合

形態素（語基）の結合用法・自立用法の関係を、M結合・W結合という観点・方法から

明らかにすべく、新聞に用いられた外来語系語基を対象に調査を行った。新聞記事を資料とするのは、(2)のように、M結合とW結合とが交替可能であるような使用が、他のジャンルやテキストタイプに比べて多いと考えられるからであり、外来語系語基を対象とするのは、和語系・漢語系の語基に比べて新しいものが多く、そのため、語彙化せずにW結合と交替可能なM結合が多いと考えられるからである。

- (2)容疑は、交際していた20代の女性に別れを告げられたことに立腹。女性が結婚した後、昨年10月9日ごろから11月19日にかけて、女性の30代の夫に対し、「(O容疑者と)交際していた時の女性の写真をインターネット上に流してもいいのか」「こんな女と結婚しているとばらすぞ」などの嫌がらせのメールを送ったり、女性の行動を監視して逐一その行動をメール送信したとしている。O容疑者は同様の内容のメールを女性にも送信していたといい、県警は女性へのストーカー行為についても調べる⁵。
(『毎日新聞』2010年2月3日付社会面)

ここでW結合(句)と交替可能なM結合(合成語)とは、林(1982)のいう「臨時一語」の一種であるが、今回の調査では、得られたM結合を臨時一語か語彙化された合成語であるか判別して前者のみを対象とする、ということは行わなかった。合成語の語彙化の程度には、これも林(1991)のいう「固定結合／安定結合／浮動結合」の3段階があるとされるが、個別のM結合についてそれがどの段階にあるかを正確かつ統一的に判別することは困難である。今回の調査で対象とする外来語系語基のM結合には、語彙化されたものから臨時一語であるものまで幅広く含まれ⁶、そのすべてがW結合と交替可能な関係にあるわけではないが、語彙化された合成語の多い和語系・漢語系の語基に比べれば、相対的に交替可能なものが多く、したがって、M結合とW結合との共時的な関係を考える上でより適していると考えたわけである。

4.1 対象とした外来語系語基

2006年の『毎日新聞』全紙面を資料として⁷、その(見出しを除く)記事本文の語彙調査を行い、使用頻度順50位までの外来語系語基を調査の対象とした。その際、人名・地名・国名などの固有名詞、ローマ字の頭字語(WTO, CD, MBSなど)、漢字表記語(護謄、瓦斯、煙草など)、単位を表すもの(キロ、ミリ、ワット、ドル、ユーロなど)は、除いた。なお、ここでの「使用頻度」とは、M結合・W結合を区別しないものである。

語基の認定にあたっては、「ホームラン」のように、(原語ではともかく日本語としては)1語基とすべきか(「ホーム」と「ラン」の)2語基とすべきか迷うようなものについては、国語辞典の見出しの示し方を参照し、辞典間でゆれがあるような場合は1語基とした⁸。また、森岡(1985,1994)では外来語系の接辞とされている「インター」や、準接辞

5 用例の下線や省略は筆者による。

6 石井(2007)のいう「既成の複合語／新造の複合語／臨時の複合語」を区別しないということでもある。

7 『CD-毎日新聞2006データ集』を、大阪大学大学院文学研究科日本語学講座が毎日新聞社と交わした利用許諾契約・覚書に基づき使用した。

8 たとえば、『新選国語辞典(第八版)』では、「ホーム」という親見出しの下に「一ラン」が子見出しとして立てられているが、『新明解国語辞典(第六版)』では「ホームラン」で単独の見出しになっている。

(接辞とは言えないが接辞に類する働きをするもの、準用の接辞)とされている「スーパー」も、本調査では語基として扱った。このほか、「パソコン」の「コン」、「ワールドカップ」の「ワールド」のように自立用法がない語基についても、調査対象から外すことはしなかった(これらはW結合の異なり語数が0となるだけである)。また、いわゆるサ変動詞語幹については、それをいわゆる動名詞語基と認めて、自立してW結合をつくる場合にはサ変動詞にもなるものと考えた。したがって、たとえば「アピール」という語基は、そのままの形でM結合(「アピールポイント」「自己アピール」など)をつくり、また、名詞としてW結合(「アピールに賛同する」「国民へのアピール」など)をつくるほか、動詞としてもW結合(「アピールする機会」「カメラにアピールする」など)をつくるものとし、「アピールする」を「アピール」と「する」とのM結合と考えることはしなかった。

4.2 結合相手(共起語基)の範囲

調査では、対象とした外来語系語基が、M結合・W結合において結びつく形態素(以下、「共起語基」⁹とする)の範囲を、漢語と外来語に限定し、和語と混種語との結合については考慮しないことにした。一般に、日本語の語構成(M結合)にはかなり強い語種の制限があり、外来語系語基は、たとえば「メールチェック」や「メール送信」は可能でも「メール調べ」や「メール送り」とは言いにくいように、同じ外来語系か漢語系の語基とはよく結合するが、和語系の語基とは結合しにくい。一方で、句構成(W結合)にはそのような語種制限はないから、外来語系語基の結合相手としての和語系語基には、M結合の共起語基にはなりにくく、W結合の共起語基にはなりやすいという不均衡があり、そのことが調査結果をゆがめる可能性がある。このほか、接辞の類は、M結合の共起形態素にはなってもW結合の共起形態素にはならないから、やはり結合相手からは除いた。ただし、語基か接辞かあいまいないわゆる造語成分(語素)は、自立用法の有無の認定が国語辞典によって異なる場合もある(山下 1995)ので、今回は、結合相手から除外していない。

4.3 M結合の抽出

ここでいうM結合は、外来語系語基が漢語系ないし外来語系の語基と直接結合して、1次結合の複合語(例「メールサービス」)を構成しているものとする。2次結合以上の複合語(例「電子メールサービス」)の場合は、外来語系語基が1次結合している部分(例「電子メール」)のみを対象とする。また、先に述べたように、主要部での結合(例「送信メール」)と非主要部での結合(例「メール送信」)とを区別する。

4.4 W結合の抽出

ここでいうW結合は、外来語系語基が単独の語として漢語系ないし外来語系の語基と結合し、文中で句を構成しているものとする。ただし、W結合の場合は、M結合と違って1次結合に限定せず、直接の結合関係が想定できる句はすべて抽出することにした¹⁰。(3)では「ブログのサービス」と「サービスを提供する」、(4)では「作品をアピールする」と「アピールする方法」という二つの句が、それぞれ、W結合として抽出されることになる。文は複合語よりも構造的に柔軟かつ複雑だから、こうした複数の句を想定することは、不

9 W結合の共起語基は、W結合の内部要素としては「共起語」と呼ぶべきであるが、ここでは、外来語系語基と結びついてW結合をつくり得る(形態素レベルにある)ものとして「共起語基」と呼ぶこととした。

10 W結合は構文解析ソフト「CaboCha」の解析結果に手作業で修正を加えて抽出した。

自然ではない。逆に、1次結合だけに限定してしまうと、想定できる多くの句を無視してしまうことになる。

- (3) ブログのサービスを提供しているサイトを利用すると誰でも手軽にホームページを公開できる。(『毎日新聞』2006年6月10日付特集面)
- (4) 「日本では作品をアピールする方法に試行錯誤する。今では不安なく創作に打ち込める」とシャーマさんは言う。(『毎日新聞』2006年6月6日付社会面)

ただし、(5)(6)のように、外来語系語基の結合する相手が語よりも大きな単位の場合には、今回は対象外とし、その主要素と外来語系語基との結びつきをW結合とすることはしていない。(5)では「パーティー参加や1口5000円の政治献金を呼びかけるよう依頼する」と「メール」、(6)では「迷惑をかける」と「チーム」が結びついているが、ここから「依頼するメール」「チームにかける」といった句をW結合として抽出することはしなかった((5)の「メールを送信する」はW結合として抽出した)。

- (5) 調べでは、石阪市長らは昨年11月29日の政治資金パーティーを控えた同5日、横浜市の局区長ら約80人に対し、パーティー参加や1口5000円の政治献金を呼びかけるよう依頼するメールを送信。(『毎日新聞』2006年6月7日付社会面)
- (6) 死球に苦しんだ昨年は、直前のヤクルト戦で暴力をふるい、交流戦で10試合出場停止となり、チームに迷惑をかけた。(『毎日新聞』2006年6月5日付スポーツ面)

W結合も、M結合と同様、主要部での結合(例「送信したメール」)と非主要部での結合(例「メールを送信する」)とを区別する。

4.5 調査結果

以上のような方法で、新聞における外来語系語基のM結合・W結合を抽出し、それぞれの異なり語数を、主要部・非主要部の用法別にカウントして、表1のような結果を得た。ただし、ここでの「異なり語数」とは、正確には、「異なり共起語基数」というべきものである。それは、たとえばW結合の場合、「メールを返信する」と「メールで返信する」では、「メール」と「返信する」との意味的な関係は同じではなく、またM結合の場合も、「メール返信」というM結合の意味が<メールを返信する>である場合と<メールで返信する>の意味である場合とがあり得るが、今回の調査では、これらを区別することはせず、いずれも「返信(する)」を共起語基とする同じW結合、同じM結合と認めたということである。共起語基を同じくするM結合・W結合における意味の違いをどう処理するかについては、今後の課題としたい。なお、W結合では、「メールを返信する」と「メールの返信」のような場合でも、サ変動詞「返信する」と名詞(サ変動詞語幹)「返信」とを同一の語基(動名詞)と認めて、同じW結合とした。これは、M結合との比較上も必要な処置である。

5. 分析: M結合・W結合の量的関係

表1の結果をもとに、M結合とW結合との量的関係の把握を試みる。分析の指標は、相

表1：外来語系語基（新聞高頻度 50 種）のM結合とW結合

No.	語基 (*動名詞)	総頻度	非主要部					主要部				
			M結合	W結合	共通	共通度	差異係数	M結合	W結合	共通	共通度	差異係数
1	チーム	8117	143	335	41	0.09	-0.40a	274	104	15	0.04	0.45a
2	リーグ	6575	85	31	5	0.05	0.47a	56	13	5	0.08	0.62a
3	テレビ	5710	149	169	32	0.11	-0.06	44	33	4	0.05	0.14
4	ネット	4985	137	155	16	0.06	-0.06	27	3	0	0	0.80a
5	グループ	4594	83	134	21	0.11	-0.24a	257	56	10	0.03	0.64a
6	センター	4581	38	55	5	0.06	-0.18	178	6	0	0	0.93a
7	テロ	4061	88	139	34	0.18	-0.22a	48	44	5	0.06	0.04
8	サッカー	3487	116	110	16	0.08	0.03	20	34	2	0.04	-0.26
9	プロ	3487	108	110	17	0.08	-0.01	13	50	0	0	-0.59a
10	トップ	3174	84	114	11	0.06	-0.15c	146	136	25	0.1	0.04
11	ボール	3041	27	53	9	0.13	-0.33b	73	34	6	0.06	0.36a
12	サービス *	2905	88	116	22	0.12	-0.14c	279	95	16	0.04	0.49a
13	システム	2890	81	183	34	0.15	-0.39a	426	116	20	0.04	0.57a
14	データ	2861	63	134	26	0.15	-0.36a	180	182	35	0.11	-0.01
15	コン	2782	0	0	0	0	-	8	0	0	0	1.00b
16	プレー *	2707	23	108	6	0.05	-0.65a	54	142	6	0.03	-0.45a
17	メーカー	2634	26	132	6	0.04	-0.67a	307	26	7	0.02	0.84a
18	ケース	2593	7	73	0	0	-0.83a	47	83	6	0.05	-0.28b
19	インター	2483	12	0	0	0	1.00a	2	0	0	0	1.00
20	メンバー	2481	22	120	7	0.05	-0.69a	167	123	25	0.09	0.15b
21	ホーム	2353	46	20	3	0.05	0.39b	19	7	1	0.04	0.46c
22	メール	2346	42	105	22	0.18	-0.43a	60	54	12	0.12	0.05
23	クラブ	2313	62	89	8	0.06	-0.18c	98	23	2	0.02	0.62a
24	テーマ	2255	12	123	4	0.03	-0.82a	63	139	18	0.1	-0.38a
25	ゴール *	2067	34	49	3	0.04	-0.18	32	45	4	0.05	-0.17
26	スタート *	2065	20	68	3	0.04	-0.55a	50	171	17	0.08	-0.55a
27	スポーツ	2055	138	101	12	0.05	0.15c	36	30	4	0.06	0.09
28	エネルギー	1994	144	97	21	0.1	0.20b	40	58	4	0.04	-0.18
29	ゲーム *	1948	82	76	10	0.07	0.04	104	40	3	0.02	0.44a
30	ワールド	1859	12	0	0	0	1.00a	6	0	0	0	1.00c
31	カップ	1841	7	3	0	0	0.40	12	1	0	0	0.85b
32	イメージ *	1827	43	125	18	0.12	-0.49a	53	222	10	0.04	-0.61a
33	カード	1792	37	59	6	0.07	-0.23c	142	41	6	0.03	0.55a
34	ページ	1749	6	18	0	0	-0.50c	23	25	0	0	-0.04
35	シリーズ	1457	33	36	2	0.03	-0.04	88	16	1	0.01	0.69a
36	スーパード	1456	77	43	2	0.02	0.28b	16	24	1	0.03	-0.20
37	スピード	1401	60	53	11	0.11	0.06	18	74	2	0.02	-0.61a
38	レース	1346	40	49	6	0.07	-0.10	80	40	2	0.02	0.33a
39	ニュース	1309	25	53	3	0.04	-0.36b	42	55	9	0.1	-0.13
40	トラブル	1303	28	51	8	0.11	-0.29b	40	49	4	0.05	-0.10
41	シーズン	1250	37	24	4	0.07	0.21	57	22	0	0	0.44a
42	オープン *	1242	40	47	4	0.05	-0.08	22	38	1	0.02	-0.27c
43	ルール	1221	23	71	12	0.15	-0.51a	110	107	16	0.08	0.01
44	ホール	1212	11	15	0	0	-0.15	14	10	0	0	0.17
45	アピール *	1195	10	82	2	0.02	-0.78a	16	169	4	0.02	-0.83a
46	マーク *	1181	3	56	1	0.02	-0.90a	45	41	2	0.02	0.05
47	ツアー	1164	57	40	5	0.05	0.18	77	15	3	0.03	0.67a
48	デビュー *	1163	24	17	3	0.08	0.17	74	44	5	0.04	0.25b
49	リード *	1029	14	51	0	0	-0.57a	7	75	3	0.04	-0.83a
50	コーチ *	894	16	29	2	0.05	-0.29	49	24	3	0.04	0.34b

関係数、共起語基の共通度、差異係数の三つである。

5.1 相関係数

はじめに、M結合数とW結合数との相関関係を、非主要部・主要部それぞれにみる。図5は、M結合・W結合それぞれの異なり語数を、前者を縦軸、後者を横軸に、縦横同じスケールで散布図としたもので、左が非主要部、右が主要部でのものである。

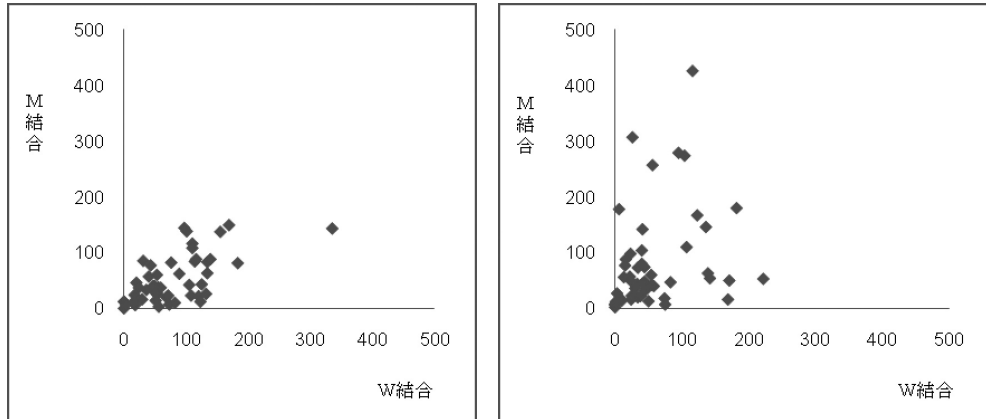


図5：M結合・W結合の散布図（左：非主要部、右：主要部）

一見して、主要部より非主要部の方が、明確な相関が認められるようである。相関係数を求めると、非主要部は0.63 ($p < 0.001$)、主要部は0.27 ($p > 0.05$)で、やはり、非主要部の方に高い正の相関が認められる。

このことは、非主要部の用法では、たとえばW結合が多い語基ほどM結合も多くなる、すなわち、句を活発につくる語基ほど複合語も活発につくるという傾向が強いわけで、W結合（自立用法）とM結合（結合用法）との間に、より強い関連性のあることが推測される。図6は、非主要部の散布図を拡大し、各語基の位置がわかるようにしたものだが、「チーム」がやや外れた位置にあるほかは、おおよそ、右肩上がりにまとまって分布している。

一方、主要部の用法では、こうした関連性は弱く、W結合の多いことが必ずしもM結合の多さに結びつかない、あるいは、W結合が多くなってもM結合が多いものがあるといった傾向がみられる。主要部の散布図を拡大した図7をみると、全体が扇形に広がった分布になっていて、とくに、図の右下、すなわち、W結合が多い割にM結合が少ないものに、「アピール（する）」「イメージ（する）」「スタート（する）」「プレー（する）」「リード（する）」といった動名詞語基（サ変動詞語幹）の多いことが目を引く。図の左上の、W結合が少ない割にはM結合が多いものには、「システム」「メーカー」「サービス（する）」「チーム」「グループ」「センター」などがある。

5.2 共起語基の共通度

次に、M結合とW結合とで、その結合相手である共起語基の共通度をみる。計算には、異なり語の重なりをみるのに適当な、宮島・水谷（1958）の「語彙の共通度」を用いる。

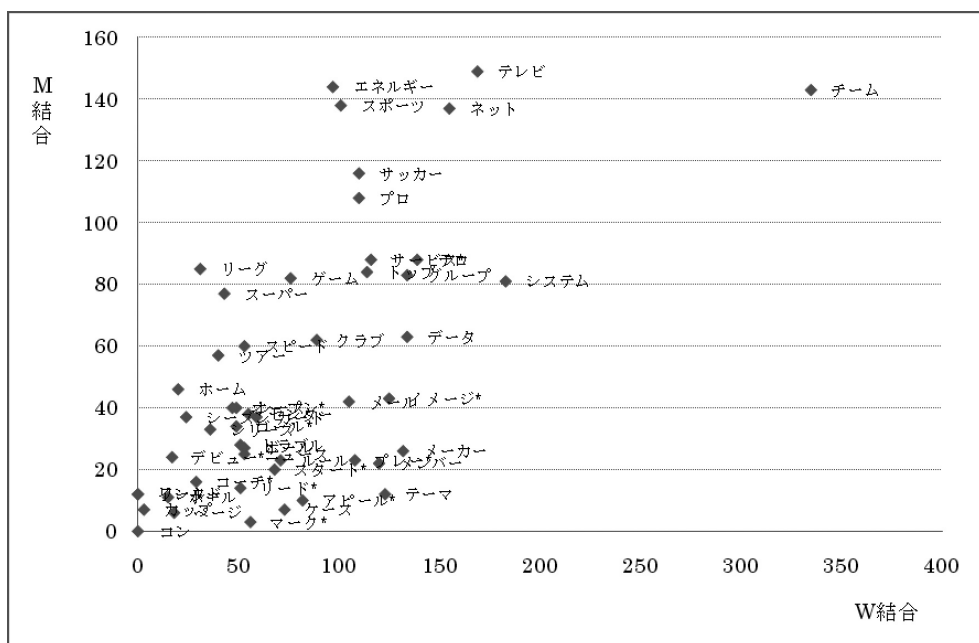


図6：非主要部の散布図（*は動名詞語基）

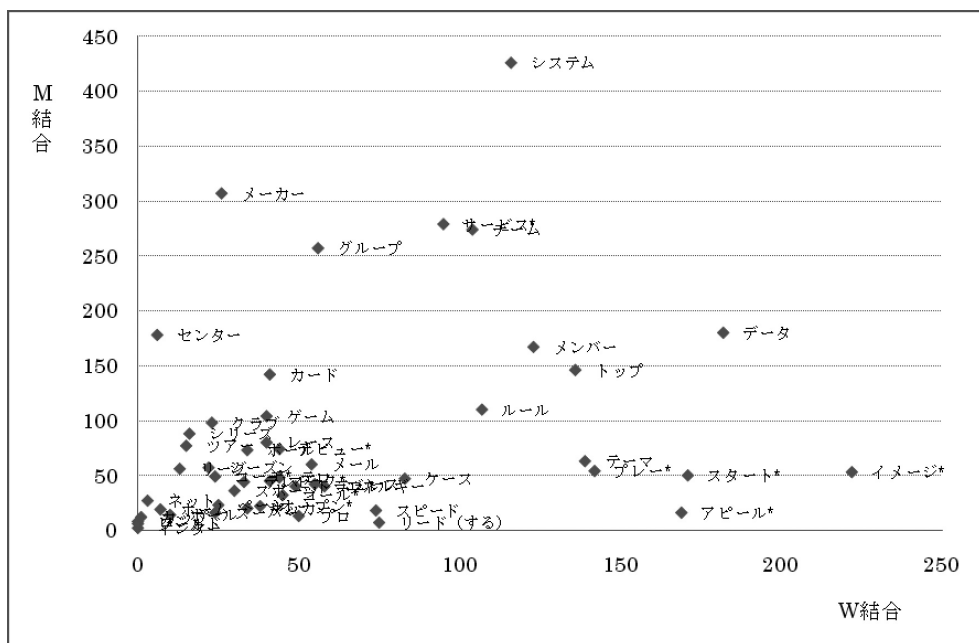


図7： 主要部の散布図（*は動名詞語基）

計算式は、以下の通り．ここで、 $a \cdot b$ は二つの語彙の見出し語（異なり語）の数、 x はこれらに共通の見出し語の数である．共通度 C_{AB} は、共通する語がまったくない0から、すべて一致する1までの間の値をとる．

$$C_{AB} = \frac{x}{a + b - x} \quad (1)$$

ここでは、語基 s について、M結合での共起語基の異なり数を a_s 、W結合での共起語基の異なり数を b_s とし、両者に共通する共起語基の異なり数を x_s として、その共通度 C_s を求める．表1では、「共通」欄に x_s の値、「共通度」欄に C_s の値を示している．

いま、非主要部と主要部とに分けて、共通度および共通する共起語基数の五数要約値を求めると、表2のようになる．最小値を別にして、いずれの値も非主要部の方が大きい．非主要部の方が、M結合とW結合との共通度が大きい、すなわち、関連性が強いわけで、これは、すでにみた相関係数の結果と矛盾しない．

表2： 共通度・共通共起語基数の五数要約値

	非主要部		主要部	
	共通度	共通共起語基数	共通度	共通共起語基数
最大値	0.18	41	0.12	35
第1四分位	0.10	16	0.05	9
中央値	0.06	6	0.04	4
第3四分位	0.03	2	0.02	1
最小値	0.00	0	0.00	0

表3： 共通共起語基数の割合の五数要約値

	M結合		W結合	
	非主要部	主要部	非主要部	主要部
最大値	0.52	0.43	0.24	0.38
第1四分位	0.29	0.13	0.17	0.16
中央値	0.15	0.07	0.12	0.09
第3四分位	0.09	0.03	0.06	0.04
最小値	0.00	0.00	0.00	0.00

また、共通する共起語基数の、M結合数・W結語数（正確には、M結合・W結合の共起語基数）に占める割合を、非主要部・主要部別に求め、その五数要約値をとると、表3のようになる（W結合を作らない「インター」「ワールド」、非主要部の用法がない「コン」を除く）．これをみると、W結合では大きな差はないが、M結合では、明らかに、非主要部の方が主要部より共通する共起語基の割合が大きい．このことも、非主要部の方が主要部より、M結合がW結合に包摂される傾向、すなわち、両者の関連性が強いことを示している．

5.3 差異係数

最後に、個別の語基について、M結合とW結合の関係をみる．指標には、差異係数を用いる．「差異係数」とは、一般に、（規模を同じくする）二つのコーパスA・Bにおける語

w の出現傾向を表す指標で、w の A における頻度を X_w 、B における頻度を Y_w とすると、その差異係数 D_w は以下の計算式で求められる（齋藤・中村・赤野 1998）。

$$D_w = \frac{X_w - Y_w}{X_w + Y_w} \quad (2)$$

ここで、 D_w は、+ 1.00 と - 1.00 の間の値をとり、正の場合には語 w がコーパス A により多く出現し、負の場合にはコーパス B により多く出現することを示す。このように、差異係数はコーパスにおける出現頻度に基づくことが多いが、ここでは、語基が M 結合と W 結合のどちらをより多くつくっているかを見るための指標として用いる。すなわち、語基 s について、その M 結合の異なり語数を X_s 、W 結合の異なり語数を Y_s とし、(2) と同様の計算式で、s の差異係数 D_s を求めるのである。ここでは、正の場合には M 結合が多く、負の場合には W 結合が多いことになる。表 1 の「差異係数」欄には、この値を示している。

差異係数の値に a・b・c の記号を添えたものは、それぞれ、有意水準 0.001, 0.01, 0.05 で有意差ありと認められたものである。ここでは、語基 s における M 結合・W 結合の生起確率、すなわち、それぞれの異なり語数は等しくなると仮定し、いずれも “ $(X_s + Y_s) / 2$ ” を期待値として、自由度 1 のカイ 2 乗検定を行った。

いま、非主要部・主要部のそれぞれで有意差の認められた語基（ $P \leq 0.05$ ）について、M 結合と W 結合との量的関係を、非主要部・主要部別に整理すると、表 4 のようになる（差異係数の降順）。これをみると、非主要部では、M 結合より W 結合の方が多い語基が多いが、主要部では、W 結合より M 結合の方が多い語基が多い。

表 4：差異係数（有意差あり）からみた外来語系語基の分布（* は動名詞語基）

	M結合＞W結合	M結合＜W結合
非主要部	ワールド、インター、リーグ、ホーム、スーパー、エネルギー、スポーツ	マーク＊、ケース、テーマ、アピール＊、メンバー、メーカー、プレー＊、リード＊、スタート＊、ルール、ページ、イメージ＊、メール、チーム、システム、データ、ニュース、ボール、トラブル、グループ、カード、テロ、クラブ、トップ、サービス＊
主要部	コン、ワールド、センター、カップ、メーカー、ネット、シリーズ、ツアー、グループ、リーグ、クラブ、システム、カード、サービス＊、ホーム、チーム、ゲーム、シーズン、ボール、コーチ＊、レース、デビュー＊、メンバー	リード＊、アピール＊、イメージ＊、スピード、プロ、スタート＊、プレー＊、テーマ、ケース、オープン＊

さらに、非主要部・主要部の両方で有意差が認められた語基について、M結合とW結合との量的関係をみると、次のようになる（*は動名詞語基）。

- 1) 非主要部・主要部の両方で、M結合＞W結合
ワールド、リーグ、ホーム
- 2) 非主要部・主要部の両方で、M結合＜W結合
アピール*、イメージ*、ケース、スタート*、テーマ、プレー*、リード*
- 3) 非主要部ではM結合＞W結合、主要部ではM結合＜W結合
なし
- 4) 非主要部ではM結合＜W結合、主要部ではM結合＞W結合
カード、クラブ、グループ、サービス*、システム、チーム、ボール、メーカー、メンバー

ここでも、最も多かったのは4)のタイプで、表4の傾向と一致する。非主要部ではW結合の方が多く、主要部ではM結合の方が多いう語基が多いのである。

6. 結論と今後の課題

以上、2006年の『毎日新聞』全紙面から得た使用頻度順50位までの外来語系語基について、そのM結合とW結合の異なり語数を調査し、相関係数・共起語基の共通度・差異係数の三つの指標を使って、M結合とW結合の量的な関係を、非主要部・主要部の用法別に整理した。その結果、非主要部では、M結合数とW結合数との間に高い正の相関が認められ、共起語基の共通度が（主要部より）大きく、M結合よりW結合の方が多い外来語系語基の多いこと、逆に、主要部では、M結合数とW結合数との間に明確な相関関係は認められず、共起語基の共通度も（非主要部より）小さく、W結合よりM結合の方が多い外来語系語基の多いことが、それぞれ、明らかになった。

この結果は、次のような推測を可能にする。すなわち、外来語系語基が非主要部に立つ用法では、M結合とW結合との関連性が強く、かつ、W結合がM結合をつくる基盤となる可能性が（主要部に比べて）大きい。このとき、M結合については、W結合を凝縮した臨時一語となること、すなわち、石井（2007）のいう「構文的な複合語形成」の傾向が強く、W結合については、M結合の基底構造となるような句である傾向が強い。一方、主要部に立つ用法では、M結合とW結合との関連性が弱く、かつ、W結合がM結合をつくる基盤となる可能性は（非主要部に比べて）小さい。このとき、M結合については、W結合の凝縮とは異なる語彙的な性格、すなわち、石井（2007）のいう「命名的な複合語形成」の傾向が強く、W結合については、M結合の基底構造となるような句である傾向は弱い、という推測である。

表5:「メール」のM結合の共起語基

	W結合と共通する	W結合と共通しない
非主要部	～チェック, ～データ, ～原案, ～交換, ～作成, ～作成者, ～受信, ～相談, ～送受信, ～送信, ～送信者, ～仲介, ～仲介者, ～提供, ～提供者, ～内容, ～入手, ～配信, ～発信, ～無料, ～問題, ～利用	～アカウント, ～アドレス, ～コンプライアンス, ～サーバー, ～サービス, ～ソフト, ～マガジン, ～以前, ～画面, ～環境, ～疑惑, ～経由, ～事件, ～自体, ～情報, ～世代, ～騒動, ～代, ～着信, ～論戦
主要部	偽造～, 携帯～, 携帯電話～, 激励～, 抗議～, 受信～, 相談～, 中傷～, 匿名～, 批判～, 返信～, 連絡～	E～, HTML～, SOS～, ウイルス～, ショート～, スカウト～, スパム～, ターゲティング～, チェーン～, デコ～, フィッシング～, ママ～, わいせつ～, 愛～, 一斉～, 応援～, 怪文書～, 確認～, 偽装～, 脅迫～, 業務～, 緊急～, 苦情～, 警告～, 広告～, 号外～, 告発～, 災害～, 社内～, 受信拒否～, 新年～, 送金～, 送金指示～, 送信～, 速報～, 注意～, 訂正～, 電子～, 登録完了～, 投稿～, 当選～, 同性愛～, 文字～, 防災～, 無料～, 迷惑～

このこと的一端を、語基「メール」について確認してみよう。いま、「メール」のM結合について、それがW結合と（その共起語基において）どの程度共通しているかをみると、表5のようになる。これをみると、非主要部の用法では、W結合と共通するものが22語、共通しないものが20語と、わずかではあるが共通するものの方が多いのに対し、主要部の用法では、W結合と共通するものは12語、共通しないものが48語と、共通しないものの方が明らかに多い。非主要部では、W結合と共通するM結合、すなわち、W結合（句）を凝縮して複合語（臨時一語）としたと想定できるM結合が多いのに対して、主要部では必ずしもW結合から作ったのではない命名的なM結合が多いと考えられるのである。

もちろん、上述の推測を確かめるためには、今回対象とした他の外来語系語基についても同様の検討を行う必要があるし、新聞というジャンルないしテキストタイプの特性、そして、なにより外来語系語基であることの特性を考慮する必要がある。また、仮に「非主要部ではM結合とW結合との間に構文的な語形成の傾向が、主要部では命名的な語形成の傾向が観察される」ということが正しいとしても、その要因の究明については、さらに質的な検討が必要になる。本稿は、そうした質的検討の前提として、「M結合・W結合」という、形態素の結合用法と自立用法との関係を把握するための新たな方法論を導入することにより、外来語系語基のM結合・W結合の量的な関係を把握しようとしたものである。

なお、今回の調査では、「ネット」「ホーム」「コン」「スーパー」「ケース」など、多義語ないし同音語の判別を行っていないため、意味の違いによるM結合とW結合の差異についてはあまり触れることができなかった。たとえば、「ケース」の場合、主要部におけるM結合「テスト-、基本-、改革-、該当-、新成長経済移行-、時効-、成長制約-、得点-、別居-」などは構文的複合語で、ほとんどが抽象名詞の「ケース」である。一方、具体名詞には「スーツ-、ショー-、ガラス-、衣装-、プラスチック-、レター-」など命名的複合語が多く、意味によってM結合とW結合の傾向が異なることが予想される。この問題については、今後の課題としたい。

文献

- 石井正彦（1992）「造語力をはかるために」『日本語学』 11-5:16-24
———（2007）『現代日本語の複合語形成論』 ひつじ書房
- 影山太郎（1993）『文法と語形成』 ひつじ書房
- 国立国語研究所（1972）『電子計算機による新聞の語彙調査Ⅲ』 秀英出版
- 斎賀秀夫（1957）「語構成の特質」『講座現代国語学Ⅱ ことばの体系』 筑摩書房（斎藤倫明・石井正彦編『日本語研究資料集 1-13 語構成』 ひつじ書房，1997 に再録） 24-45
- 齋藤俊雄・中村純作・赤野一郎編（1998）『英語コーパス言語学—基礎と実践—』 研究社出版
- 斎藤倫明（2004）『語彙論的語構成論』 ひつじ書房
- 林 四郎（1982）「臨時一語の構造」『国語学』 131:15-26（林四郎『漢字・語彙・文章の研究へ』，明治書院，1987 に再録）
- 林 四郎（1991）「日本語造語要素結合度の3段階性—その社会言語学的意味」『明海大学外国語学部論集』 3:1-13
- 宮島達夫・水谷静夫（1958）「語イの共通度について」『計量国語学』 7:40-42
- 森岡健二（1985）「外来語の派生語彙」『日本語学』 4-9:43-53
- 山下喜代（1995）「形態素と造語成分」『日本語学』 14-5:54-63

（2013 年 11 月 6 日受付，2014 年 1 月 31 日再受付）

Paper

Bound-form Usage (Word Formation) and Free-form Usage (Phrase Formation) of Morphemes:

The Quantitative Relation between Loanword Stems

KIM Juyoung (Momoyama Gakuin University, part-time instructor)

Abstract:

In the present paper bound usage of morphemes was regarded as morpheme-based word-formation and free usage of morphemes – as morpheme-based phrase formation. The quantitative relation between these 2 usages, denoted as “M-combination” and “W-combination” respectively, was measured by number of types of morphemes to which stems are attached. This analysis was carried out by choosing 50 high frequency loanword stems, found in “Mainichi Newspaper” Corpus in all articles published in 2006. Number of Types of M-combinations and W-combinations that these stems form with Sino-Japanese and loanword stems, were counted. Then the relation between 2 usages of chosen stems, divided into 2 groups (head usage and non-head usage), was analyzed with 3 parameters: correlation coefficient, similarity of co-occurring stems, difference coefficient of co-occurring stems.

Keywords: M-combination, W-combination, combination coefficient, correlation coefficient, lexical similarity, difference coefficient