

辞書見出し語の9,10文字漢字熟語を対象とした語基構成の解析

梅木 定博, 後藤 智範
神奈川大学 理学部 情報科学科

日本語のテキストにおいて、主要な概念・テーマは漢字熟語または漢字熟語を含む名詞句に表現されることが多い。特に数文字以上の漢字熟語は、より短い漢字熟語、すなわち語基(word base)から構成され、統語的、意味的構造を有している。大規模な漢字熟語集合について、これらの構造を分析することは、漢字熟語の造語特性・形態素解析・関連語の選定など様々な自然言語解析およびオントロジーの開発に有用な基礎データを提供するものと考えられる。

本研究は、一般辞書および専門用語辞書の見出し語から9および10文字の漢字熟語を対象に、構成語基の観点から品詞列パターンおよび構成語基の係り受けパターンについて調査・分析した。

Analysis to Nine or Kanji Compound Words in Entry Terms of Dictionaries

Sadahiro Umeki, Tomonori Gotoh

Department of Information Science, Kanagawa University

Kanji compound words are generally used to explain key concepts or themes in Japanese texts. Especially long kanji compound words have these characteristics in academic papers or patent documents. Long kanji compound word, which has five letters more consists of short word bases and has syntactic and semantic structures. It should be much beneficial to study a large set of long kanji words based on word base sequence patterns.

Our research examines the patterns of the large set of kanji compound words with nine or ten letters which are contained in entry terms of the various kinds of dictionaries. This paper reports the occurrences of kanji compound words and the number of parts of speech sequence per a word base sequence pattern.

1. はじめに

日本語のテキストにおいて、主要な概念・テーマは漢字熟語または漢字熟語を含む名詞句に表現される。特に数文字以上の漢字熟語は、複雑な構造を有し、その構造を分析することは高精度の形態素解析、関連語の選定、未知語の推定などに有効と考えられる。

漢字熟語を対象とした複合語の構造解析の研究は、野村の研究[1][2]を嚆矢とし、大規模な医学用語データベースを対象とした調査[3][4]、語彙概念構造に基づく解析の研究[5][6]などがある。

本研究では、コーパスとしては大規模となる総計 10^4 のオーダーの漢字熟語集合を対象とし、漢字熟語を表層的な特性、具体的には、その長さ(文字数)、構成語基の長さ/構成語基数の観点から、構成語基毎に下記の項目を明らかにしようとするものである。

- (1) 出現熟語数
- (2) 構成語基パターン毎の品詞列パターン
- (3) 構成語基パターン毎の係り受け

5文字から8文字熟語の上記項目の調査分析結果については既に報告した[7][8][9][10]。本研究では、9および10文字漢字熟語を対象とし上記項目の解析結果について報告する。

2. コーパスおよび解析手順

コーパスおよび対象漢字熟語群の特性は、上述の5~8文字漢字熟語群と同様であり、詳細は省略する。本研究の対象漢字熟語集合は、9文字熟語1277、10文字熟語593であったが、現在使用されていないと推定される語を除外し、それぞれ666、283となった[11]。

解析手順、使用された品詞体系についても5~8文字熟語集合のときと同一であり省略する。

3. 結果

対象熟語は 9 文字、10 文字熟語共に 4、5、6、7 語基のいずれかに分割された。

3.1 語基構成パターン

9 文字および 10 文字漢字熟語の構成語基数毎の出現頻度をそれぞれ表 3.1a、表 3.1b に示す。

表 3.1a 構成語基数毎の熟語数(9 文字熟語)

語基数	4	5	6	7
出現数	10	385	263	8
比率	1.50	57.81	39.49	1.20

9 文字熟語の語基数は、上の表から 5 語基が最も多く、次いで 6 語基で、両者を合わせると 97% に達し他は全体の 3% にも満たないことがわかる。

表 3.1b 構成語基数毎の熟語数(10 文字熟語)

語基数	4	5	6	7
出現数	1	52	195	35
比率	1	18	69	12

10 文字熟語の語基数は、上の 2 つの表から 6 語基が最も多く、次いで 5 語基で、両者を合わせると 87% に達し、4 語基は 1 事例しかなかったことがわかる。

表 3.2a および表 3.2b に 9、10 文字の可能な語基構成パターンと実際に出現したパターンを対比した数値をそれぞれ示す。

表 3.2a 構成語基数毎の構成パターン数(9 文字)

語基数	可能なパターン数	出現パターン数	比率	語基数	可能なパターン数	出現パターン数	比率
1	1	0	0	6	56	17	30
2	8	0	0	7	28	5	18
3	28	0	0	8	8	0	0
4	56	3	5.4	9	1	0	0
5	70	13	19				
					256	38	15

表 3.2a から 9 文字熟語では可能なパターンのうち 15% しか出現しなかったことがわかる。5、6 語基の比率が高いのは、表 3.1a から漢字熟語が多いことと一致する。

同様に、表 3.2b から 10 文字熟語では可能なパターンのうちわずか約 8% しか出現しなかったことがわかる。この値は 9 文字の 1/2 であるが、10 文字漢字熟語の総数 283 が 9 文字の総数 666 の半数に近いという事実からは妥当と言えるであろう。

表 3.2b 構成語基数毎の構成パターン数(10 文字)

語基数	可能なパターン数	出現パターン数	比率	語基数	可能なパターン数	出現パターン数	比率
1	1	0	0	6	126	19	15
2	9	0	0	7	84	13	15
3	36	0	0	8	36	0	0
4	84	1	1.2	9	9	0	0
5	126	7	5.6	10	1	0	0
					512	40	7.8

3.2 品詞列パターン

表 3.2 に挙げた構成語基数毎に出現した品詞列(並び)パターンの数を表 3.3a および表 3.3b に挙げる。

表 3.3a 語基数毎の品詞列パターン数(9 文字)

語基数	4	5	6	7
品詞列数	4	126	123	7
比率	1.54	48.46	47.31	2.69
出現数/品詞列数	2.50	3.06	2.14	1.14

表 3.1a および表 3.3a から、7 語基を除き、平均して同一の品詞列パターンの熟語が 2 以上あることがわかる。

表 3.3b 語基数毎の品詞列パターン数(10 文字)

語基数	4	5	6	7
品詞列数	1	28	115	33
比率	0.56	15.82	64.97	18.64
出現数/品詞列数	1.00	1.86	1.70	1.06

表 3.1b および表 3.3b から、いずれの語基数でもほぼ漢字熟語毎に品詞列が異なることを示している。前章で挙げたように 10 文字漢字熟語の調査対象漢字は 283 と非常に少なく、一方表 3.2b に示されるように、可能な分割パターンの組み合わせは多いためこのような結果になったと推測される。

表 3.4a 構成語基数の先頭品詞毎の品詞列パターン数(9 文字)

	名詞	サ変	形動	動詞	形容	接頭	数詞	計
4語基	3	1	0	0	0	0	0	4
	75	25	0	0	0	0	0	
5語基	60	34	18	0	6	6	1	125
	48	27.2	14.4	0	4.8	4.8	0.8	
6語基	64	20	3	2	6	23	5	123
	52	16.3	2.44	1.63	4.88	18.7	4.07	
7語基	2	1	0	0	0	3	1	7
	28.6	14.3	0	0	0	42.9	14.3	
計	129	56	21	2	12	32	7	259
	49.8	21.6	8.11	0.77	4.63	12.4	2.7	

各語基で上段は出現数、下段は出現比率を示す。

表 3.4a および表 3.4b は、構成語基数単位に品

詞列を整列した結果の一端を表している。この表から、語基数(行)および先頭品詞(列)の観点からその特性を明確にするために、それぞれの比率を算出した。

表 3.4a から 9 文字熟語では、先頭の語基が動詞、数詞となる熟語は非常に少ないことがわかる。

表 3.4b 構成語基数の先頭品詞毎の品詞列
パターン数(10 文字)

	名詞	サ変	形動	動詞	形容	接頭	数詞	計
4語基	1	0	0	0	0	0	0	1
	100	0	0	0	0	0	0	
5語基	13	7	6	1	1	0	0	28
	46.4	25	21.4	3.57	3.57	0	0	
6語基	69	21	10	1	2	8	4	115
	60	18.3	8.7	0.87	1.74	6.96	3.48	
7語基	14	10	1	1	1	5	1	33
	42.4	30.3	3.03	3.03	3.03	15.2	3.03	
計	97	38	17	3	4	13	5	177
	54.8	21.5	9.6	1.69	2.26	7.34	2.82	

表 3.4b から 10 文字熟語も同様に先頭の語基が動詞、数詞となる熟語は非常に少ないことがわかる。

3.3 係り受けパターン

本節では、構成語基数毎の係り受けパターンと品詞列パターンとの出現傾向についての結果を述べる。

表 3.5a および表 3.5b に 9、10 文字の語基数毎の出現した係り受けパターンの種類を挙げる。

表 3.5a 係り受けパターン比率(9 文字)

語基	A	B	C	D	E	F	G	H	I	J
4	3	1	1							
(3)	60	20	20							
5	52	34	18	25	1	3	2	38	1	3
(13)	26	17	8.9	12	0.5	1.5	1	19	0.5	1.5
6	16	9	6	14	21	7	2	2	23	5
(15)	13	7.1	4.7	11	17	5.5	1.6	1.6	18	3.9
7	1	1	1							
(3)	33	33	33							

	K	L	M	N	O
5	5	16	5		
	2.5	7.9	2.5		
6	7	2	1	10	2
	5.5	1.6	0.8	7.9	1.6

・語基(列)のカッコ内の数字は、語基毎に出現した係り受けパターンの種類を示している。

・各語基毎に、上段は当該係り受けパターンの出現数(漢字熟語数)、下段はその総数に対する比率を示す。

・表 3.5b についても同様である。

表 3.5a から 9 文字熟語について、5 語基については、13 種類の係り受けパターンのうち、上位 4 種

のパターン A、B、D、H で約 74% に達することがわかる。一方、6 語基については、そのような偏差が無いことがわかる。

図 3.1 および図 3.2 にこれらの係り受けパターンを挙げる。

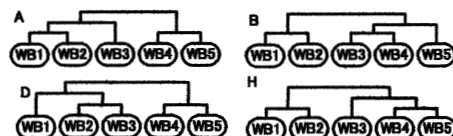


図 3.1 5 語基の係り受けパターンの実列
(9 文字熟語、高出現品詞列)

・WB_i は語基(出現順)を表している。以下同様。

表 3.5b 係り受けパターン比率(10 文字)

語基	A	B	C	D	E	F	G	H	I	J
5	7	8	3	6	1	1	6	1	1	1
(10)	20	23	8.6	17	2.9	2.9	17	2.9	2.9	2.9
6	7	8	6	20	1	14	4	1	1	4
(25)	5.6	6.5	4.8	16	0.8	11	3.2	0.8	0.8	3.2
7	2	3	5	3	2	4	1	2		
(8)	9.1	14	23	14	9.1	18	4.5	9.1		

	K	L	M	N	O	P	Q	R	S	T
6	27	1	1	5	3	1	8	1	2	3
	22	0.8	0.8	4	2.4	0.8	6.5	0.8	1.6	2.4

	U	V	W	X	Y
6	1	2	1	1	1
	0.8	1.6	0.8	0.8	0.8

表 3.5b から 10 文字熟語について 5 語基については、10 種類の係り受けパターンのうち、上位 4 種のパターン A、B、D、G で 77% 以上に達することがわかる。一方、6 語基については、25 種類の係り受けパターンのうち、上位 3 種のパターン D、F、K で 49% に達することがわかる。

図 3.2 および図 3.3 に、それぞれ上述の 5 語基、6 語基の係り受けパターンを示す。

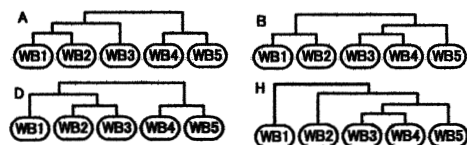


図 3.2 5 語基の係り受けパターンの実列
(10 文字熟語、高出現品詞列)

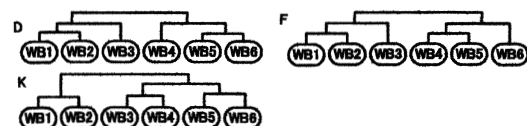


図 3.3 6 語基の係り受けパターンの実列

4. 考察

4.1 語基構成パターン

表3.2aおよび表3.2bに9,10文字熟語についてそれぞれの語基数毎の熟語数を挙げた。表4.1aおよび表4.1bにより小さい粒度での特性を明らかにするため、語基構成パターン毎の熟語数と品詞列パターン数を示す。

表 4.1a 語基数毎の語基構成パターン(9 文字)

語基構成パターン	出現数	品詞	語基構成パターン	出現数	品詞列数
4			1-2-1-1-2-2	4	4
2-1-4-2	1	1	1-2-1-1-3-1	1	1
2-3-2-2	5	4	1-2-1-2-1-2	5	5
2-4-2-1	4	2	1-2-1-2-2-1	38	30
5			1-2-2-1-1-2	6	5
1-1-4-2-1	1	1	1-2-2-1-2-1	9	8
1-2-2-2-2	31	18	1-2-2-2-1-1	1	1
1-2-3-2-1	1	1	1-2-1-1-1-2-2	3	3
1-4-1-2-1	1	1	1-2-1-1-2-1-2	10	8
2-1-1-2-3	2	2	2-1-1-2-2-1	47	25
2-1-2-1-3	1	1	1-2-1-2-1-1-2	8	7
2-1-2-2-2	92	38	2-1-2-1-2-1	93	33
2-1-2-3-1	1	1	1-2-1-2-2-1-1	3	3
2-1-3-1-2	1	1	1-2-2-1-1-2-1	24	11
2-1-3-2-1	1	1	1-2-2-1-2-1-1	7	5
2-2-1-2-2	89	34	7		
2-2-2-1-2	41	25	1-1-1-1-2-2-1	2	2
2-2-2-2-1	122	38	1-1-1-2-1-1-2	1	1
2-3-1-2-1	1	1	1-1-2-1-1-2-1	1	1
6			1-2-1-1-1-2-1	1	1
1-1-2-1-2-2	1	1	2-1-1-1-1-2-1	3	2
1-1-2-2-2-1	3	3			

5語基では、特に下記の5種類のパターンが多く出現し、表3.1から実際に出現したパターンの38%(5/13)、可能なパターンの7%(5/70)が、5語基の漢字熟語の97%(375/385)になっていることを示している。

1-2-2-2-2、 2-1-2-2-2、 2-2-1-2-2、
2-2-2-1-2、 2-2-2-2-1

さらにこれらのパターンでは、平均して2個以上の熟語が同一の品詞列となることを示している。

6語基については、特に下記の4種類のパターンが多く出現し、表3.1から実際に出現したパターンの24%(4/17)、可能なパターンの7%(4/56)が、6語基の漢字熟語の77%(202/263)に達することを示している。

2-1-2-1-2-1、 2-1-1-2-2-1
1-2-1-2-2-1、 2-2-1-1-2-1

これらのパターンのうち、平均して2個以上の熟語が同一の品詞列となるのは、上位2つのパターンのみである。

表 4.1b 語基数毎の語基構成パターン(10 文字)

語基構成パターン	出現数	品詞	語基構成パターン	出現数	品詞列数
4			2-1-2-2-2-1	46	26
1-4-2-3	1	1	2-1-3-1-2-1	2	1
5			2-2-1-1-2-2	8	6
1-2-2-2-3	1	1	2-2-1-2-1-2	11	7
2-1-2-1-4	1	1	2-2-1-2-2-1	41	25
2-1-2-2-3	1	1	2-2-2-1-1-2	3	3
2-1-4-2-1	2	1	2-2-2-1-2-1	15	7
2-2-2-2-2	45	23	3-1-2-1-2-1	1	1
2-2-2-3-1	1	1	7		
2-2-3-2-1	1	1	1-1-2-1-2-2-1	2	2
6			1-1-2-2-1-2-1	1	1
1-1-2-2-2-2	2	2	1-2-1-1-2-2-1	1	1
1-2-1-2-1-3	1	1	1-2-1-2-1-1-2	1	1
1-2-1-2-2-2	4	4	1-2-1-2-1-2-1	4	4
1-2-1-2-3-1	1	1	2-1-1-1-2-2-1	1	1
1-2-2-1-2-2	4	4	2-1-1-2-1-1-2	2	2
1-2-2-2-1-2	4	3	2-1-1-2-1-2-1	9	9
1-2-2-2-2-1	7	5	2-1-2-1-1-1-2	1	1
2-1-1-2-2-2	8	7	2-1-2-1-1-2-1	8	7
2-1-1-3-2-1	1	1	2-1-2-1-2-1-1	2	2
2-1-2-1-2-2	24	19	2-1-2-2-1-1-1	2	2
2-1-2-2-1-2	12	9	2-2-2-1-1-1-1	1	1

5語基では、2-2-2-2-2だけが極端に多く表3.1aから、5語基の漢字熟語の88%(45/52)に達し、約2個の熟語が同一の品詞列となることを示している。

6語基については、特に下記の4種類のパターンが多く出現し、表3.1から実際に出現したパターンの24%(4/17)、可能なパターンの7%(4/56)が、6語基の漢字熟語の65%(126/195)を占める。

2-1-2-2-2-1、 2-2-1-2-2-1
2-1-2-1-2-2、 2-2-2-1-2-1

語基構成について、特定の語基構成パターンだけが極端に多く出現するという特徴は、6~8文字漢字熟語のそれと類似の傾向を示している[8][9][10]。

4.2 品詞列パターン

表 3.4a、3.4b を単なる語基数から、語基構成パターン単位に粒度を下げた、これを表 4.2a、4.2b に示す。

9文字漢字について、4.1節で明らかにしたように特定の語基パターンが顕著に出現するという特性を踏まえ、5、6語基について、以下の語基構成パターンに着目する。

5語基: 1-2-2-2-2、 2-1-2-2-2、 2-2-1-2-2、

2-2-2-1-2、 2-2-2-2-1
 6語基: 2-1-2-1-2-1、 2-1-1-2-2-1
 1-2-1-2-2-1、 2-2-1-1-2-1

表 4.2a 構成語基パターン毎の品詞列パターン数
 (9文字)

	名詞	サ変	形動	動詞	形容	接頭	数詞
4語基							
2142	1	0	0	0	0	0	0
2322	3	1	0	0	0	0	0
2421	2	0	0	0	0	0	0
5語基							
11421	1	0	0	0	0	0	0
12222	5	0	0	0	6	6	1
14121	1	0	0	0	0	0	0
21123	1	1	0	0	0	0	0
21213	1	0	0	0	0	0	0
21222	17	13	8	0	0	0	0
21231	1	0	0	0	0	0	0
21312	1	0	0	0	0	0	0
21321	1	0	0	0	0	0	0
22122	19	11	4	0	0	0	0
22212	16	5	4	0	0	0	0
22221	21	12	5	0	0	0	0
23121	1	0	0	0	0	0	0
6語基							
112122	0	0	0	0	0	0	1
112221	2	0	0	0	0	1	0
121122	1	0	0	0	0	2	1
121131	0	0	0	0	0	1	0
121212	1	0	0	0	0	3	1
121221	9	0	0	1	4	14	2
122112	1	0	0	0	2	2	0
122121	3	0	0	1	2	2	0
122211	0	0	0	0	0	1	0
211122	1	2	0	0	0	0	0
211212	7	1	0	0	0	0	0
211221	15	9	1	0	0	0	0
212112	6	1	0	0	0	0	0
212121	24	8	1	0	0	0	0
212211	3	0	0	0	0	0	0
221121	10	1	0	0	0	0	0
221211	4	0	1	0	0	0	0
7語基							
1111221	1	0	0	0	0	0	1
1112112	0	0	0	0	0	1	0
1121121	0	0	0	0	0	1	0
1211121	0	0	0	0	0	1	0
2111121	1	1	0	0	0	0	0

上の表から、5語基では1-2-2-2-2以外は先頭語基の品詞は名詞、サ変名詞、形容動詞で90%以上を占める。一方、6語基では1-2-1-2-2-1以外は、先頭語基は名詞、サ変名詞だけが顕著に多く、形容動詞が僅かにあることがわかる。また、9文字熟語全体で、先頭語基の品詞が接頭辞、数詞にな

ることは非常に稀であることが判明する。

表 4.2b 構成語基パターン毎の品詞列パターン数
 (10文字)

	名詞	サ変	形動	動詞	形容	接頭	数詞
4語基							
1423	1	0	0	0	0	0	0
5語基							
12223	0	0	0	0	1	0	0
21214	1	0	0	0	0	0	0
21223	1	0	0	0	0	0	0
21421	1	0	0	0	0	0	0
22222	10	6	10	1	0	0	0
22231	1	0	0	0	0	0	0
22321	0	1	0	0	0	0	0
6語基							
112222	0	0	0	0	0	1	1
121213	0	0	0	0	0	1	0
121222	0	0	0	1	0	2	1
121231	0	0	0	0	1	0	0
122122	1	0	0	0	0	2	1
122212	1	0	0	0	0	2	0
122221	2	0	0	0	1	1	1
211222	5	1	1	0	0	0	0
211321	0	0	1	0	0	0	0
212122	13	5	1	0	0	0	0
212212	6	2	1	0	0	0	0
212221	17	6	3	0	0	0	0
213121	1	0	0	0	0	0	0
221122	5	1	0	0	0	0	0
221212	5	1	1	0	0	0	0
221221	18	5	2	0	0	0	0
222112	3	0	0	0	0	0	0
222121	5	1	1	0	0	0	0
312121	1	0	0	0	0	0	0
7語基							
1121221	1	0	0	0	0	1	0
1122121	0	0	0	0	0	1	0
1211221	0	0	0	0	0	1	0
1212112	0	0	0	0	0	1	0
1212121	0	0	0	1	1	1	1
2111221	1	0	0	0	0	0	0
2112112	2	0	0	0	0	0	0
2112121	4	4	1	0	0	0	0
2121112	0	1	0	0	0	0	0
2121121	3	4	0	0	0	0	0
2122111	2	1	0	0	0	0	0
2122111	1	0	0	0	0	0	0
2221111	1	0	0	0	0	0	0

10文字熟語についても9文字熟語と同様の視点で特性を捉えてみる。表 4.2b から、2-2-2-2-2以外は、9文字と同様に先頭語基は名詞、サ変名詞、形容動詞で95%を以上占める。10文字熟語

全体で、先頭語基の品詞が接頭辞、数詞、形容詞になることは非常に稀であることが判明する。

先頭語基の品詞として接頭辞、数詞、形容詞が極端に少ないことは9、10文字熟語に固有の現象であり、6～8文字熟語ではこのような現象は見られなかった[8][9][10]。

4.3 係り受けパターンの構造

係り受け構造は、木構造的に捉えることができる。図3.1から図3.3に挙げた事例に見られるように、語基数が増えるとレベル(高さ)が高くなる傾向になる。

3.3節に挙げた係り受けパターンとそのレベルに着目してみたい。表4.3aおよび表4.3bは、それぞれ9文字、10文字熟語について、表3.5aおよび表3.5bに挙げた係り受けパターンのレベルを示したものである。

表4.3a 係り受けパターンとそのレベル(9文字)

レベル\語基数	2	3	4	5	6
4	1	2	×	×	×
5	-	6	7	×	×
6	-	6	8	1	×
7	-	1	2	-	-

・×は論理的にあり得ない、-は出現しなかったことを示している。以下同様。

図3.1に9文字熟語の5語基の最多出現の係り受けパターンの構造を示したが、いずれもレベルは3である。

表4.3b 係り受けパターンとそのレベル (10文字)

レベル\語基数	2	3	4	5	6
5	-	6	4	×	×
6	1	6	15	3	×
7	-	2	5	1	-

図3.2および図3.3に10文字熟語の5語基および6語基のそれぞれ最多出現の係り受けパターンの構造を示した。5語基では、Hを除きレベルは3、6語基では、3種類と共にレベルは3である。

図3.1から図3.3の例からもわかるように、2つの語基間の係り受けが多いとレベルは抑えられ、一方少ないと図3.2のパターンHのように高くなる。上の2つ表から、最多の係り受けパターンの多くは前者であったことがわかる。このことは、語長が長く語基数が多い場合には、2語基を1つの構

成単位とすることにより熟語全体の構造が捉え易くなると考えられる。

5. 終わりに

6～7文字熟語に対しても、前節で言及した係りあう2語基を1つの構成単位として、その構造を再調査することもその造語特性を理解する上で意義があると考えられる。

謝辞: 本研究を進めるにあたり語基分割と係り受け作業に協力していただいた情報科学科4年生の樽林正浩さんに感謝します。

註・参考文献

- [1] 野村雅昭. 三字漢字の構造. 秀英出版. 国立国語研究所報告. No.51. pp.37-62(1973).
- [2] 野村雅昭. 四字漢字の構造. 秀英出版. 国立国語研究所報告. No.54. pp.36-80(1974).
- [3] 小山照夫, 大江和彦. 医学専門用語の構造解析. 学術情報センター紀要. No.6, pp.115-124(1994).
- [4] 小山照夫, 大江和彦. 日本語医学専門用語の構造解析. 情報知識学会第2回研究報告会講演論文集. pp.17-20(1994).
- [5] 竹内孔一, 内山清子, 吉岡真治, 影浦峯, 小山照夫. 語彙の制約を考慮した複合語解析モデルの構築. 情報処理学会. 情報学基礎研究会報告. No.57, pp.71-78(2000).
- [6] 竹内孔一, 内山清子, 吉岡真治, 影浦峯, 小山照夫. 語彙概念構造を利用した複合名詞内の係り関係の解析. 情報処理学会論文誌. Vol.43, No.5, pp.1446-1456(2002).
- [7] 郭恩東, 森本貴之, 後藤智範. 辞書見出し語の5文字漢字熟語を対象とした語基構成の解析. 言語処理学会第13回年次大会. pp.348-351(2007).
- [8] 梅木定博, 森本貴之, 後藤智範. 辞書見出し語の6文字漢字熟語を対象とした語基構成の解析. 情報処理学会. NL no.184, pp.13-18(2008).
- [9] 梅木定博, 後藤智範. 辞書見出し語の7文字漢字熟語を対象とした語基構成の解析. 情報処理学会. NL no.184, pp.113-118(2008).
- [10] 梅木定博, 後藤智範. 辞書見出し語の8文字漢字熟語を対象とした語基構成の解析. 情報処理学会. NL no.187, pp.1-6(2008).
- [11] 意味がわからない熟語については、全てGoogleを使用して検索を行った(引用符で囲まれた語そのものを探索する方式を採用)。その結果、9、10文字では半数が見つからない結果となった。以前の研究5～8文字についても同様の処理をしたが僅かであった。