

## 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

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8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 Eight-Kanji Compound Words in Entry Terms of Dictionaries

Sadahiro Umeki, Tomonori Gotoh

Department of Information Science, Kanagawa University

Kanji compound words or noun phrase consisted in them intend 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 have syntactically and semantically structures. It should be much beneficial to study to 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 eight 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な漢字熟語集合(10<sup>4</sup>)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7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]。本研究では、8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5~7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、すなわち同一の特性をもつ漢字熟語の総数は3132であった。

解析手順、使用された品詞体系についても5~7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3、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

漢字熟語の構成語基数毎の出現頻度を表 3.1 に示す。

| 構成基数  | 3    | 4     | 5     | 6    | 7    |
|-------|------|-------|-------|------|------|
| 熟語数   | 8    | 1245  | 1797  | 81   | 1    |
| 比率(%) | 0.26 | 39.75 | 57.38 | 2.59 | 0.03 |

上の表から5語基が最も多く、次いで4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 3.2 構成語基数毎の構成パターン数

| 分割パターン       | 種類 | 5語基(35)      | 6語基(21) | 7語基(1) |
|--------------|----|--------------|---------|--------|
| 3語基(21)      |    |              |         |        |
| 1[2]6[1]     | 3  | 1[4]4[1]     | 5       |        |
| 1[1]2[1]5[1] | 6  | 1[3]2[1]3[1] | 20      |        |
| 1[1]3[1]4[1] | 6  | 1[2]2[3]     | 10      |        |
| 2[2]4[1]     | 3  | 6語基(21)      |         |        |
| 2[1]3[2]     | 3  | 1[5]3[1]     | 6       |        |
| 4語基(35)      |    | 1[4]2[2]     | 15      |        |
| 1[3]5[1]     | 4  | 7語基(1)       |         |        |
| 1[2]2[1]4[1] | 12 | 1[6]2[1]     | 1       |        |
| 1[2]3[2]     | 6  | 8語基(1)       |         |        |
| 1[1]2[2]3[1] | 12 | 1[8]         | 1       |        |
| 2[4]         | 1  |              |         |        |

表 3.2 は、8 字漢字熟語の構成語基の全ての組み合わせを表している。実際に出現したのは約 100 パターンで、詳細は表 4.1 に示す。

### 3.2 品詞列パターンの分析

表 3.2 に挙げた構成語基数毎に出現した品詞列(並び)パターン数を表 3.3 に挙げる。

| 構成基数   | 3    | 4     | 5     | 6     | 7    |
|--------|------|-------|-------|-------|------|
| パターン数  | 3    | 85    | 372   | 69    | 1    |
| 比率(%)  | 0.57 | 15.88 | 70.32 | 13.04 | 0.19 |
| 出頻/品詞列 | 2.67 | 14.65 | 4.83  | 1.17  | 1.00 |

表 3.3 から、5 語基が品詞列パターンで最多となった。また、表 3.1 と見比べると、4 語基では 85 の品詞列パターンで 1245 語の熟語に対応できるのに対して、6~7 語基では、熟語毎に異なった品詞列となっていることがわかる。

表 3.4 は、構成語基数単位に品詞列を整理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 3.4 構成語基数の先頭品詞毎の品詞列パターン数

| 語基数 | サ変  | 形動 | 形容 | 接頭 | 動詞 | 名詞  | 数詞 |
|-----|-----|----|----|----|----|-----|----|
| 3語基 | 0   | 0  | 0  | 0  | 0  | 3   | 0  |
| 4語基 | 22  | 17 | 1  | 0  | 3  | 42  | 0  |
| 5語基 | 84  | 45 | 27 | 50 | 8  | 139 | 19 |
| 6語基 | 5   | 2  | 3  | 16 | 3  | 25  | 15 |
| 7語基 | 0   | 0  | 0  | 0  | 0  | 0   | 1  |
| 計   | 106 | 64 | 31 | 66 | 14 | 209 | 35 |

表 3.5 構成語基の先頭品詞毎の品詞列パターン数比率

| 語基数 | サ変   | 形動   | 形容  | 接頭   | 動詞  | 名詞   | 数詞   |
|-----|------|------|-----|------|-----|------|------|
| 3語基 | 0    | 0    | 0   | 0    | 0   | 100  | 0    |
| 4語基 | 25.9 | 20   | 1.2 | 0    | 3.5 | 49.4 | 0    |
| 5語基 | 22.6 | 12.1 | 7.3 | 13.4 | 2.2 | 37.4 | 5.1  |
| 6語基 | 7.2  | 2.9  | 4.3 | 23.2 | 4.3 | 36.2 | 21.7 |
| 7語基 | 0    | 0    | 0   | 0    | 0   | 0    | 100  |

表 3.5 は語基数単位で、特定品詞で開始する品詞列パターンの全体の比率を表している。この表および表 3.4 に基づいて、以下に語基数毎にその特性について記述する。

- (1) **3 語基:** 殆当該の全ての熟語が名詞で始まることわかる。
- (2) **4 語基:** 約半数が名詞、形動、サ変も 20% 以上の比率で先頭品詞となっていることがわかる。また、数詞、接頭を先頭に持つ語は出現しなかった。
- (3) **5 語基:** 名詞とサ変名詞の比率が大きな事はわかるが、この表の中では全ての品詞を先頭品詞に持っていることがわかる。
- (4) **6 語基:** 名詞、接頭、数詞が先頭品詞となる比率は 80% を超えていることがわかる。同時に 6 語基で形動、サ変の比率が大きく低下していることがわかる。
- (5) **7 語基:** 出現頻度が 1 語だけであるため特徴について言及することはできない。

以上から、先頭語基の品詞について、語基数の増加により以下の全般的傾向が見られることがわかる。

- 減少: 名詞、サ変、形容動詞
- 増加: 数詞、接頭辞

### 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3.6に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 3.6 係り受けパターン比率

| 語基数 | A    | B    | C    | D    | E    | F    | G    | H    | I    | J    |
|-----|------|------|------|------|------|------|------|------|------|------|
| 3   | 2    |      |      |      |      |      |      |      |      |      |
|     | 100  |      |      |      |      |      |      |      |      |      |
| 4   | 56   | 36   | 37   | 35   | 44   | 8    | 1    |      |      |      |
|     | 25.8 | 16.6 | 17.1 | 16.1 | 20.3 | 3.69 | 0.46 |      |      |      |
| 5   | 106  | 94   | 25   | 66   | 3    | 3    | 2    | 78   | 11   | 5    |
|     | 23   | 20.4 | 5.42 | 14.3 | 0.65 | 0.65 | 0.43 | 16.9 | 2.39 | 1.08 |
| 6   | 8    | 12   | 3    | 7    | 1    | 27   | 5    | 5    | 9    | 7    |
|     | 6.45 | 9.68 | 2.42 | 5.65 | 0.81 | 21.8 | 4.03 | 4.03 | 7.26 | 5.65 |
| 7   | 1    | 1    |      |      |      |      |      |      |      |      |
|     | 50   | 50   |      |      |      |      |      |      |      |      |

| 語基数 | K    | L    | M    | N    | O    | P    | Q    | R    | S    | T    |
|-----|------|------|------|------|------|------|------|------|------|------|
| 5   | 6    | 15   | 33   | 5    | 4    | 3    | 2    |      |      |      |
|     | 1.3  | 3.25 | 7.16 | 1.08 | 0.87 | 0.65 | 0.43 |      |      |      |
| 6   | 4    | 1    | 3    | 2    | 1    | 9    | 2    | 2    | 1    | 3    |
|     | 3.23 | 0.81 | 2.42 | 1.61 | 0.81 | 7.26 | 1.61 | 1.61 | 0.81 | 2.42 |

| 語基数 | U    | V    | W    | X    | Y    | Z    | 2A   | 2B   | パターン数 |
|-----|------|------|------|------|------|------|------|------|-------|
| 6   | 1    | 1    | 2    | 3    | 1    | 1    | 1    | 2    | 2     |
|     | 0.81 | 0.81 | 1.61 | 2.42 | 0.81 | 0.81 | 0.81 | 1.61 | 比率    |

6語基と7語基については係り受けの同定が一意に決定できない事例が大多数あった。このため、表3.6の6、7語基については係り受け総数は、品質列数よりも多くなっている。この現象を考慮しても、語基数により、係り受けパターン数の偏差が非常に大きくなっていることがわかる。

(1) 3語基: 図3.1に示す係り受けパターンが見られた。

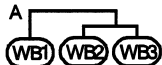


図 3.1 3語基の係り受けパターン

(WB\*は出現する順序での語基を表している)

(2) 4語基: 図3.2に示す7種類の係り受けパターンが見られた。パターンF・Gは並列構造である。

(3) 5語基: 図3.3に示す17種類の係り受けパターンが見られた。N、P、Qは並列構造であり、その構造も少ない語基数のそれと比較すると複雑化している。

(4) 6語基: 図3.4に示す29種類の係り受けパターンが見られた。ただし、表3.7より1つの係り受けパターンに対する品詞列パターン数はかなり少ないことになる。6語基そのものが全体で69語のみであったために、2語に1つの品詞列パターンが存在することになり、その傾向はほとんど見られなかった。S、W、Z、2A、2Bは並列

構造を持っている。5語基と比べても多くの並列構造を持つ係り受けのパターンが出現していることがわかる。

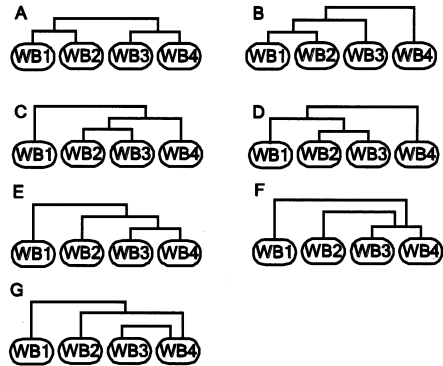


図 3.2 4語基の係り受けパターン

(5) 7語基: 図3.3に示す2種類の係り受けパターンが見られた。しかしこのパターンは1語に対して、有効解となりそうな係り受けパターンである。

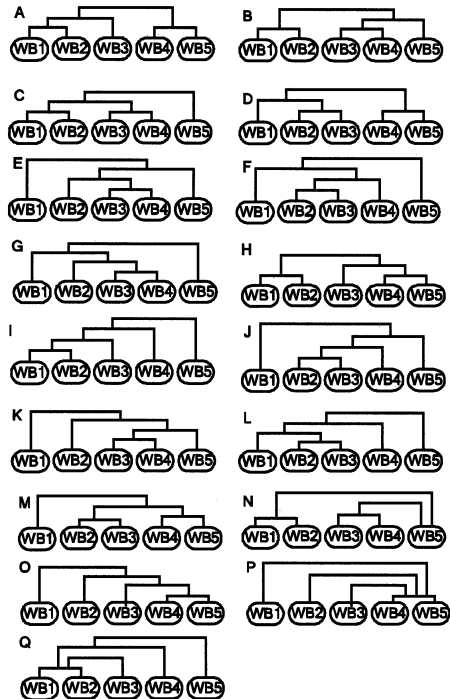


図 3.3 5語基の係り受けパターン

表3.3に構成語基数毎の品詞列パターンの数を挙げたが、より小さい粒度での特性、すなわち構成語基と品詞列の関係はこの表からは明らかではない。表4.1に、構成語基毎の熟語の出現頻度および品詞列パターンの種類を示す。

表 4.1 語基構成パターン

| 構成語基           | 熟語数         | 品詞列       | 構成語基               | 熟語数        | 品詞列       |
|----------------|-------------|-----------|--------------------|------------|-----------|
| 2-4-2          | 2           | 1         | 1-3-1-2-1          | 2          | 2         |
| 4-2-2          | 6           | 3         | 2-1-1-2-2          | 111        | 57        |
| 1-2-2-3        | 1           | 1         | 2-1-1-3-1          | 2          | 1         |
| 1-2-3-2        | 1           | 1         | 2-1-2-1-2          | 197        | 58        |
| 1-3-2-2        | 2           | 2         | <b>2-1-2-2-1</b>   | <b>589</b> | <b>88</b> |
| 1-4-2-1        | 3           | 1         | 2-2-1-1-2          | 61         | 35        |
| 2-1-2-3        | 14          | 5         | <b>2-2-1-2-1</b>   | <b>489</b> | <b>81</b> |
| 2-1-3-2        | 11          | 4         | 2-2-2-1-1          | 18         | 11        |
| 2-2-1-3        | 3           | 2         | 3-1-1-2-1          | 2          | 2         |
| <b>2-2-2-2</b> | <b>1140</b> | <b>60</b> | <b>1-1-1-1-2-2</b> | <b>2</b>   | <b>2</b>  |
| 2-2-3-1        | 8           | 7         | 1-1-1-2-1-2        | 1          | 1         |
| 2-3-1-2        | 6           | 5         | 1-1-1-2-2-1        | 7          | 7         |
| 2-3-2-1        | 26          | 9         | 1-1-2-1-1-2        | 2          | 2         |
| 3-1-2-2        | 9           | 7         | 1-1-2-1-2-1        | 7          | 7         |
| 3-2-1-2        | 4           | 3         | 1-1-2-2-1-1        | 1          | 1         |
| 3-2-2-1        | 16          | 5         | 1-2-1-1-1-2        | 2          | 2         |
| 4-1-2-1        | 1           | 1         | 1-2-1-1-2-1        | 20         | 17        |
| 1-1-2-2-2      | 22          | 17        | 1-2-1-2-1-1        | 7          | 7         |
| 1-1-2-3-1      | 1           | 1         | 2-1-1-1-1-2        | 4          | 3         |
| 1-2-1-2-2      | 106         | 49        | 2-1-1-1-2-1        | 21         | 15        |
| 1-2-1-3-1      | 2           | 2         | 2-1-1-2-1-1        | 5          | 4         |
| 1-2-2-1-2      | 51          | 34        | 2-2-1-1-1-1        | 2          | 2         |
| 1-2-2-2-1      | 144         | 54        | 1-1-1-1-1-2-1      | 1          | 1         |

この表から、8文字漢字熟語において、構成語基と漢字熟語数の非常に特徴的な傾向が見取れる。

8文字熟語では2-2-2-2パターンが最も多く、次に2-1-2-2-1パターン、2-2-1-2-1パターンが多い。2-2-2-2パターンを除くと、5語基の特定の語基構成が顕著に多いことがわかる。

全般的に語基数毎に特定の語基構成が顕著に多い、具体的にはどの語基構成が最も多く出現するかが判明した。すなわち、以下ようになった。

- 4語基: 2-2-2-2、
- 5語基: 2-1-2-2-1、2-2-1-2-1  
2-1-2-1-2、2-1-1-2-2
- 6語基: 1-2-1-1-2-1、2-1-1-1-2-1

また、2-2-2-2パターンと2-1-2-2-1および2-2-1-2-1パターンはさらに顕著な特性を示している。すなわち、語基数が同数の他のパターンと比較して、品詞列パターン数が顕著に少ないことである。

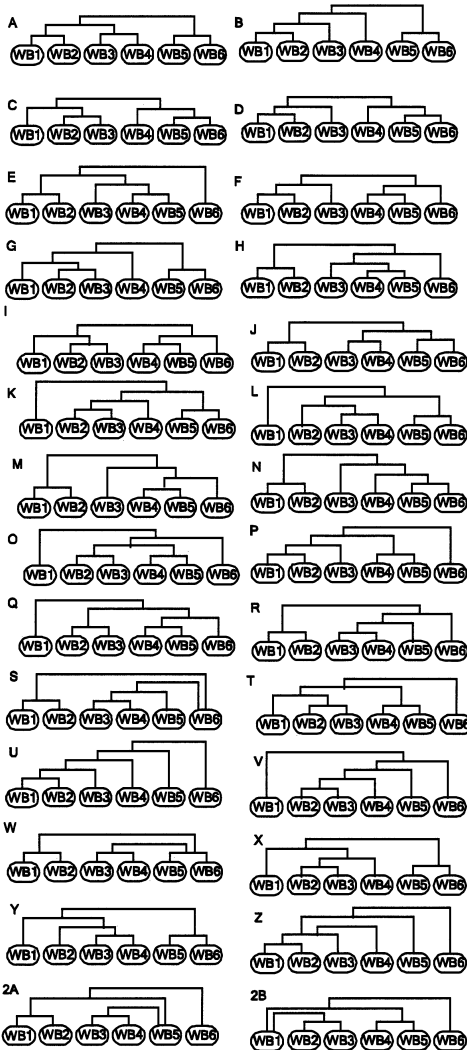


図 3.4 6語基の係り受けパターン

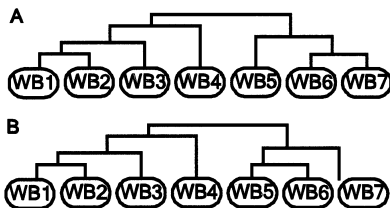


図 3.5 7語基の係り受けパターン

## 4. 考察

### 4.1 語基構成パターン

語基構成についての上述の特徴は、6文字、7文字漢字熟語のそれと類似の傾向を示している[8][9]。

#### 4.2 品詞列パターン

表 3.4 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 4.2 に示す。

表 4.2 構成語基パターン毎の品詞列パターン数

|   | 名詞                 | 動詞        | 接頭       | 数詞       | 形容       | 形動        | サ変        |
|---|--------------------|-----------|----------|----------|----------|-----------|-----------|
| 3 | 2-4-2              | 1         | 0        | 0        | 0        | 0         | 0         |
|   | 4-2-2              | 3         | 0        | 0        | 0        | 0         | 0         |
| 4 | 1-2-2-3            | 0         | 0        | 0        | 0        | 1         | 0         |
|   | 1-2-3-2            | 1         | 0        | 0        | 0        | 0         | 0         |
|   | 1-3-2-2            | 1         | 0        | 0        | 0        | 1         | 0         |
|   | 1-4-2-1            | 1         | 0        | 0        | 0        | 0         | 0         |
|   | 2-1-2-3            | 2         | 0        | 0        | 0        | 0         | 3         |
|   | 2-1-3-2            | 3         | 0        | 0        | 0        | 0         | 1         |
|   | 2-2-1-3            | 2         | 0        | 0        | 0        | 0         | 0         |
|   | <b>2-2-2-2</b>     | <b>26</b> | <b>3</b> | <b>0</b> | <b>0</b> | <b>14</b> | <b>17</b> |
|   | 2-2-3-1            | 5         | 0        | 0        | 0        | 2         | 0         |
|   | 2-3-1-2            | 2         | 0        | 0        | 0        | 2         | 1         |
|   | 2-3-2-1            | 4         | 0        | 0        | 0        | 2         | 3         |
|   | 3-1-2-2            | 6         | 0        | 0        | 0        | 0         | 2         |
|   | 3-2-1-2            | 3         | 0        | 0        | 0        | 0         | 0         |
|   | 3-2-2-1            | 4         | 0        | 0        | 0        | 0         | 1         |
|   | 4-1-2-1            | 1         | 0        | 0        | 0        | 0         | 0         |
| 5 | 1-1-2-2-2          | 1         | 0        | 5        | 11       | 0         | 0         |
|   | 1-1-2-3-1          | 0         | 0        | 0        | 1        | 0         | 0         |
|   | 1-2-1-2-2          | 12        | 3        | 23       | 1        | 10        | 0         |
|   | 1-2-1-3-1          | 1         | 0        | 1        | 0        | 0         | 0         |
|   | 1-2-2-1-2          | 8         | 2        | 13       | 0        | 11        | 0         |
|   | 1-2-2-2-1          | 14        | 5        | 18       | 6        | 10        | 1         |
|   | 1-3-1-2-1          | 1         | 0        | 0        | 1        | 0         | 0         |
|   | <b>2-1-1-2-2</b>   | <b>28</b> | <b>0</b> | <b>0</b> | <b>0</b> | <b>4</b>  | <b>25</b> |
|   | 2-1-1-3-1          | 1         | 0        | 0        | 0        | 0         | 0         |
|   | <b>2-1-2-1-2</b>   | <b>32</b> | <b>1</b> | <b>0</b> | <b>0</b> | <b>6</b>  | <b>19</b> |
|   | <b>2-1-2-2-1</b>   | <b>43</b> | <b>0</b> | <b>0</b> | <b>0</b> | <b>17</b> | <b>28</b> |
|   | 2-2-1-1-2          | 21        | 0        | 0        | 0        | 0         | 7         |
|   | <b>2-2-1-2-1</b>   | <b>40</b> | <b>0</b> | <b>0</b> | <b>0</b> | <b>18</b> | <b>23</b> |
|   | 2-2-2-1-1          | 9         | 0        | 0        | 0        | 0         | 2         |
|   | 3-1-1-2-1          | 0         | 0        | 0        | 0        | 0         | 2         |
| 6 | 1-1-1-1-2-2        | 1         | 0        | 0        | 1        | 0         | 0         |
|   | 1-1-1-2-1-2        | 0         | 0        | 0        | 1        | 0         | 0         |
|   | 1-1-1-2-2-1        | 0         | 0        | 2        | 5        | 0         | 0         |
|   | 1-1-2-1-1-2        | 1         | 0        | 0        | 1        | 0         | 0         |
|   | 1-1-2-1-2-1        | 1         | 0        | 2        | 4        | 0         | 0         |
|   | 1-1-2-2-1-1        | 0         | 0        | 0        | 1        | 0         | 0         |
|   | 1-2-1-1-1-2        | 0         | 0        | 1        | 1        | 0         | 0         |
|   | <b>1-2-1-1-2-1</b> | <b>5</b>  | <b>3</b> | <b>7</b> | <b>0</b> | <b>2</b>  | <b>0</b>  |
|   | 1-2-1-2-1-1        | 1         | 0        | 4        | 1        | 1         | 0         |
|   | 2-1-1-1-1-2        | 2         | 0        | 0        | 0        | 0         | 1         |
|   | <b>2-1-1-1-2-1</b> | <b>11</b> | <b>0</b> | <b>0</b> | <b>0</b> | <b>0</b>  | <b>2</b>  |
|   | 2-1-1-2-1-1        | 2         | 0        | 0        | 0        | 0         | 2         |
|   | 2-2-1-1-1-1        | 2         | 0        | 0        | 0        | 0         | 0         |
| 7 | 1-1-1-1-1-2-1      | 0         | 0        | 0        | 1        | 0         | 0         |

表 4.2 は、語基数毎の特性を表している表 3.2、表 3.3 の結果以上の顕著な事象はもたらさない。

前節で明らかになったように、個々の語基構成パターンで出現頻度が多いので、そうでない語基構成パターンの漢字熟語については、表 4.2 の結果について言及するのは不適切であろう。

しかしながら、前節で明らかになった特に 5 語基の高出現頻度の語基構成パターン、すなわち 2-1-1-2-2、2-1-1-2-2、2-1-2-1-2、2-1-2-2-1、2-2-1-2-1 では、先頭語基の品詞は名詞とサ変名詞でほぼ 90% になる。2-2-2-2 パターンについても、類似の傾向をもつことがわかる。別の見方をすれば、先頭語基がサ変名詞である事例は、上記の語基構成パターンとなる場合が高いことがわかる。

#### 4.3 係り受けパターン

表 4.3 以降に語基数毎の先頭語基の品詞による係り受けパターンの種類を示す。

表 4.3 3 語基構成熟語の係り受け構造  
毎の先頭語基品詞による品詞列数

| 名詞 | 動詞 | 接頭 | 数詞 | 形容 | 形動 | サ変 |
|----|----|----|----|----|----|----|
| A  | 1  | 0  | 0  | 0  | 0  | 0  |

表 4.4 4 語基構成熟語の係り受け構造  
毎の先頭語基品詞による品詞列数

| 名詞 | 動詞 | 接頭 | 数詞 | 形容 | 形動 | サ変 |    |
|----|----|----|----|----|----|----|----|
| A  | 28 | 1  | 0  | 0  | 1  | 9  | 17 |
| B  | 18 | 1  | 0  | 0  | 0  | 7  | 10 |
| C  | 16 | 0  | 0  | 0  | 0  | 9  | 12 |
| D  | 18 | 0  | 0  | 0  | 0  | 7  | 10 |
| E  | 21 | 1  | 0  | 0  | 1  | 11 | 10 |
| F  | 5  | 0  | 0  | 0  | 0  | 2  | 1  |
| G  | 1  | 0  | 0  | 0  | 0  | 0  | 0  |

4 語基において並列構造を持つ F、G 以外での先頭品詞による係り受けの大きな偏差はない。

表 4.5 5 語基構成熟語の係り受けパターン  
毎の先頭語基品詞による品詞列数

| 名詞 | 動詞 | 接頭 | 数詞 | 形容 | 形動 | サ変 |    |
|----|----|----|----|----|----|----|----|
| A  | 36 | 6  | 22 | 6  | 10 | 10 | 16 |
| B  | 43 | 1  | 8  | 4  | 10 | 4  | 24 |
| C  | 9  | 0  | 4  | 1  | 5  | 0  | 6  |
| D  | 27 | 0  | 15 | 2  | 2  | 8  | 12 |
| E  | 2  | 0  | 0  | 0  | 0  | 0  | 1  |
| F  | 1  | 0  | 0  | 0  | 1  | 0  | 1  |
| G  | 1  | 0  | 0  | 0  | 0  | 1  | 0  |
| H  | 32 | 2  | 10 | 3  | 7  | 6  | 18 |
| I  | 5  | 0  | 2  | 3  | 0  | 1  | 0  |
| J  | 3  | 0  | 1  | 0  | 0  | 1  | 0  |
| K  | 4  | 0  | 0  | 0  | 0  | 1  | 1  |
| L  | 9  | 0  | 3  | 0  | 0  | 3  | 0  |
| M  | 13 | 0  | 0  | 0  | 0  | 9  | 11 |
| N  | 2  | 0  | 0  | 0  | 2  | 0  | 1  |
| O  | 2  | 0  | 1  | 0  | 0  | 1  | 0  |
| P  | 3  | 0  | 0  | 0  | 0  | 0  | 0  |
| Q  | 1  | 0  | 1  | 0  | 0  | 0  | 0  |

5語基では、A,D,Hでは、名詞の割合が多い他は接頭語の先頭品詞の比率が多くなるが、Bに関しては名詞・サ変の割合に比べ、接頭語が先頭品詞となる比率が小さくなることからわかる。

表4.6 6語基構成熟語の係り受けパターン  
毎の先頭語基品詞による品詞列数

|    | 名詞 | 動詞 | 接頭 | 数詞 | 形容 | 形動 | サ変 |
|----|----|----|----|----|----|----|----|
| A  | 5  | 0  | 0  | 4  | 0  | 0  | 0  |
| B  | 6  | 0  | 2  | 3  | 1  | 0  | 0  |
| C  | 0  | 1  | 2  | 0  | 0  | 0  | 0  |
| D  | 3  | 1  | 2  | 1  | 0  | 0  | 0  |
| E  | 1  | 0  | 0  | 0  | 0  | 0  | 0  |
| F  | 12 | 1  | 8  | 3  | 1  | 0  | 2  |
| G  | 1  | 0  | 3  | 0  | 1  | 0  | 0  |
| H  | 2  | 0  | 0  | 0  | 0  | 0  | 3  |
| I  | 2  | 1  | 3  | 2  | 1  | 0  | 0  |
| J  | 2  | 0  | 0  | 4  | 0  | 1  | 0  |
| K  | 1  | 0  | 2  | 0  | 0  | 1  | 0  |
| L  | 1  | 0  | 0  | 0  | 0  | 0  | 0  |
| M  | 1  | 0  | 0  | 0  | 0  | 0  | 2  |
| N  | 2  | 0  | 0  | 0  | 0  | 0  | 0  |
| O  | 1  | 0  | 0  | 0  | 0  | 0  | 0  |
| P  | 4  | 1  | 2  | 1  | 0  | 0  | 1  |
| Q  | 2  | 0  | 0  | 0  | 0  | 0  | 0  |
| R  | 1  | 0  | 0  | 0  | 0  | 0  | 1  |
| S  | 1  | 0  | 0  | 0  | 0  | 0  | 0  |
| T  | 0  | 1  | 1  | 0  | 1  | 0  | 0  |
| U  | 0  | 0  | 1  | 0  | 0  | 0  | 0  |
| V  | 0  | 0  | 1  | 0  | 0  | 0  | 0  |
| W  | 0  | 0  | 1  | 1  | 0  | 0  | 0  |
| X  | 0  | 0  | 3  | 0  | 0  | 0  | 0  |
| Y  | 0  | 0  | 1  | 0  | 0  | 0  | 0  |
| Z  | 0  | 0  | 0  | 1  | 0  | 0  | 0  |
| 2A | 0  | 0  | 0  | 1  | 0  | 0  | 0  |
| 2B | 0  | 0  | 0  | 1  | 1  | 0  | 0  |

6語基でB,F,I,Jを見るとB,Fでは名詞の比率が大きいことがわかるが、I,Jでは名詞より接頭語や数詞の占める比率が大きくなっていることがわかる。

8文字熟語では、語基数が増えるにつれ、品詞列や係り受け構造が複雑化している。特に6語基では顕著で、多くの係り受けパターンでは、その係り受けパターンをなす品詞列は1種のみとなっている。

## 5. 終わりに

3.3節で述べたように、係り受けが同定困難な漢字熟語が多数存在しており、下記にその実例を一例として挙げる。

実例1(5語基)： 合成 軌道 角 運動 量  
品詞パターン：サ変 名詞 名詞 サ変 名詞



上の例では、「合成軌道角」と「軌道角運動量」が語基より大きなレベルでの単位語として考えられるため、どちらの語が全体の語の意味の中心となるかにより、係り受け構造が変わってくる。

7文字熟語の分析[9]でも指摘したが、<名詞>または<サ変>の連続が、8文字熟語では、多くの品詞列パターンに見られた。

実例2(5語基)： 二 極 管 整流 増幅 計  
品詞パターン：数詞 名詞 名詞 サ変 サ変 名詞

実例1および実例2もこの現象を含み、係り受け構造の決定に困難をもたらした。

謝辞： 本研究を進めるにあたって語基分割と係り受けのチェックをして戴いた榊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 3(2008).
- [9] 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 17(2008).