

# 文章分解項目

分解は次の三段階で行われる。

- i
- |      |       |      |
|------|-------|------|
| 1    | 2     | 3    |
| 元文字数 | 分析文字数 | Sen数 |
- 元文字数は、分析対象となった文章の文字数である。空白、記号、絵文字等々をすべて含めた文字数である。分解文字数は、空白、記号等を削除した後の文字数である。

Sen:センテンス KW:キーワード

- ii
- |      |      |         |         |      |         |     |      |     |      |       |   |    |     |      |      |      |
|------|------|---------|---------|------|---------|-----|------|-----|------|-------|---|----|-----|------|------|------|
| 1    | 2    | 3       | 4       | 5    | 6       | 7   | 8    | 9   | 10   | 11    | 12  | 13 | 14  | 15   | 16   | 17   |
| 体言数  | 全体名詞 | 普通・固有名詞 | 数詞を除く名詞 | 普通名詞 | 非分類普通名詞 | 成語  | 季語   | 時期  | その他  | 固有名詞数 | 人名  | 社名 | 地名  | 国名   | 元号   | 歴史用語 |
| 18   | 19   | 20      | 21      | 22   | 23      | 24  | 25   | 26  | 27   | 28    | 29  | 30 | 31  | 32   | 33   | 34   |
| その他2 | 代名詞数 | 人称代名詞   | 指示代名詞   | その他3 | 数詞      | 年月日 | 金額   | 単位  | その他4 | 副詞数   | 状態  | 程度 | 陳述  | その他5 | 接続詞数 | 順接   |
| 35   | 36   | 37      | 38      | 39   | 40      | 41  | 42   | 43  | 44   | 45    | 46  | 47 | 48  | 49   | 50   | 51   |
| 逆説   | 因果   | 添加      | 並列      | 補足   | 選択      | 転換  | その他6 | 連体詞 | 感動詞  | 感動    | 呼びかけ  | 応答 | 掛け声 | 挨拶   | 用言数  | 形容詞数 |
| 52   | 53   | 54      | 55      | 56   | 57      | 58  | 59   | 60  | 61   | 62    | 63  | 64 | 65  | 66   | 67   | 68   |
| 動詞   | 自動詞  | 他動詞     | 一般      | 付属語数 | 助動詞     | 使役  | 受身   | 打消し | 推量1  | 打消推量  | 過去  | 希望 | 丁寧  | 様態   | 伝聞   | 比況   |
| 69   | 70   | 71      | 72      | 73   | 74      | 75  | 76   | 77  | 78   | 79    | ii は単語単位に分解される。単語別の頻度数がカウントされ、使われた単語の頻度総数と種類数の2種類で表される。 |    |     |      |      |      |
| 断定   | 推量2  | 丁寧断定    | 助詞数     | 格助詞  | 接続助詞    | 副助詞 | 係助詞  | 終助詞 | 間投助詞 | 連語数   |   |    |     |      |      |      |

- iii
- |       |          |          |          |         |         |         |         |        |           |           |       |                             |       |       |       |       |
|-------|----------|----------|----------|---------|---------|---------|---------|--------|-----------|-----------|-------|-----------------------------|-------|-------|-------|-------|
| 1     | 2        | 1        | 2        | 3       | 4       | 5       | 6       | 7      | 8         | 9         | 10    | i と ii の後、分析値を求める前の計算が行われる。 |       |       |       |       |
| 計算文字数 | 1単語平均文字数 | 総頻度数     | 出現単語数    | 否定語数    | 否定ワケ    | 非不単語数   | ない数     | 助動詞否定  | 感想単語      | 感想名詞数     | 感想用言数 |                             |       |       |       |       |
| 11    | 12       | 13       | 14       | 15      | 16      | 17      | 18      | 19     | 20        | 21        | 22    | 23                          | 24    | 25    | 26    | 27    |
| 同義語数  | 反意語数     | 同頻度数     | KW数      | KW第1重複数 | KW第2重複数 | KW第3重複数 | KW第4重複数 | KW感想1  | KW感想2     | KW感想9     | 中心KW  | 補完単語数                       | 補完感想1 | 補完感想2 | 補完感想9 | 中心補完語 |
| 28    | 29       | 30       | 31       | 32      | 33      | 34      | 35      | 36     | 37        | 38        |       |                             |       |       |       |       |
| 1頻度単語 | 1頻度単語感想1 | 1頻度単語感想2 | 1頻度単語感想9 | 中心1頻度   | 第1単語頻度  | 第2単語頻度  | 第3単語頻度  | 第4単語頻度 | Sen単語値最大値 | Sen単語値平均値 |       |                             |       |       |       |       |

# 分解項目の解説

分解する最小単位は単語である。その基準は品詞区分になり、単語種類になる。一つ一つの単語が基準になって、品詞に当てはめられ、カウントされる。品詞のカウント数、単語単位のカウント数(頻度数)である。

最初に文章の文字数がカウントされる。空白、記号を含めた書かれた時の文章の文字数である。次に記号、空白を削除した文字数がカウントされる。前の文字数を元文字数、記号削除後の文字数を分析文字数とする。報告文、論文等では、元文字数と分析文字数の差はほとんどない。メール文、SNS、ブログなど書かれた文章では差が生じる。この差を見るだけでも文章の性質が分かる。

品詞単位で単語の種類数と、単語単位の頻度数が計算される。1単語の頻度数よりも、品詞単位での単語の種類数が、分析では大きな役割をする。語彙量、類語、反意語の使い方で、知識分野や思考構成の方法と程度を窺い知れる。

複数文章、または分野やテーマの異なる文章を比較できる分析でなければ、基準値が求められない。仮に求められたとしても、基準値が適切であるかの判断材料が得られない。如何なる文章であっても、同じ視点、同じ基準で、分析が得られなければならない。

日本語の文章は、漢字、仮名、カタカナ、アルファベットが混在する。日本語の単語の音は短く、外来語等のカタカナ語は音が長い。文章を構成する単語群の形状によって文字数が変化する。単語形状によって左右されない文章の文字数を計算しているのが、計算文字数である。計算文字数は、1単語平均文字数、総単語数から求められる。

日本語は、時制、推量、断定などは、助動詞で判断できる。内容の構成は、自立語である体言と用言で見分けられる。日本語には近似文がある。近似文は1文から助動詞、助詞を除いた文で、およその意味を表せる。主語は助詞によって設定されるが、主語の概念が乏しく、主体と客体で区分される。用言ではサ変動詞があり、「育成する」などの名詞と「する」が合体して動詞となる。複合化されたサ変動詞は1単語としてカウントされる。

カウントの方法が2種類あって、一つが総頻度数、他が単語種類数である。すべての品詞でカウントされるが、近似文を構成する自立語が主な分析値を構成している。総頻度数は出現した自立語の総数である。単語種類数は、使われた単語の種類の数である。総頻度数と単語種類数の比率は、1単語の平均重複率を表す。

使われた単語がすべて異なる頻度数にならず、同じ頻度の単語がいくつもある。同じ頻度の単語種類数を表したのが同頻度数としている。

『分解項目』の ii から iii の間で単語重量値の計算がされる。同頻度数と1単語の頻度数、計算文字数で計算され、キーワード群、補完単語群、1頻度単語群に分類される。単語重量値は『基本分析機能1』に記載している。

分析辞書には、体言辞書、用言辞書、類語辞書、反意語辞書がある。1文章で使われた、類語、反意語が単語重量値と共に抽出される。類語、反意語は、語彙量等の計算に用いられている。また、作者を特徴づけたり、テーマに対しての領域を明らかにするために活用される。

センテンス単語値平均値は、1センテンス内で使われた自立語の単語重量値の平均を示している。キーワード群と、このセンテンス単語値平均値がキーセンテンスを求める参考になる。