

# 基本分析機能1

## A 基本分解分析機能

基本分析は一定の文字量の文章を必ず分解し、分析をする項目を挙げている。すべての分析結果には、ここで挙げられている分析内容はすべて実施されている。文章を分解するために辞書が必要となる。体言辞書、用言辞書、反意語・類語辞書などがある。カタカナ、アルファベットで登録されていない単語群は自動で分解される。

1

### 文章を単語単位で品詞設定

a

最小品詞単位で分解する。複合語は、複合化される前の単語と複合化された単語群に区別される。複合化されて意味が変化する場合は複合語で1つと判断されている。単語別、品詞別、分野別、活用の一部の意味別にカウントする。

2

### 文章中単語の意味の重さ

a

すべての各単語はカウントされ、文章全体の文字量に合わせてカウント比率が求められる。基準が文章文字数になっているので単語の文字数を計算し、他の文章との比較ができる数値に置き直す。相対的比率を分析値に置き換える。この値を**単語重量値**と言い、文章中で作者による単語の意味の重さとする。

3

### 基準単語への変換

a

表現の仕方で、短縮形などを標準の表現に置き換えている。動詞、形容詞、助動詞は、終止形に置きかえてから分析を行う。データは元文と標準にした単語群で構成された2つを持つ。

a

1文章単体で分析できる項目、複数文章も同じ

g

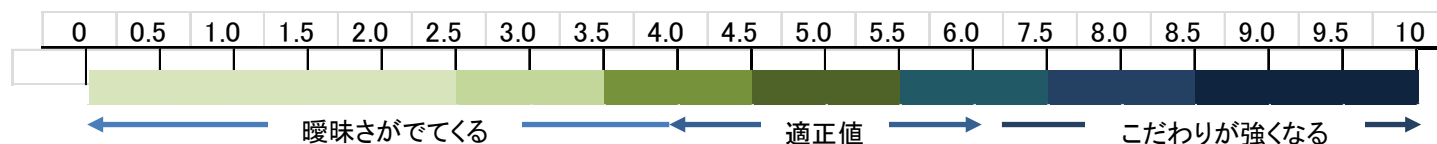
複数文章で分析できる項目

## 基本分析機能2

### 4 他文章比較用分析値抽出

a

35項目の分析値がある。正規分布している分析値を16項目取り出し、常に表示できる値として活用している。表される分析値のすべては、異なる文章、異なる文字量の文章などと比較しても同じ意味で分析値を比較できる値にしている。各分析値の意味も同様の扱いになっている。分析値の基準は5.0であり、この値より大きくなるとこだわりが出てきて、小さくなると曖昧になっていると判断できる。



### 5 同意語・反意語の抽出

a

分析時に使用された単語群に対して、同じ文章中で同意語と反意語が抽出されている。これらは、伝達意識を計測したり、他データの検索要因として使われる。

### 6 習慣使用単語の抽出

a

1つの文章で、よく使われている連体詞、接続詞、副詞、感嘆詞、などが抽出されている。

### 7 キーワードの抽出

a

第1キーワード群

第1補完単語群

一頻度単語群

右のように、1文章に表された単語を7つのブロックに分類している。第1から第3のキーワード群が適正に書かれている文章で4%程

第2キーワード群

第2補完単語群

第3キーワード群

第3補完単語群

度、1頻度単語群が80%前後の比率で構成されている。この構成比で、文章表現の適正さが検証できる。趣旨、キーセンテンスもこの構成から計算される。

## 基本分析機能3

8

### キーセンテンスの抽出

a

文章中での単語の意味の重さ(=単語重量値 以下 $W_g$ と表す)を用いて1センテンス重量値を現す。1センテンス重量値は感嘆詞、連体詞、接続詞を除く体言と用言の $W_g$ 合計とこれらの単語数から平均を求める。さらに、1センテンス内にキーワードが含まれているセンテンスで、1センテンス重量値の大きいものをキーセンテンスとする。

⇒単語重量値の説明は『文章分析<文道>の分析方法』を参照

9

### 思考&テーマ分野の抽出

a

体言辞書の単語の1つ1つに分野が設定されている。文章で使われた単語群で辞書に分野が設定されている単語の出現数とそれらの $W_g$ から分野が計算される。1分野の固定ではなく、複数分野の比率で表される。

10

### 近似文への変換

a

センテンスから付属語(助詞、助動詞)を除いたセンテンスを近似文と言う。近似文は他のセンテンスの同一性の発見、他言語への変換に用いられる。

11

### テーマ分野自動設定

a

特定分野について、自動設定されている。分野辞書を用いるのではなく、政治、経済、国際、金融、社会、事件についてのみ自動設定される。これらは複数分野の特定ではなく、1分野特定に用いられる。