

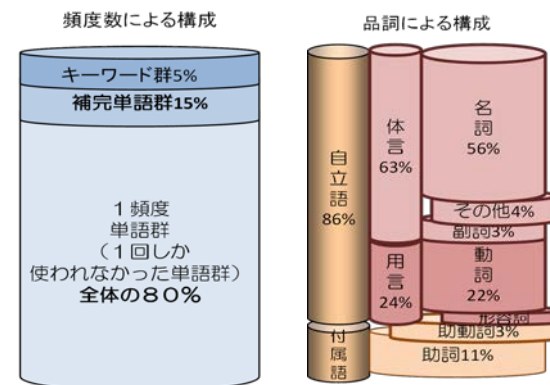
文章分析で、判別できる、分かる事柄

- ☑01 表現ジャンル区分ができる。
- ☑02 異分野、他文章との表現レベル比較が可能である。
- ☑03 文章データの 카테고리区分ができる。
- ☑04 文章の知識&要素を抽出構成できる。
- ☑05 読みやすさを客観的に判定できる。
- ☑06 趣旨抽出ができる。
- ☑07 キーワード群が抽出できる。
- ☑08 論旨構成が表現できる。
- ☑09 標準からの比較で論旨難易度が測定できる。
- ☑10 テーマに対しての得意性が判定できる。
- ☑11 反意語検索、類語検索が可能である。
- ☑12 意識的形式知、暗黙知が抽出できる。
- ☑13 意識的曖昧さを確定できる。
- ☑14 単語に対しての作者の意識の強さが表される。順位が付けられる。
- ☑15 同じ単語を、別の文章データの中での意識の強さが比較できる。
- ☑16 作者、グループ等の意識変化が抽出できる。
- ☑17 文章の作者が特定できる。
- ☑18 知識習熟度が測定できる。
- ☑19 時間経過しても、社会表現形態が変化しても同じレベル、同じ基準で分析ができる。時代比較ができる。
- ☑20 異なる趣旨の複数文章での比較が可能である。
- ☑21 異なる作者の意識、性格等の相関を測定できる。
- ☑22 文章の趣旨から同類の文章検索ができる。(文章から文章の検索)

etc.

言葉を素材にした分析論

文章分析<<文道>>の仕組み



800 字以上の数万の文章を分解、分析し、異常値を取り除き、平均を求めた値が左図になる。日本語文章の構造の一つである。

2000 年 12 月以降、朝日、産経、日経、毎日、読売の社説を毎月、分析し、表現構造の変化がないかを確認している。社説の分析だけですでに 4 万件を超えている。

文章分析の形

文章分析では基本となる分析値基準の持ち方が大切である。1つの文章を分析しても、多数の文章を分析しても、同じ基準で結果が出てこなければ判断ができない。文章には異なるジャンルがある。論文、報告文、エッセイ、小説、紀行文、日記があり、文語文、口語文がある。無作為に文章を分析して、結果が小説と論文等のジャンル区別がつかなくてはならない。同じジャンルでも作者の区別がつかなくてはならない。逆に、異なるジャンルであっても、類似テーマの抽出が出来なくてはならない。小説と論文で同じテーマ、同じ趣旨の文章が引き出せなければ、分析できているとは言えない。

弊社文章分析<<文道>>システムでは、分析基準を求めるために、各文学ジャンルの文章を合わせて数十万件を分析した。その中では、報告文に類する文章が最も多い。作者年齢では中学生以上の文章であり、社会人が大半を占める。分析結果は、「数と言葉」で表す。「数」は、統計で言うところの正規分布をなす 16 項目を見い出した。「言葉」はキーワード群とキーセンテンスで表し、カテゴリー分類ができるようにした。数値は表現、意識の強さを表す。複数の分析値の組み合わせで人材スタイルを見出す。言葉は単語を基準として、単語単位で、単語に対しての意識の強さを数値で表せる。分析値は異なる文章ジャンル、異なるテーマ間で比較できるようにした。

開発にあたり、目的や仮説を持たないようにした。眼の前にある文章をあるがままに、分析を試みた。文章の分解から始まり、分析値では単語間の相関、相反を捉えようとした。表現されている姿の発見の連続だった。人と文章、集団と文章、集団と集団などの比較ができていく。現在では、分析した文章は 100 万件を超す。現在もなお、研究を行っている。「文は人なり」を実感する日々である。

文章分析の対象と目的



言葉があれば、分析できないモノはない。

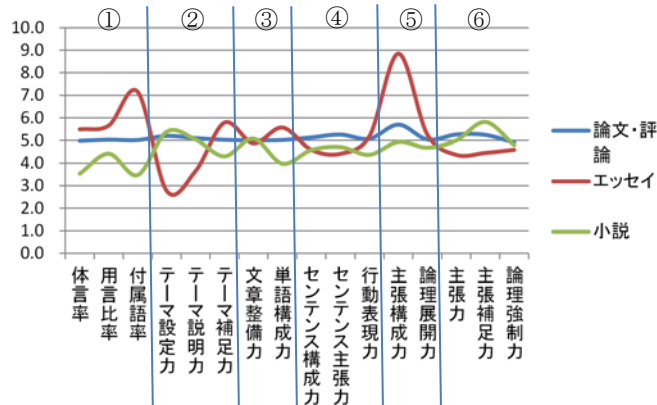
言葉が一定以上の量があり、対象言語であれば、分析できる。文字量によって精度が影響される。文章であれば分析できる。1単語であれば、分析でなく検索になる。分析エンジンが持っている辞書内容検索になる。分析理論上、200文字あれば分析対象になるが、精度を保証できるのは400字以上の文章である。昇格等など、精度の高い分析を望むのであれば、800字以上の文章になる。

文章分析エンジンには、辞書が必要である。体言辞書、用言辞書、類語辞書、反意語辞書、カテゴリ辞書などがある。単語数は全体で20万語程度になるだろう。文法形式を示すアルゴリズムがある。文法は、言葉表現の後追いになっており、曖昧な箇所が多い。文章分析用の汎用性の高い文法ができてくる。プログラムコードで書かれる場合と辞書として用意されるものがある。これらを含めて文章分析エンジンになる。

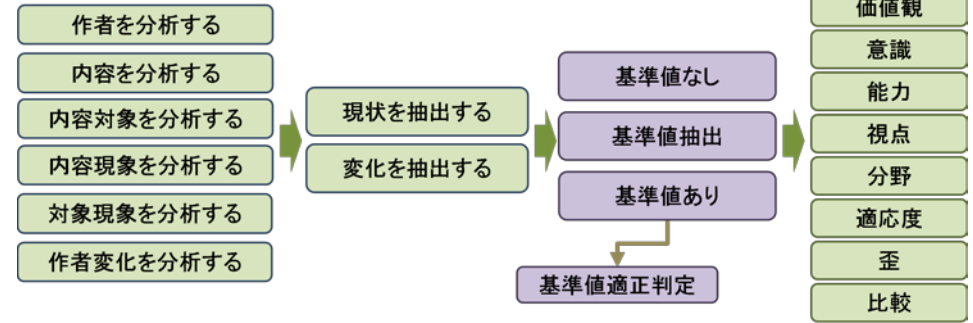
言葉を人が綴る。一人で綴ることはなく、何らかの刺激があつて綴る。一つの文章は、一つではなく、他との関係性が必ず存在している。複数の文章、複数人の文章を分析すれば、そのグループの状態が現れる。ここに分析の目的に多様性が現れる。

文章分析の分析値

下記グラフに表しているように、16の分析値を導きだしている。基準値は5.0に設定されている。評価点に置き換えるとすれば、5.0が満点で、5.0からの差が減点になる。5.0より大きくなるとこだわりが強くなり、5.0より小さくなると曖昧さが大きくなる。但し、5.0が常に良いのではなく、文章ジャンルによって、5.0になる基準値は変化する。右に表しているグラフは多数の論文の論理的ターゲットを基準にして、分析値を求めている。16すべての分析値が5.0になると評価点は満点になるが、読後感としては面白くない文章になる。第三者的表現と解釈する。読み手を意識した文章、喜怒哀楽が現れている文章、好感が持てる文章など、それぞれにスタイルがある。



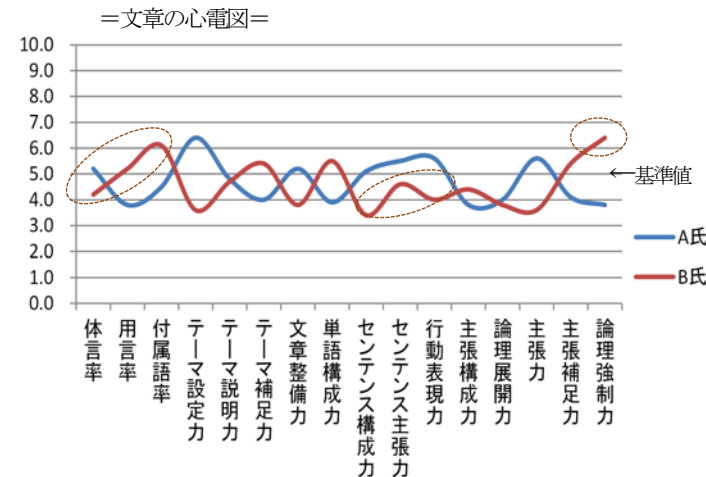
文章分析の目的



文章の心電図

下記の左右ページのグラフを「文章の心電図」としている。16項目の分析値を折れ線グラフにした。分析値は、左下のグラフに示したように6つのブロックに分かれている。6つの計算方法、扱い項目が異なっている。①は対人性と関わりを持ち、③は基本となる文章力、④は強調性と関わりを持っている。グラフ形状と分析値から、表現されたテーマに対しての人材像が浮かび上がる。複数人材の集合を取り出せば、互いの関係程度が伺える。

	体言率	用言率	付属語率	テーマ設定力	テーマ説明力	テーマ補足力	文章整備力	単語構成力	センテンス構成力	センテンス主張力	行動表現力	主張展開力	主張力	主張補足力	論理強制力	
A氏	52	38	45	64	48	40	52	39	51	55	56	38	40	56	41	38
B氏	42	52	61	36	47	54	38	55	34	46	40	44	38	36	54	64



【或る事例】

左グラフの2人の分析値を評価点に換算すると、90点を超え、ほぼ同じ点数であった。AB両氏は共に優秀であった。ところが、B氏は優秀であるが、A氏は、協調性がなく、独善的で優秀ではないと社内で評価された。

B氏はグラフからは、控え目で馴染みややすい話し方をし、意見を畳みかける。○で囲っているところで示されている。

A氏のグラフはB氏とは逆である。A氏は一人で仕事をする方が向いている様である。思考&表現スタイルが違っているために、行動観察では、優劣がつかってしまった。