

言葉を素材にした分析論

文章分析へのアプローチ

(株)シードウィン

文章分析のアプローチ

人が 書いて **人が** 読む。

いろんな 人が **いろんな** 気持ちで **いろんな** 状態で書いて、
いろんな 人が **いろんな** 気持ちで **いろんな** 状態で読む。

これらの組み合わせで、無数の結果が現れてくる。

無数の 文字データが、出来事が

毎日 出現してきて絡みあい、 **無数の** 現象を引き起こす。

いつ、どこで、何が、どんな形で、どのように、
自分の仕事や、生活に影響してくるのか？

私たちに関係するモノを引き出すのも大変

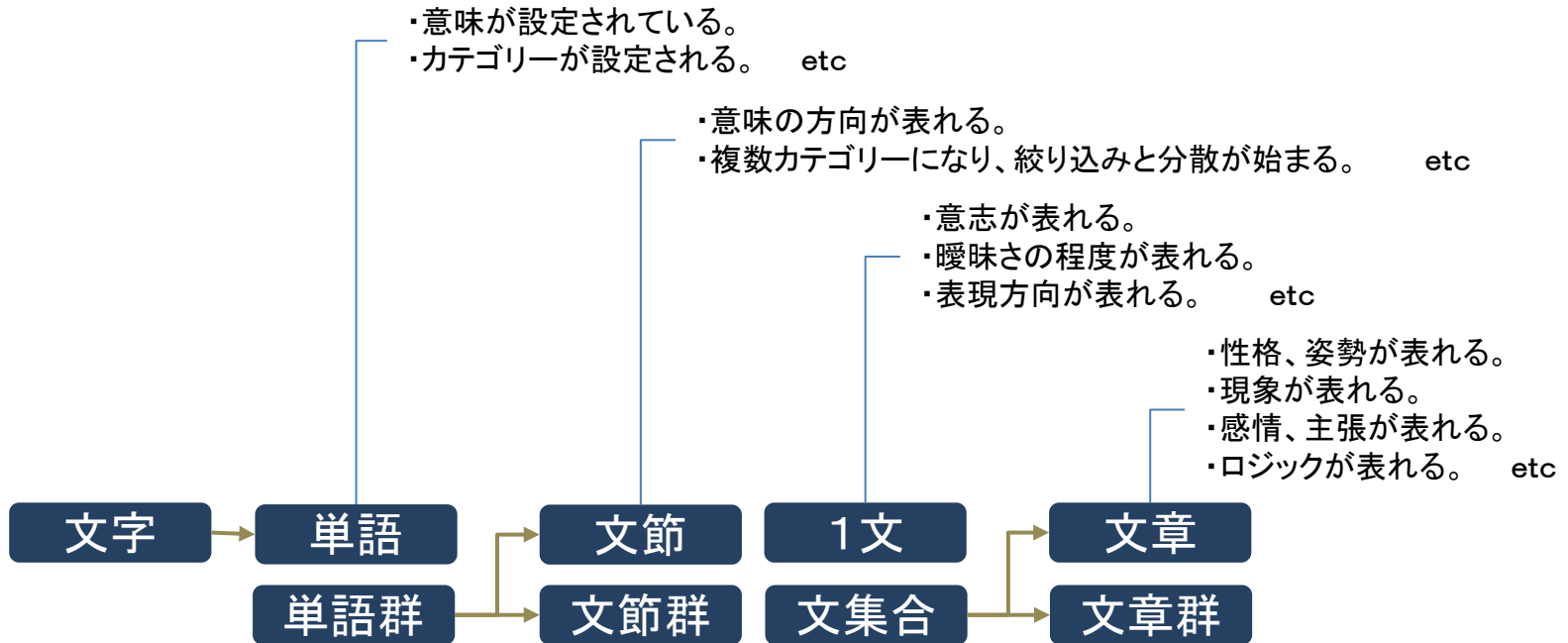
結果から方向を見定めるのも大変

1人の人、複数の人、たくさんの人
の気持ちと行動を解きほぐし、
私たちの進むべき方向を導きだす。

文章分析は、そのような指針を導きだす。

分析対象—文字データ

1文字でも、その1文字が何を現し、何を求めているか、認識できるのが一番良い。残念ながら1文字は分析対象にならない。1文字が1単語を表せば分析対象になる。認識できれば、対象に対して如何なる対応が可能かを探りあてるのが目的となる。



複数文節群、1文の集合の分析などはアンケート分析などに活用される。
複数の文章群になる時系列変化、意識変化、価値変化などの変化が現れる。

文字データ分析の最大の利点は、無限と言えるほどの文字データであっても、一括で意味を把握できる。推測を確信に変える。

文章の表現ジャンル

表現ジャンル別に、形式が培われ、共通の構造をもつようになった。

表現ジャンル		
報告文	評論	論文
日記	紀行文	エッセイ
		小説
挨拶文	Eメール	手紙
法律文	特許文	職業文

●文章スタイルは大きく、論文形態、エッセイ形態、小説形態に分類される。

評論、報告文は論文形態と類似する。論文は学会等の習慣的表現が強く現れている場合が多い。

●日記文は、読者を意識して書かれた場合と、誰にも読ませる意識がない場合によって、表現形態、論旨構造が違ってくる。ここで区分しているのは、読者を意識しているとして分類した。

●職業文は職種によって表現構造が異なる。概して、その職種外の人には読みにくい。

■の枠内で表現されている組み合わせが分かりやすく読みやすい文章になる。

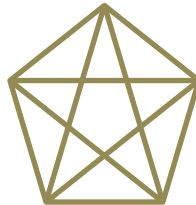
論理構造のある文章

論理構造のない文章

論理が単数構造の文章

論理が並列表現の文章

論理が複数構造の文章



命題が明瞭な文章

命題が曖昧な文章

命題が不明の文章

因果関係が明瞭な文章

因果関係のない文章

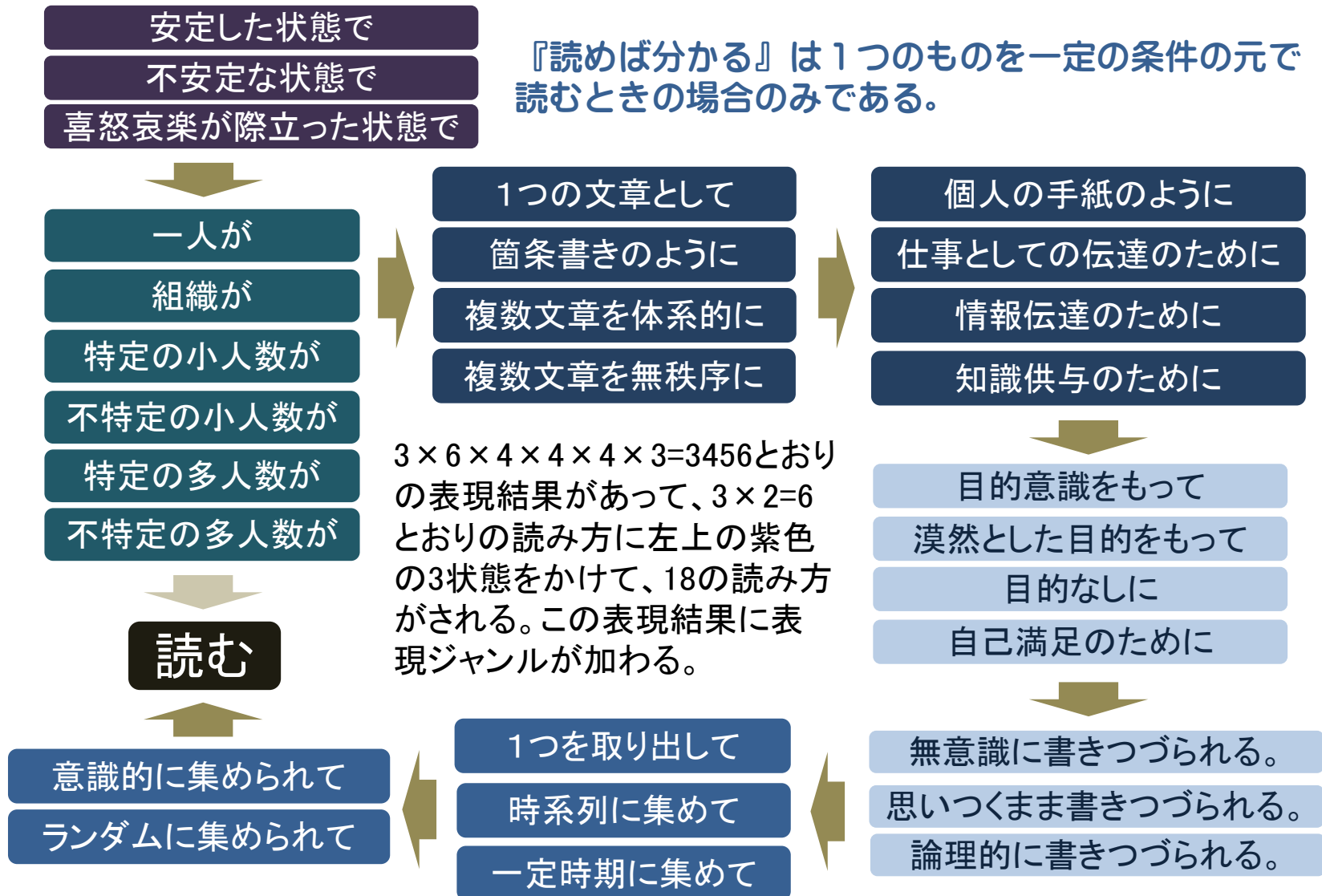
表現技法の優れた文章

表現技法の拙い文章

●手紙文は個人の習慣が良く現れる文章である。ジャンルに分類されるような特定の表現構造を持たない。

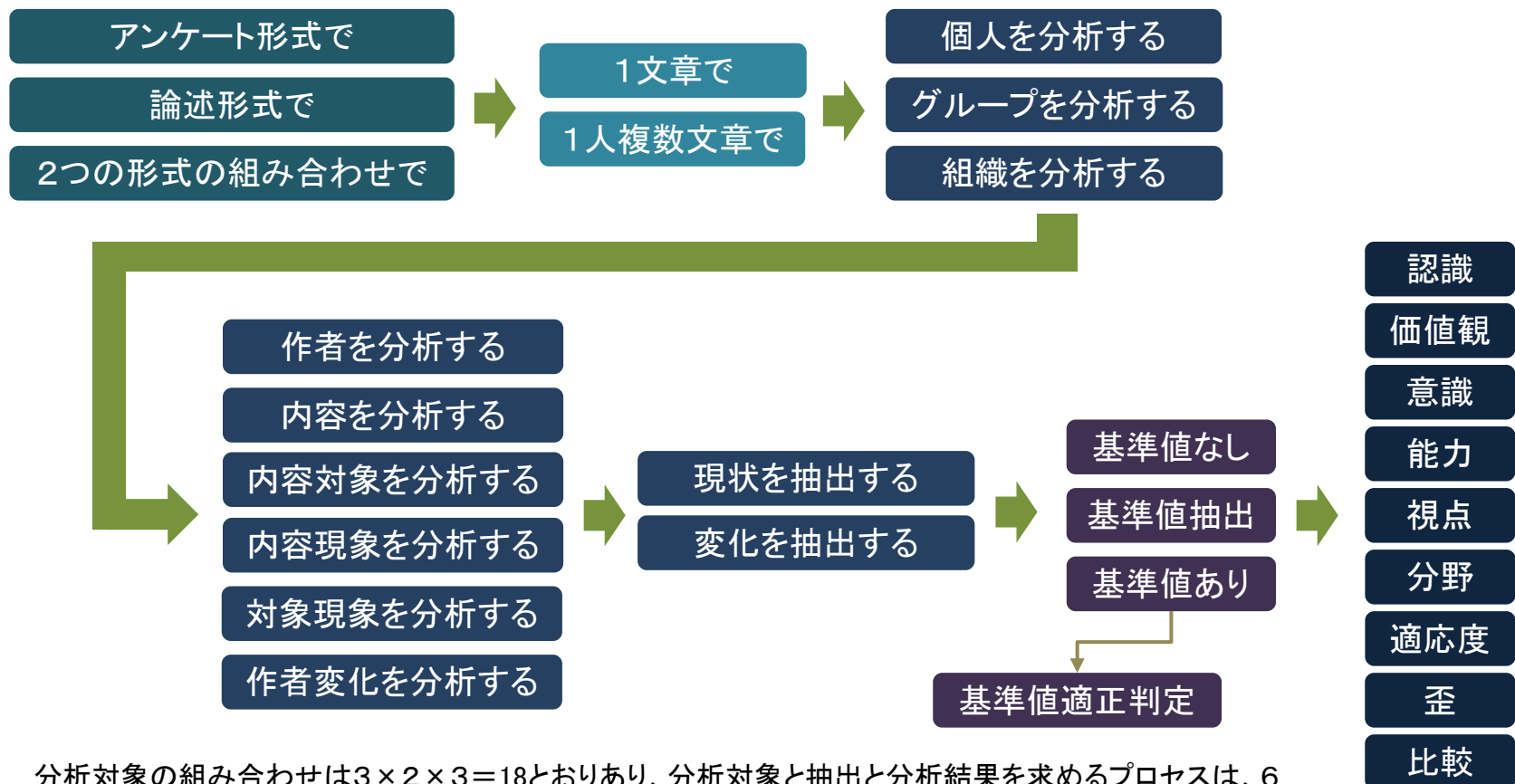
●挨拶文は形式に則っている場合が多く、無味乾燥なイメージを受けやすい。読み物としては最下位のレベルである。

表現された文章の背景



文字データ分析対象と分類

分析目的は、現状認識と確認であり、意思決定の材料抽出が目的となるはずである。

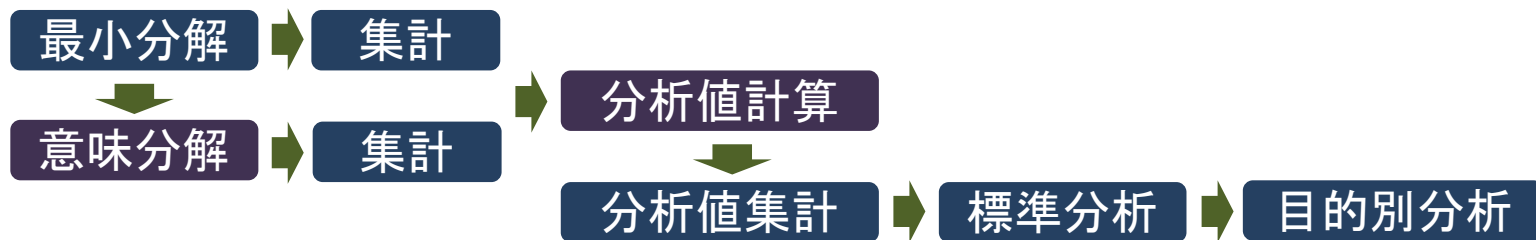


分析対象の組み合わせは $3 \times 2 \times 3 = 18$ とおろりあり、分析対象と抽出と分析結果を求めるプロセスは、 $6 \times 2 = 12$ とおろりで、求める項目は $3 \times 9 = 27$ ある。全体を合わせると $18 \times 12 \times 27 = 5832$ とおろりとなる。求める項目を2つ、3つと組み合わせを複合化させると限りなくあると言ってもよいだろう。分析を求める目的に対して絞り込みも、そのための材料分析が必要となる。

いかに分析するか

基準となる『はかり』がなければ分析のしようがなく、基準がなければ仮説を設けるしかない。仮説はどこまでも仮説で、実証されなければ実践では活用できない。すべての仮説が実証されるとは限らず、実証できた範囲から実践し、データ収集を繰り返しながら、実証範囲を広げていく。

分析する工程



●品詞単位、単語単位、活用単位、用言活用単位、複合単語、類語、反意語単位での分解とそれぞれの区分別の集計を行う。

●1文章の文字数、ひらがな、漢字、カタカナ、アルファベットでの文字数が、文章によって異なるので、計算上、1単語の文字数を標準化する。

●分析値は、異なる文章で同じ基準で比較できる数値にしなければならない。

●分析項目が異なっても分析値が同じ基準になっていなければならない。

●1言語での統一基準、表現ジャンル別基準、1文章文字量区分での基準などがあり、基準同士での比較ができないなければならない。

●分析値、表現スタイル、趣旨、キーワード群、表現難易度、主張姿勢とレベル、言語表現としての適正などが、基本分析となる。

●目的に応じて、分解、分析の組み合わせ、分析表見が変わる。

「文字データ分析対象と分類」で表されている項目などによって、標準分析を踏まえうえで、処理形式が変わる。

分析基準を設ける

1言語全体での基準

様々な表現ジャンル、老若男女、職種などすべてを含めたとき、日本語としての基準が必要である。⇒表現全体基準

表現ジャンル別基準

論文、報告文、エッセイなど、様々な表現ジャンルがある。表現ジャンルがあるのは、表現構造、表現展開が異なるからである。これらのジャンルの基準がある。

職種別基準

職業によって、思考構造が変わる。同じ言語であるのにも関わらず職種で変わる。研究職、営業職、経営者など、考える範囲と対象、日常に接する人々によって、表現構造、思考構造が構成されていく。

読みやすい基準

多くの人に親しまれやすい表現構造がある。分かりやすいだけでなく、リズム、肯定・否定、断定・推量などの組み合わせなどにより変化する。

●これらの基準は、表現全体基準があって、全体基準と比較して現れる。基準を設けるには、多量の文章群があってターゲットが求められる。

●表現は、思考スタイルと生活スタイル、習得されている知識範囲に影響される。望まれる表現構造は論理的に求められる場合もある。仮説設定であり、仮説ができてから、多量の文章群との比較検証は必要である。

分析精度&レベルチェック

分析基本

- 01 表現ジャンル区分ができるか。
- 02 異分野、他文章との表現レベル比較が可能か。
- 03 文章データの 카테고리区分ができるか。
- 04 読みやすさを客観的に判定できるか。
- 05 趣旨抽出ができるか。
- 06 キーワード群が抽出できるか。
- 07 論旨構成が表せられるか。
- 08 標準からの比較で論旨難易度が測定できるか。
- 09 テーマに対しての得意性が判定できるか。
- 10 反意語検索、類語検索が可能か。
- 11 意識的形式知、暗黙知が抽出できるか。
- 12 意識的曖昧さを確定できるか。
- 13 単語に対しての意味付けの強さが表されるか。
順位が付けられるか。
- 14 同じ単語を別の文章データでの意味付けの強さが比較できるか。
- 15 作者、グループ等の意識変化が抽出できるか。
- 16 文章の作者が特定できるか。
- 17 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か。時代比較ができるか。
etc

文章分析は様々な要件が絡み合って、意思決定ができる状態になる。

言葉の曖昧さを分析するのではない。意味の曖昧さを分析する。

前ページであらわされた
「文章の表現ジャンル」
「表現された文章の背景」
「文字データ分析対象と分類」
の項目と内容などを分析できるようにして、様々な現象をつかめられる。
言葉を材料にした分析が可能となる。