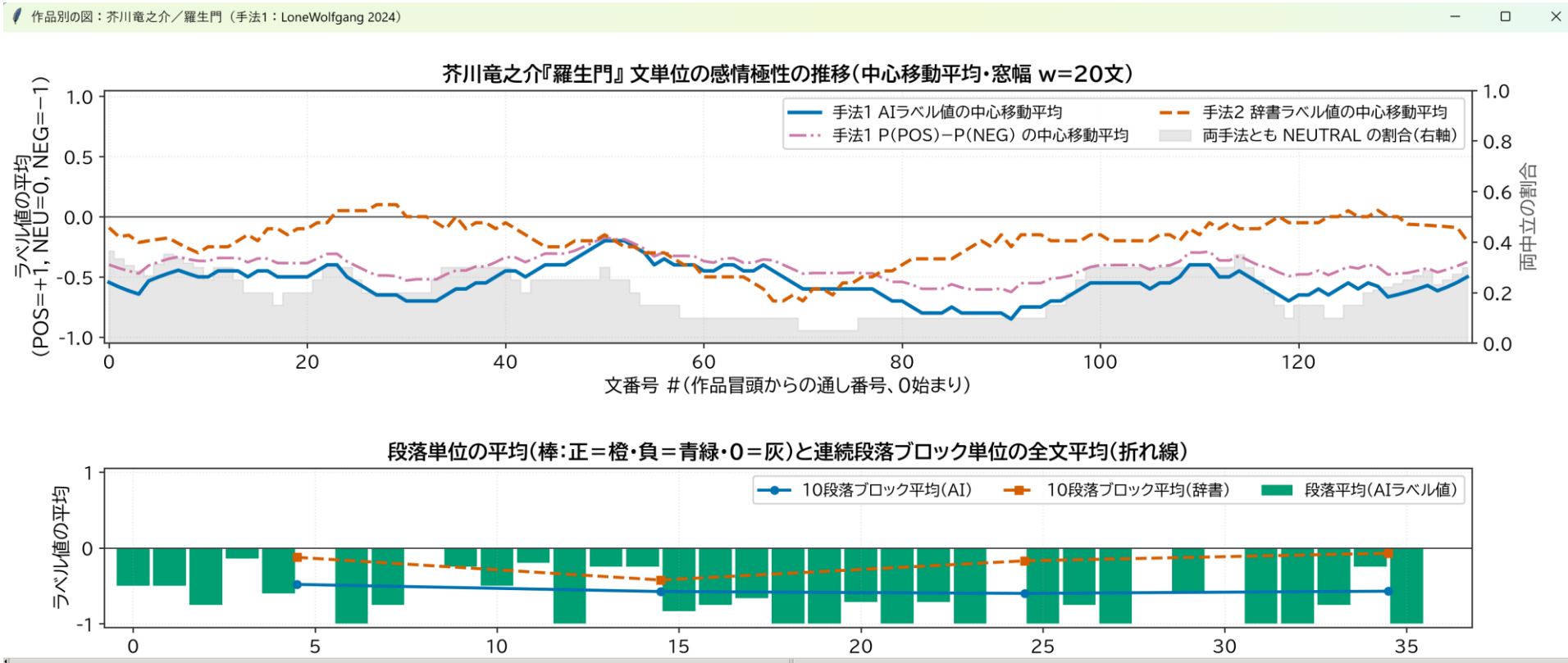
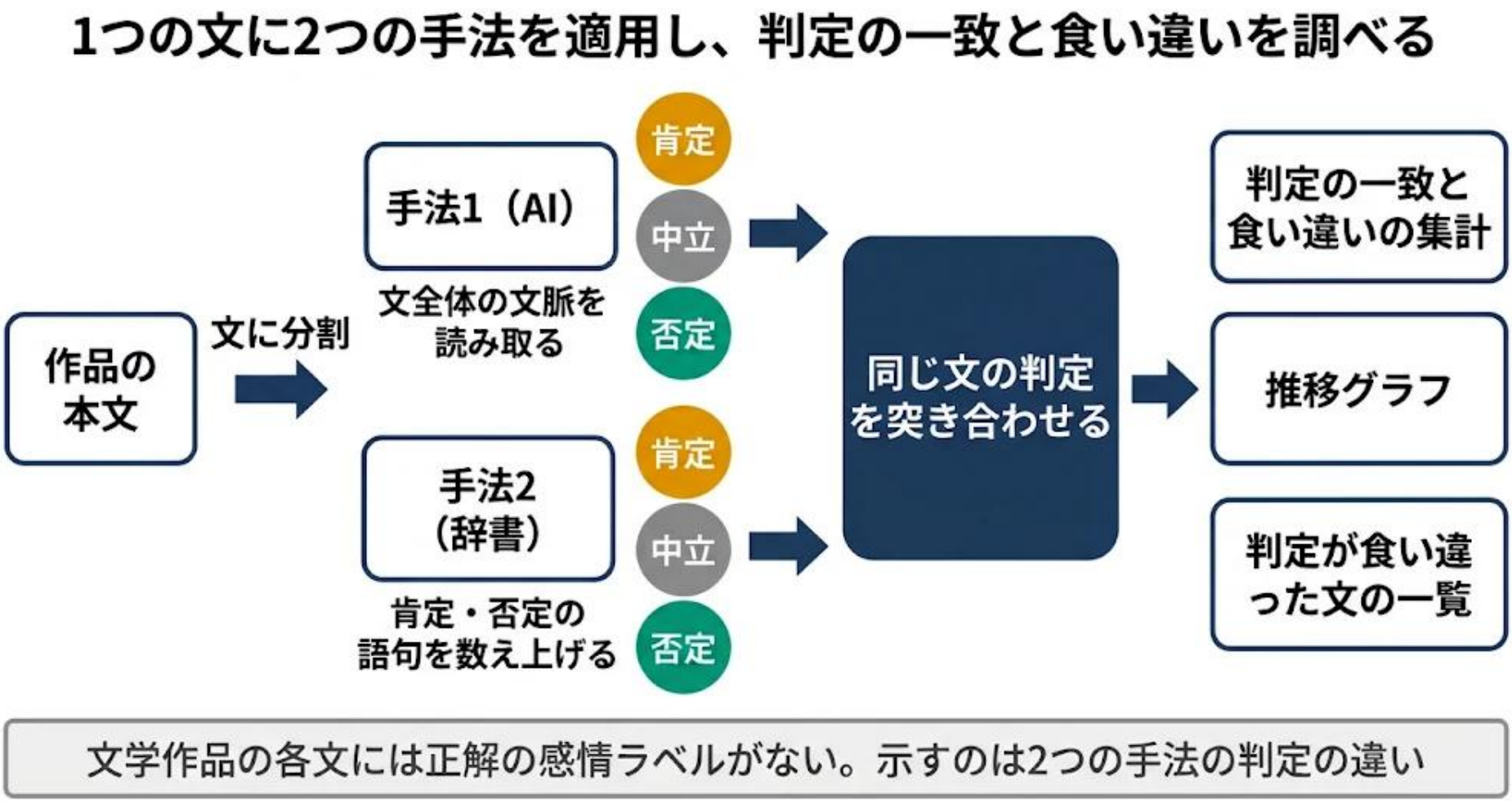
日本語文学作品の感情極性分析



資料中の図などは Google Nano Banana 2 を用いて作成

全体の仕組み



文学作品の各文には正解の感情ラベルがない。示すのは2つの手法の判定の違い

手法1（AI）の判定の仕組み

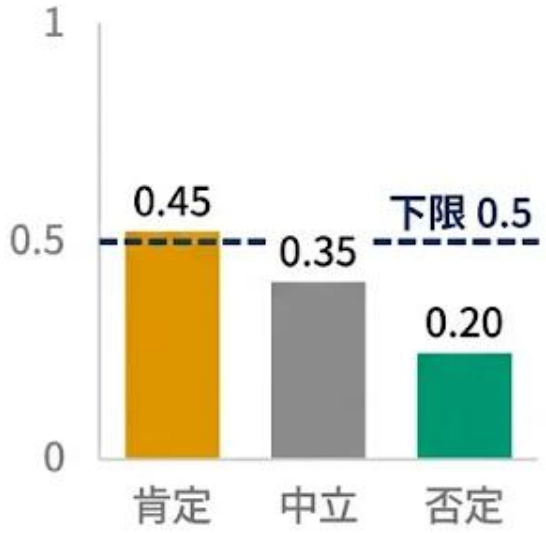
**3クラスの確率を出し、最大クラス確率が下限より低ければ中立、
それ以外は最も確率の高いクラスと判定する**

「今日は楽しかった。」 → 手法1（AI） → 肯定・中立・否定の3クラスの確率（合計1）



肯定

最大クラス確率
0.70 は
下限以上



中立

最大クラス確率
0.45 は
下限より低い

最大クラス確率は必ず1/3以上のため、下限の初期値0.34ではほとんど置き換えは起きない

下限を0.5にした例。数値は説明用

手法2（辞書）の判定の仕組み

極性表現の極性値を否定表現で反転し、その合計（初期値）を文極性値 v としてしきい値θと比べる

1 形態素に分ける

それ	は	美しく	ない	。

→

2 辞書と照合する

美しく

↓

基本形：美しい

↓

+1

文中で辞書と一致した語句を極性表現と呼ぶ

→

3 否定表現を見る

美しく　ない

直後3形態素以内に否定表現が奇数個あれば極性を反転する

+1 → −1

→

4 しきい値θと比べる

文極性値 v ＝ −1

θ＝1 の例

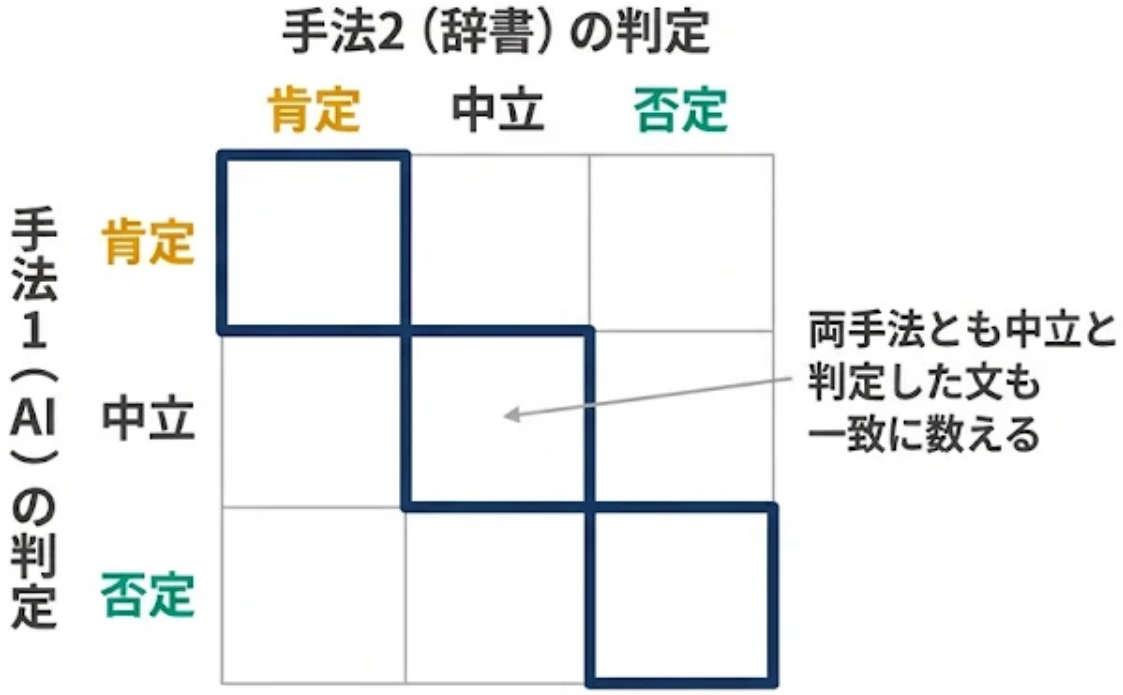


v が −θ 以下なので否定

極性表現が0件の文は中立

3×3分割表と一致度の読み方

各文を、手法1（AI）の判定（行）と手法2（辞書）の判定（列）の組み合わせのマスに数える



観測一致率

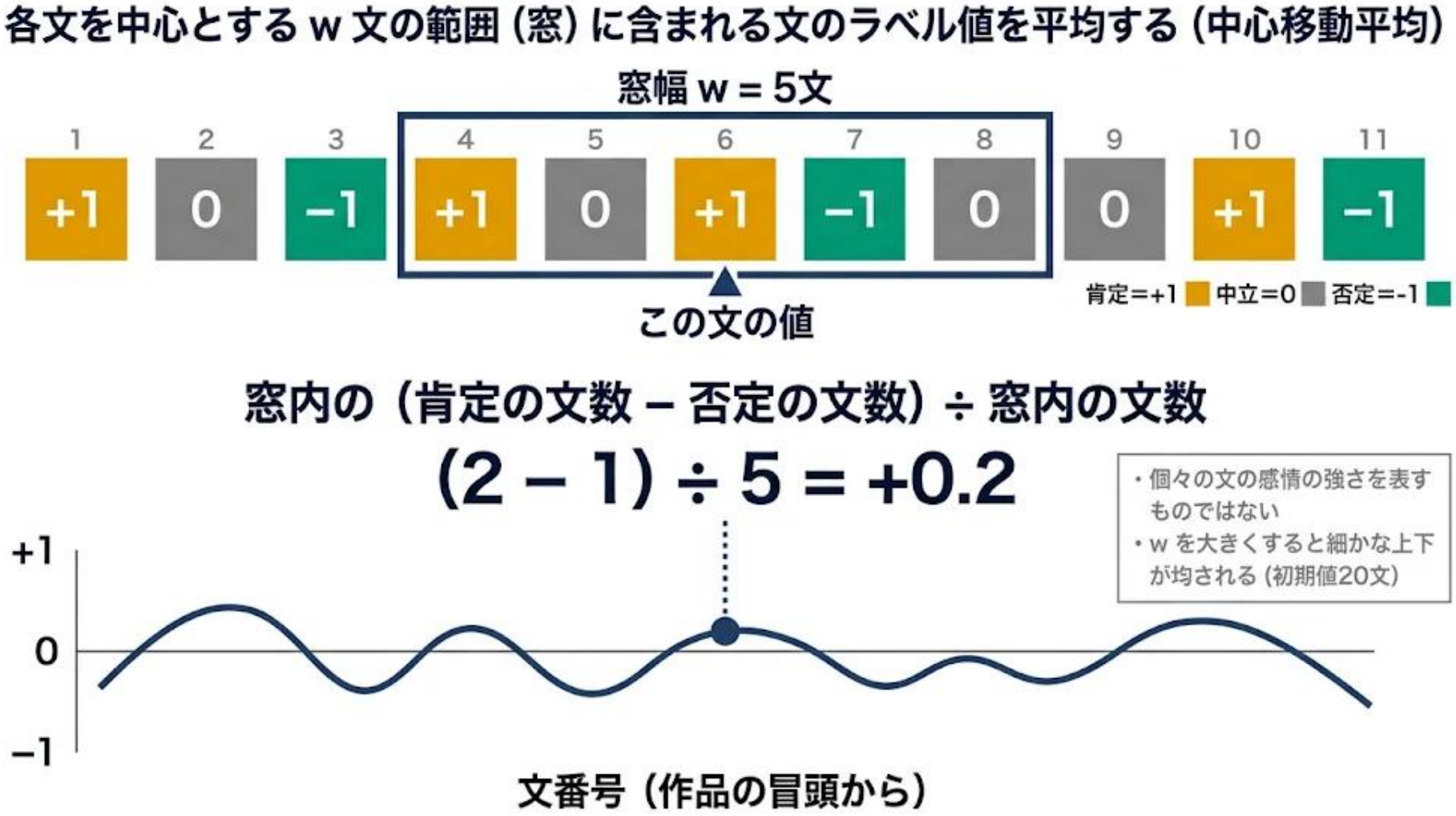
判定が一致した文の割合
（対角のマスの文数 ÷ 全文数）

Cohen の κ

偶然の一致の期待値を差し引いた一致度。1は完全な一致、0は偶然と同程度の一致

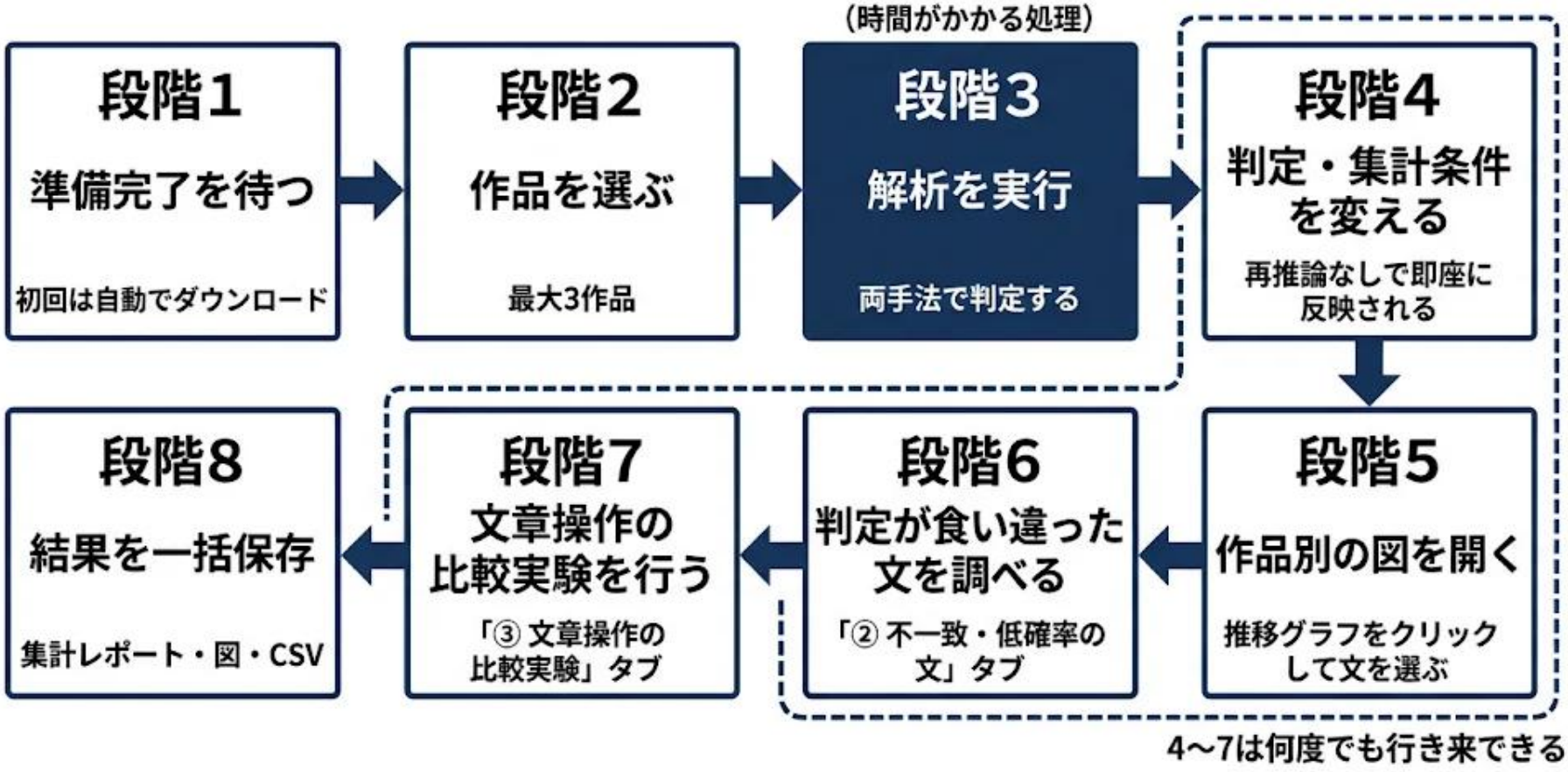
一致度は、どちらかの正しさを示すものではない

推移グラフ（中心移動平均）の読み方



進め方（画面の操作順）

一度解析すれば、判定・集計条件は再推論なしで何度でも変えられる



文章操作の比較実験（例と選べる操作）

文章操作を作品の全文に適用し、原文と変換後の文の判定を比較する

⑥ 否定表現の除去

原文	変換後の文
それは美しくない。	それは美しい。

	原文	変換後の文
手法2（辞書）	● 否定	● 肯定
手法1（AI）	?	?

手法1（AI）の判定がどう変わるかを実験で確かめる

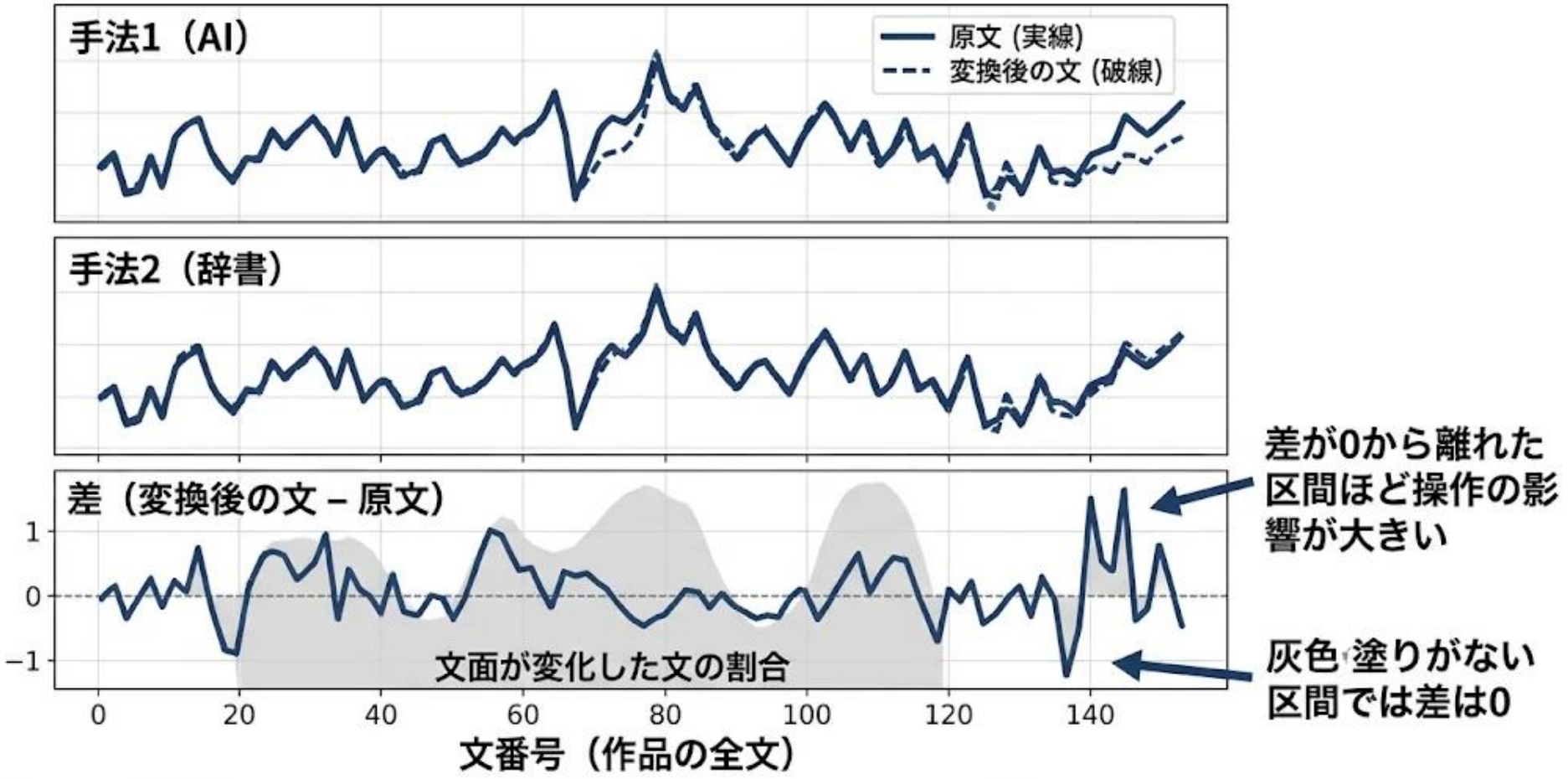
文章操作（1つを選ぶ）

① 括弧類の除去	⑥ 否定表現の除去
② 感嘆符・疑問符・三点リーダ等の除去	⑦ 修飾語の除去
③ である調（常体）への統一	⑧ 読点の除去
④ です・ます調（敬体）への統一	⑨ 感情語の伏せ字化
⑤ 文末の終助詞・間投助詞の除去	⑩ 全文のひらがな化

変換は規則に基づく処理であるため、変換されない文もある

文章操作の比較図の読み方

原文と変換後の文の推移を重ね、差が出た区間を見る

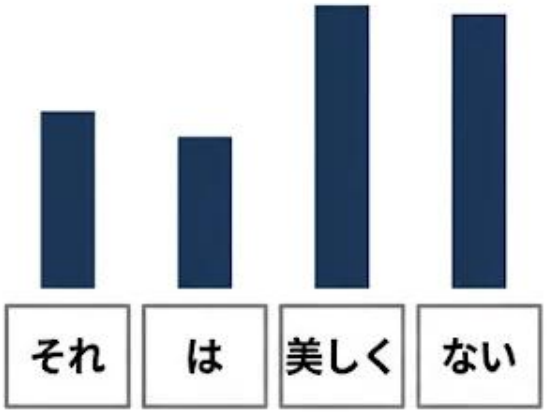


手法1（AI）の判定の理由を調べる3つの方法

AIが文中のどのトークンに注目したか、どのトークンが判定にどれだけ寄与したかを調べる

トークン：AIが扱う文字列の単位

自己注意の重み



最終層で、文全体を代表する位置が各トークンにどれだけ注目したか

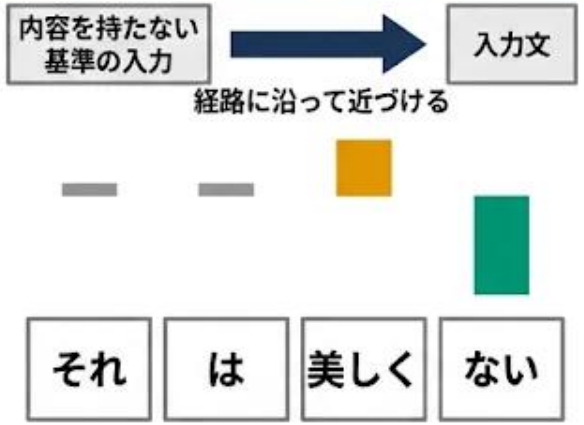
そのままでは判断の理由を表さない

遮蔽法



確率差 $P(\text{肯定}) - P(\text{否定})$ がどれだけ変わるか

統合勾配



寄与の合計 ≒
確率差 $P(\text{肯定}) - P(\text{否定})$ の変化量

分点数 m を増やすと両者の差が小さくなる

結果は手法による判定の説明であり、文の感情の正解ではない（棒とトークン分割は説明用の例）

判定・集計条件の変更とAIモデルの切り替え

変えた設定によって、再推論なしで即座に反映されるか、再解析が必要かが決まる

④ 判定・集計条件

- 最大クラス確率の下限
- 判定しきい値 θ
- 否定表現による極性反転
- 中心移動平均の窓幅 w
- など

↓

再推論なしで即座に反映される

③ 推論条件と作品の選択

- 解析する文数の上限 N
- GPU を使用する
- 作品の選択
- など

↓

再解析が必要
(「▶ 解析を実行」を押す)

② 手法1のAIモデル

一方での解析が終わると、もう一方での推論が裏で自動的に進む

計算済み	→ 即座に結果が切り替わる
推論中	→ 推論が完了した時点で自動的に切り替わる
未計算	→ 裏の推論が始まり、完了後に自動的に切り替わる

↓

AIモデルを切り替えると、文章操作の比較と統合勾配の結果は消去される

結果を解釈する際の注意

