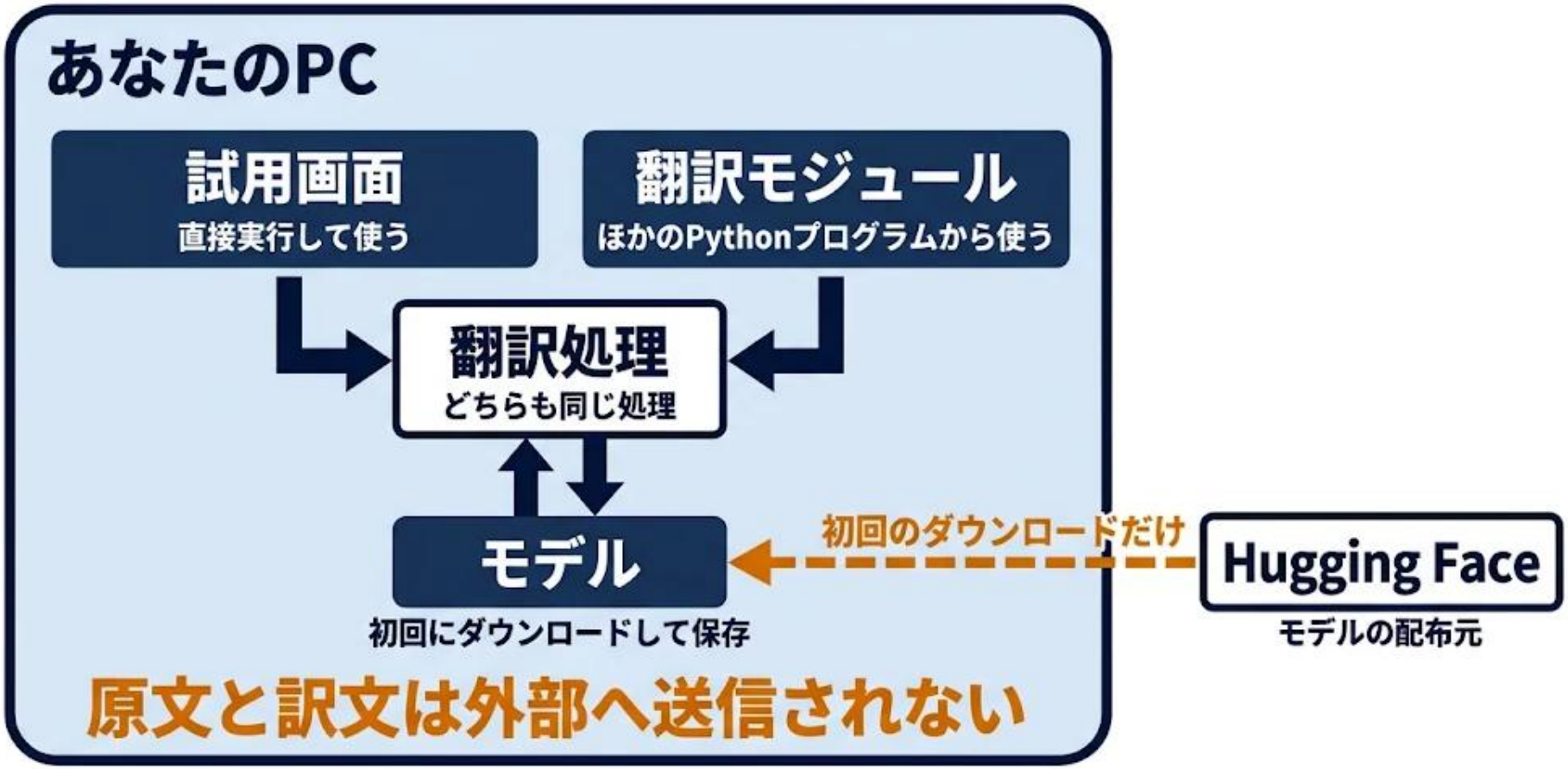
ローカル翻訳エンジンと試用 GUI（日本語・英語・中国語（簡体字）・インドネシア語）



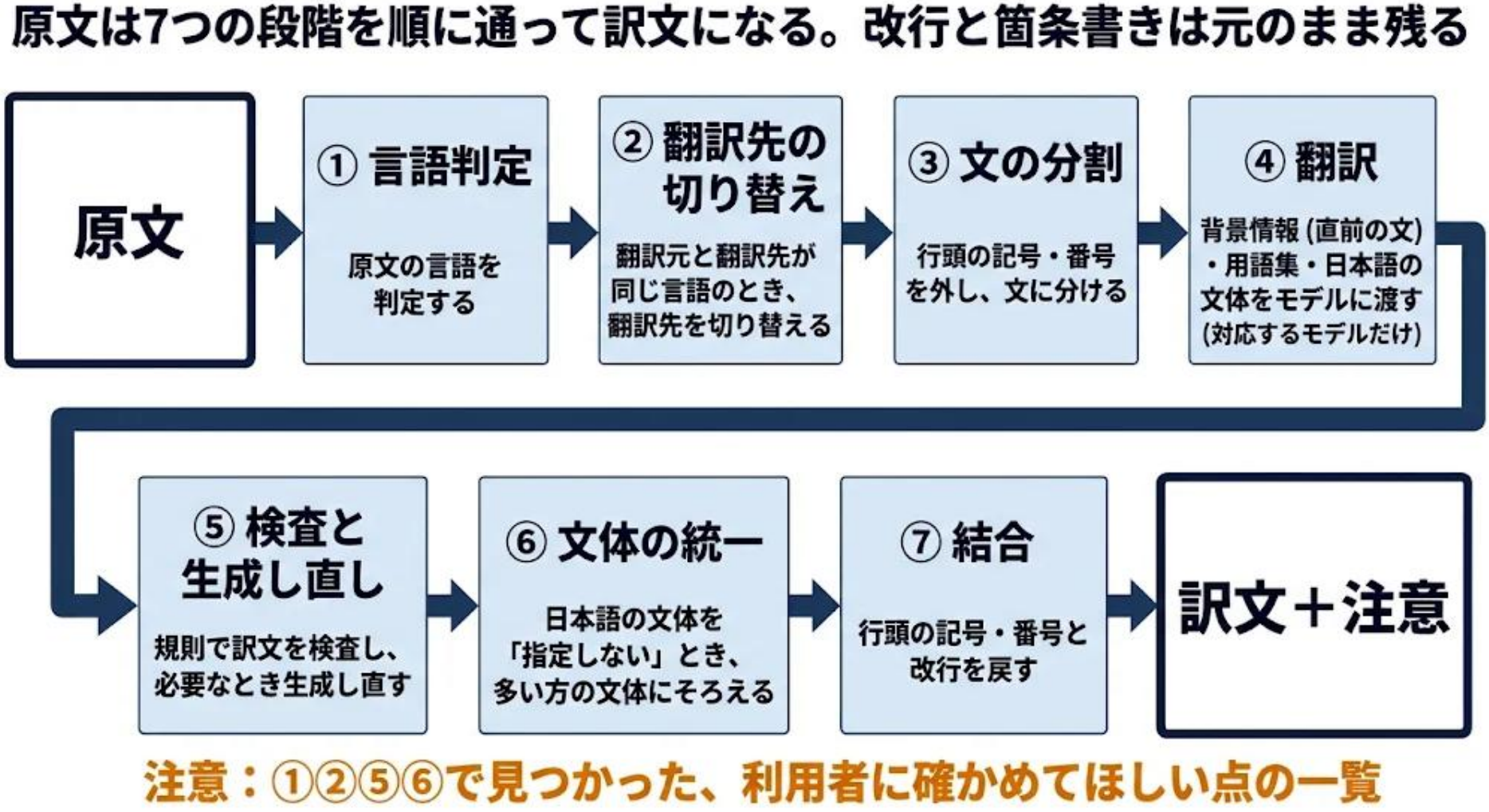
資料中の図などは Google Nano Banana 2 を用いて作成

全体像

日本語・英語・中国語（簡体字）・インドネシア語の4言語を、このPCの中で翻訳する

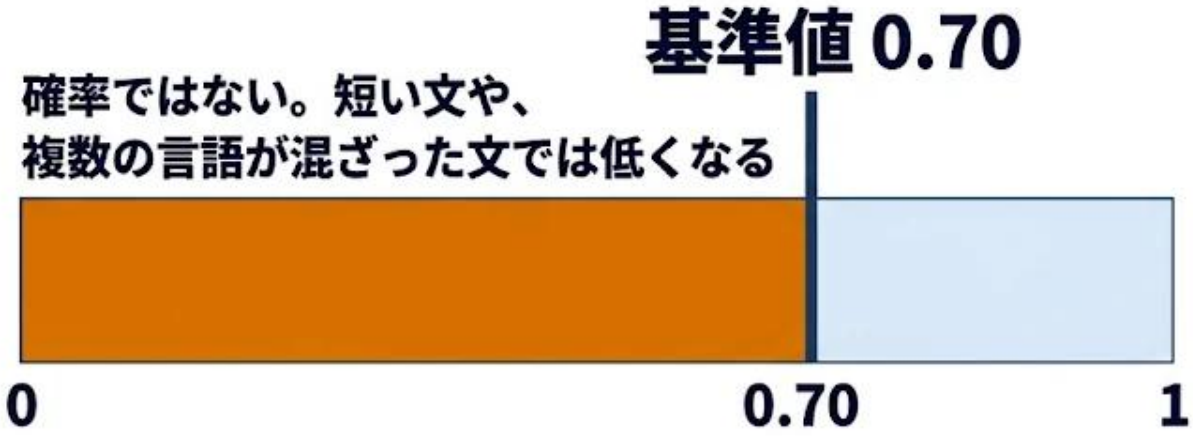


処理の流れ


原文は7つの段階を順に通って訳文になる。改行と箇条書きは元のまま残る
原文
① 言語判定
原文の言語を判定する
② 翻訳先の切り替え
翻訳元と翻訳先が同じ言語のとき、翻訳先を切り替える
③ 文の分割
行頭の記号・番号を外し、文に分ける
④ 翻訳
背景情報 (直前の文)・用語集・日本語の文体をモデルに渡す (対応するモデルだけ)
⑤ 検査と生成し直し
規則で訳文を検査し、必要なとき生成し直す
⑥ 文体の統一
日本語の文体を「指定しない」とき、多い方の文体にそろえる
⑦ 結合
行頭の記号・番号と改行を戻す
訳文＋注意
注意：①②⑤⑥で見つかった、利用者に確かめてほしい点の一覧

言語判定と信頼度スコア

翻訳元が「自動判定」のとき、原文の言語を判定して翻訳元にする。
信頼度スコアは、判定の確からしさを表す0〜1の値



基準値を下回る

判定した言語で翻訳し、
注意を表示する

基準値以上

判定した言語で
翻訳する

判定できない

- 文字を含まない
  (例：12:30)
  → 原文のまま返す
- 4言語以外の文字が
  半分を超える
  (例：ハングル)
  → 翻訳しない

翻訳先の切り替えと、文ごとの扱い

翻訳先と同じ言語の原文は、翻訳先を切り替えるか、その文を原文のまま残す

翻訳元と翻訳先が同じ言語のとき

翻訳先を切り替え、注意を表示する

日本語の原文 →	英語に翻訳する
英語・中国語・インドネシア語の原文 →	日本語に翻訳する

翻訳元が「自動判定」のとき（文ごとに判定する）

例：翻訳先は日本語

原文	
Please submit the report by Friday. →	英語 → 日本語に翻訳する
The office closes at 5 pm. →	英語 → 日本語に翻訳する
締切は金曜日です。 →	日本語（信頼度スコアが基準値以上）原文のまま残す

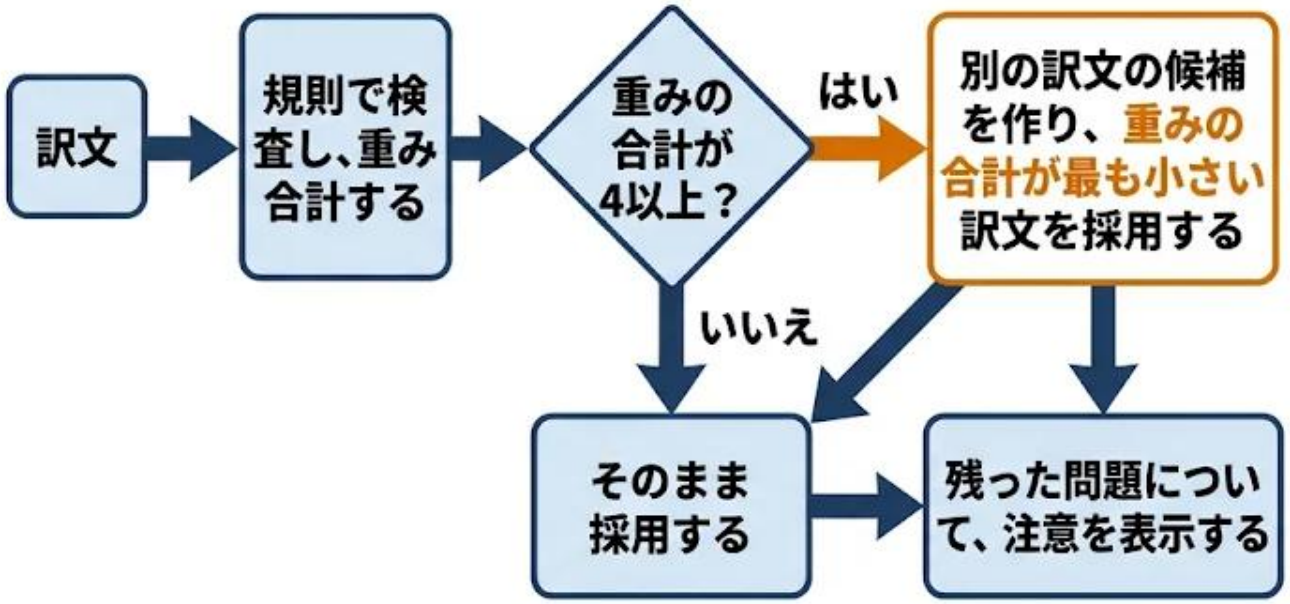
原文のまま残る文があるのは、この規則による

検査と再生成

検査項目ごとの重みを合計し、重みの合計が4以上の文だけを生成し直す

主な検査項目と重み	
訳文の言語の誤り	8
URLの欠落	6
数値の欠落	5
用語集の訳語の欠落	5
日本語の文体の不一致	4
英字の語や型番の欠落	2

重みは、問題の深刻さに応じて決めてある

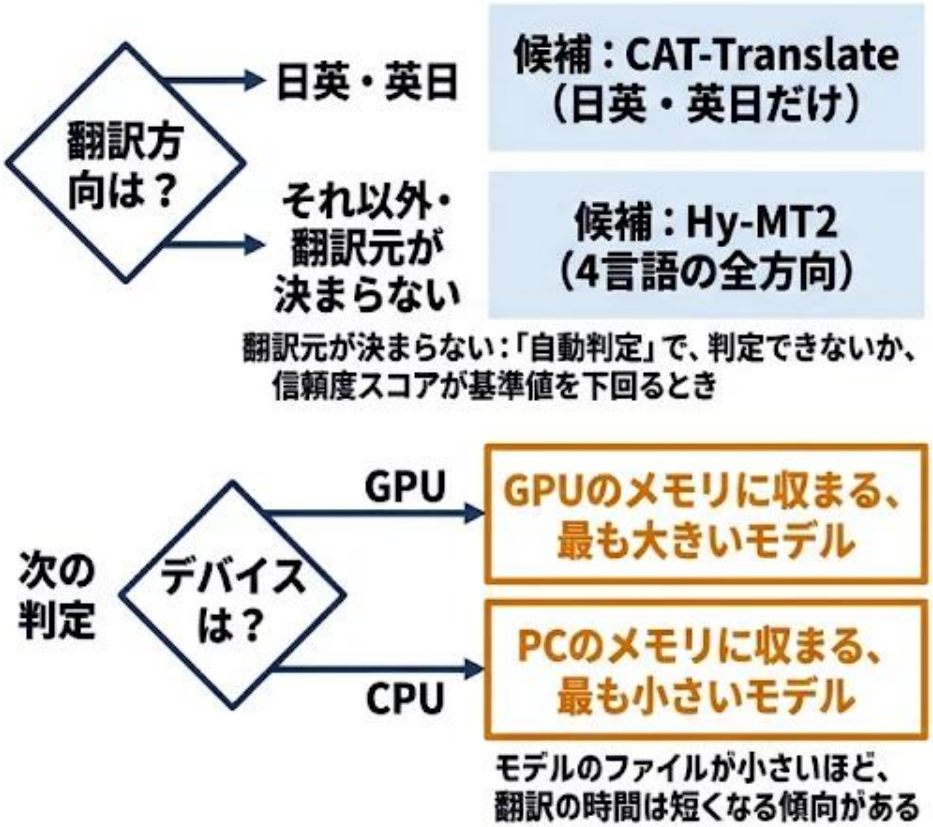


例：数値の欠落（重み5）→ 生成し直す
例：英字の語の欠落（重み2）だけ → 生成し直さず、注意を表示する

規則による検査であり、誤訳が無いことは保証しない

推奨のモデル

試用画面は、翻訳方向・デバイス・メモリから推奨のモデルを示す。モデルを選ぶのは利用者



メモリに収まるモデルが無いときは、推奨できるモデルが無いことを表示する

モデル	モデルのファイの大きさ（ダウンロードの容量）	必要なメモリの目安
Hy-MT2-1.8B	約4.1 GB	約5.7 GB
Hy-MT2-7B	約16.1 GB	約17.7 GB
CAT-Translate--1.4b	約2.8 GB	約4.4 GB
CAT-Translate--3.3b	約6.7 GB	約8.3 GB
CAT-Translate-7b	約15.0 GB	約16.6 GB

必要なメモリの目安 ＝ モデルのファイルの大きさ + 約1.6 GB
目安は見積もりであり、長い原文では足りなくなることがある

試用画面の配置



① モデルとデバイスを選び、読み込む

② 推奨のモデルと理由を表示する。
ボタンで推奨のモデルに切り替わる

③ 翻訳元・翻訳先・日本語の文体を選ぶ

④ 翻訳を始める・止める

⑤ 原文を入力する。用語集の欄に、
1行に1組「原語 = 訳語」と書く

⑥ 言語判定の行は、信頼度スコアが基準値をを下回るときと、判定できないときに赤字になる。ボタンを押すと、2番目の候補の言語を翻訳元にして翻訳し直す

⑦ 処理の状態とエラーを表示する

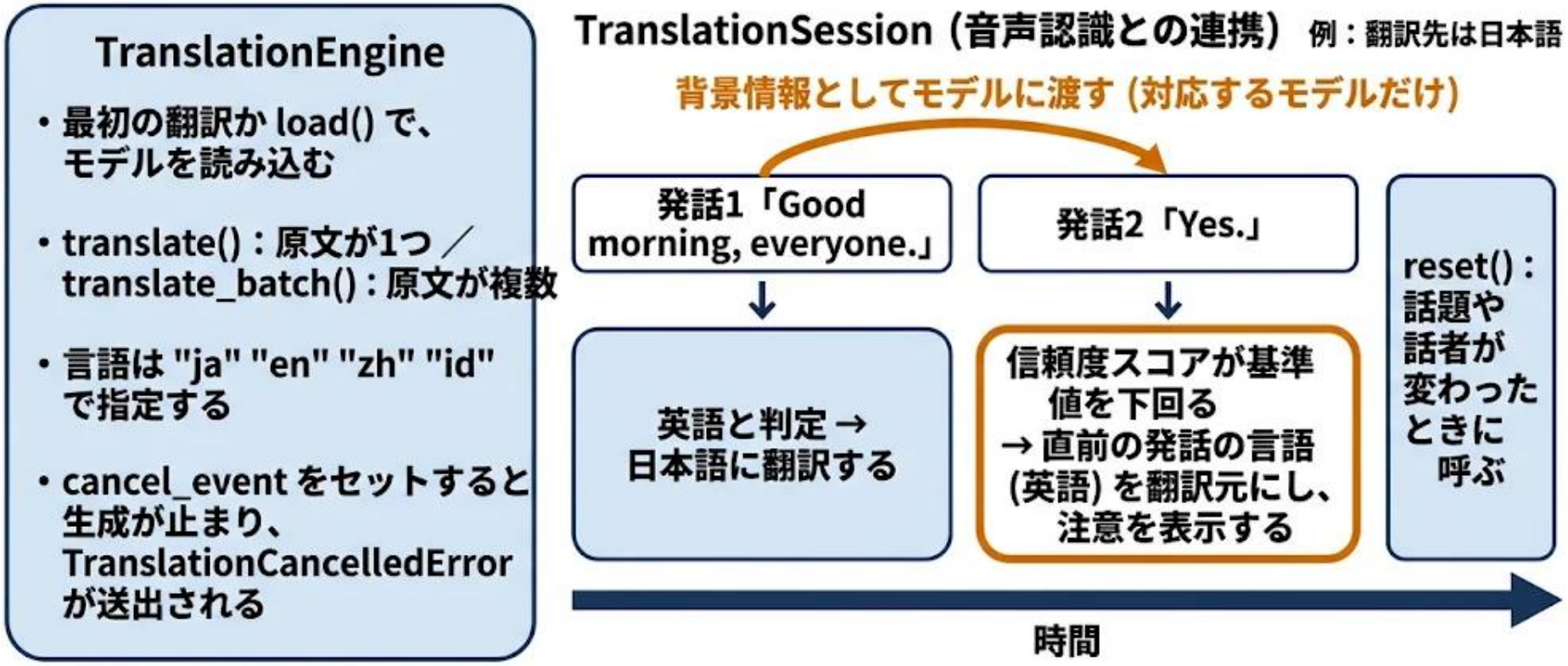
「対訳」タブ：文ごとに、原文・訳文・検査の結果を並べて表示する

キー操作

- Ctrl+Enter（原文の欄）：翻訳を始める
- Ctrl+Z（原文・用語集の欄）：入力を元に戻す
- F1：ヘルプを開く
- Esc：ヘルプから戻る

翻訳モジュールとしての利用

ほかのPythonプログラムから読み込んで使う。
音声認識との連携では、直前の発話を背景情報としてモデルに渡す



翻訳結果：訳文/実際に使われた翻訳元と翻訳先/言語判定の結果/処理時間/使用デバイス/使用モデル/注意の一覧