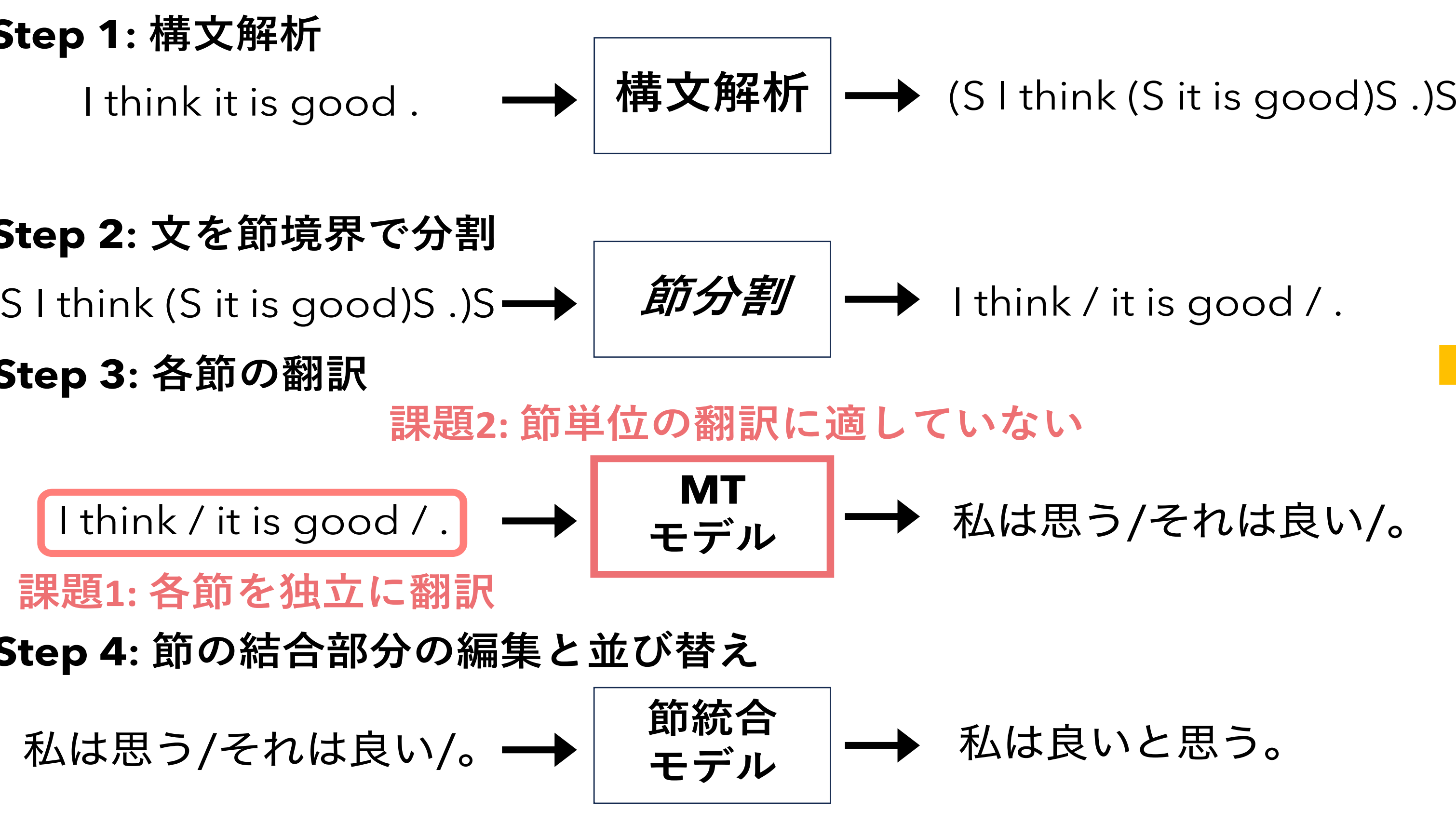


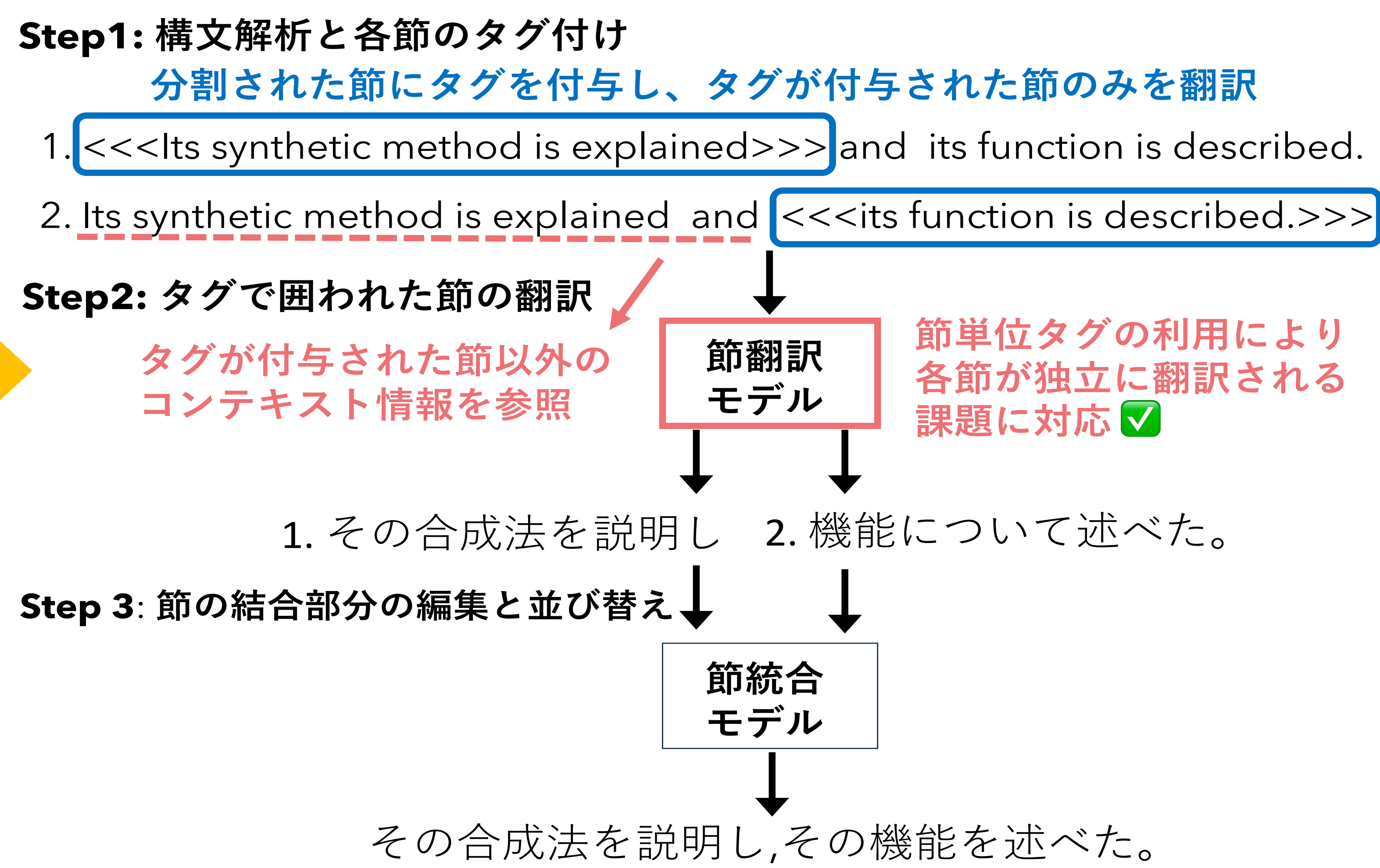
## 概要

- 分割統治NMT: 長文を節に分割して翻訳し、各節翻訳の結果をseq2seqモデルで整える手法
- 課題1: 分割により文全体のコンテキストの利用ができない → 節単位タグ利用によるコンテキスト補完
- 課題2: 節の翻訳に文レベルのNMTを利用している → 節単位の対訳データによる節翻訳モデルのFine-tuning

### 分割統治NMT

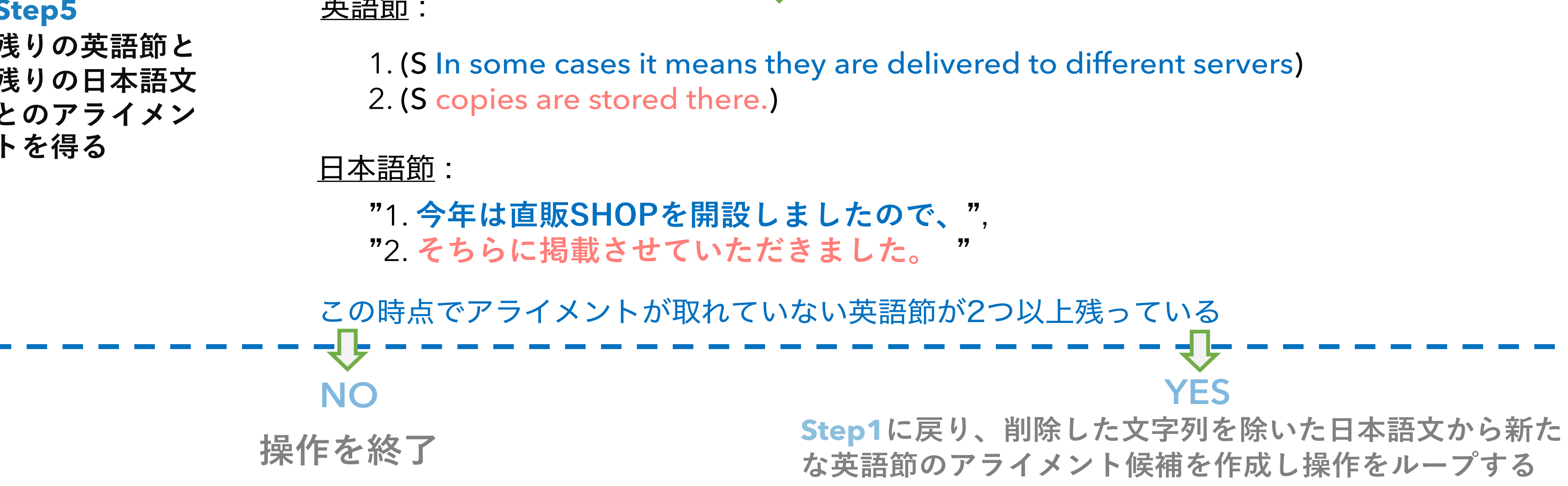
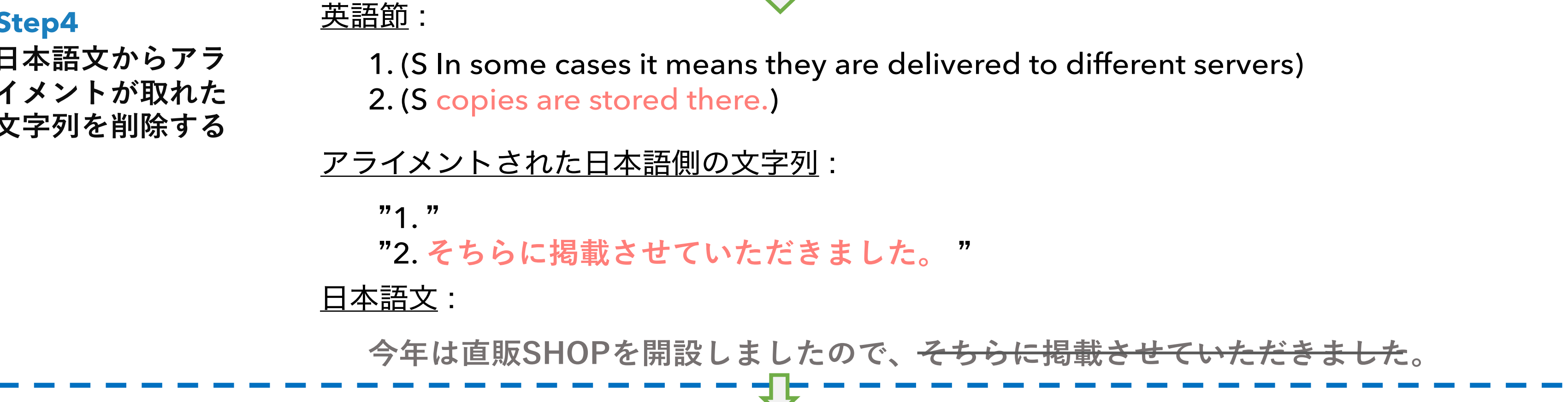
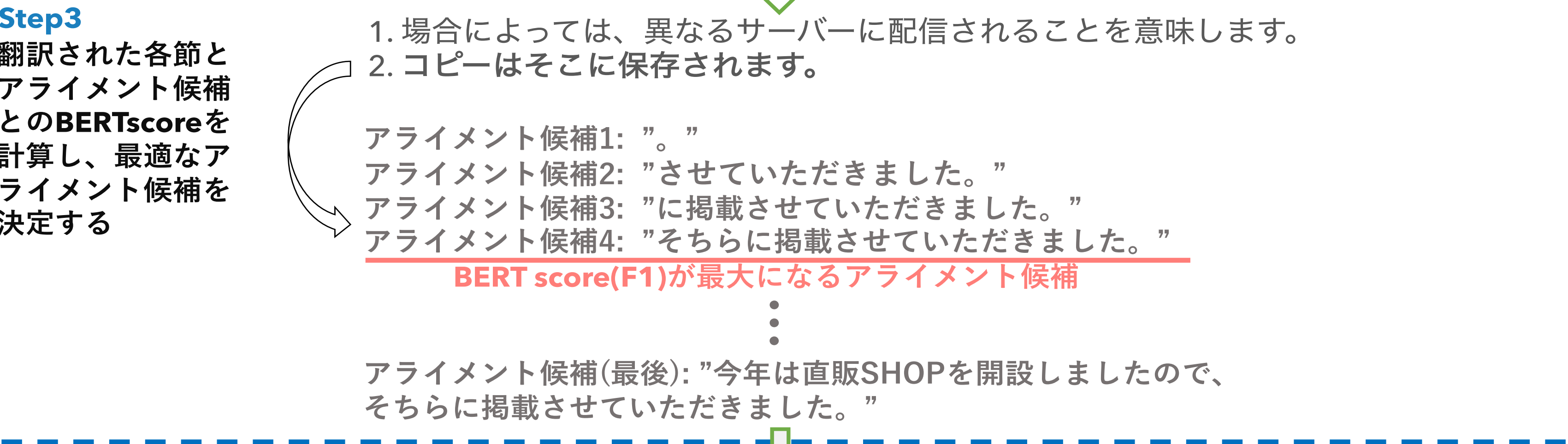
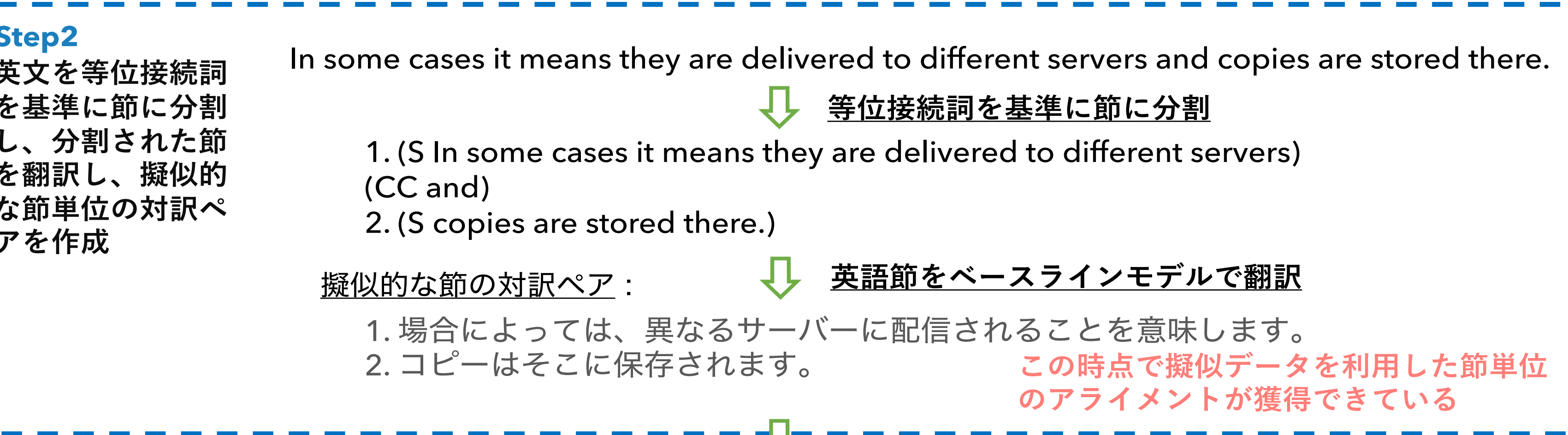
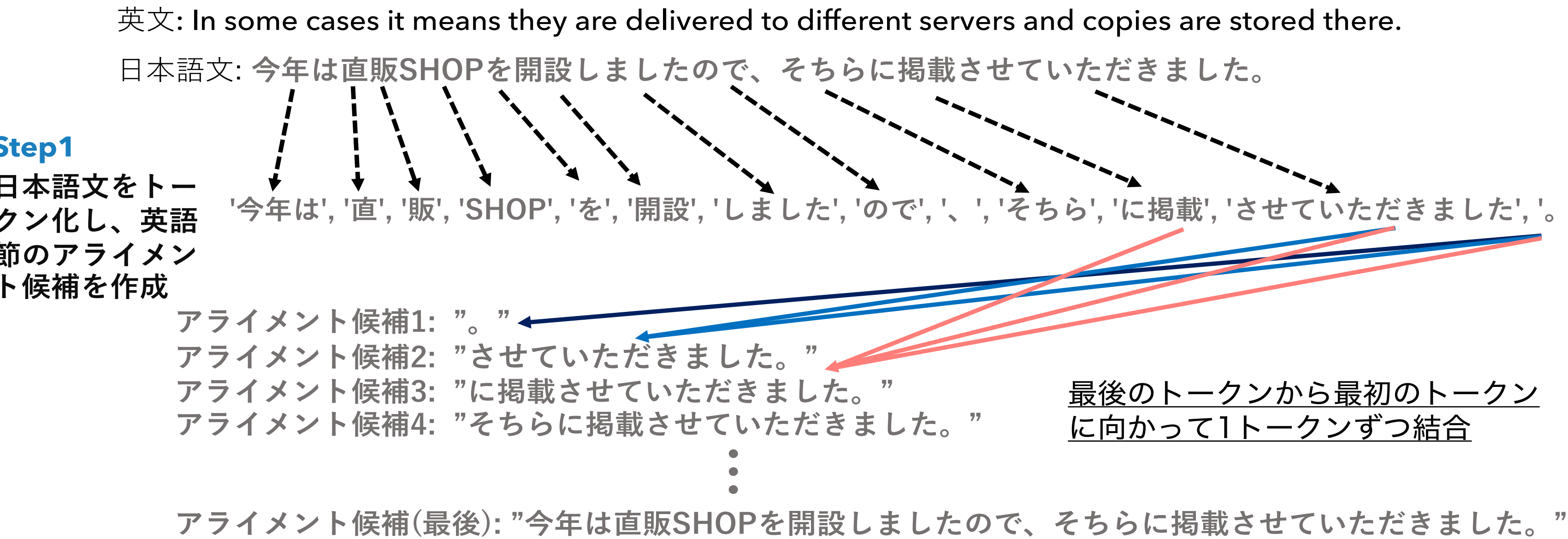


### 文内コンテキストを利用した分割統治NMT



## 節アライメント

### ■ 節アライメントの取得: 英日の対訳文ペアから分割された英語節と日本語文とのアライメントを取得する



## 結果と分析

### ■ 単語数41以上の長文の翻訳でBLEUが改善

提案手法が適応された文のBLEU

単語数	1-20	21-40	41-60	61-	全て
ベースライン	49.7	44.8	40.1	31.4	44.7
単純連結	51.0	44.6	40.8	34.2	44.8
提案手法	48.8	44.9	42.5	32.5	45.0

テストセット全体のBLEU

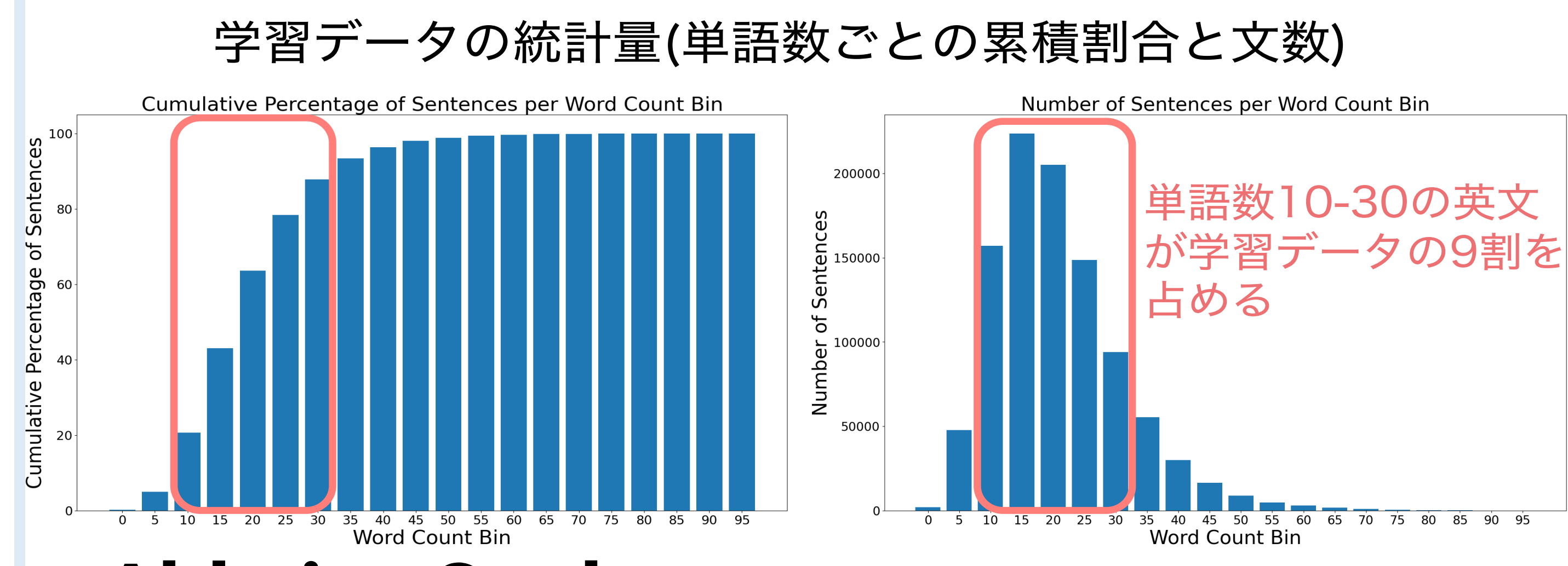
単語数	1-20	21-40	41-60	61-	全て
ベースライン	40.3	42.1	40.0	46.5	41.3
単純連結	40.4	42.1	40.2	47.5	41.3
提案手法	40.2	42.1	40.7	47.8	41.4

単語長ごとのテストデータの文数

単語数	1-20	21-40	41-60	61-	全て
全文	932	784	89	7	1812
分割された文	54	179	24	2	259

### ■ 節単位の翻訳によりモデルが翻訳しやすい単語数単位のテストデータの節と文の平均単語数

テストデータ	1-20	21-40	41-60	61-
節の平均単語数	8.6	13.4	21.6	24.4
文の平均単語数	14.8	27.7	46.8	68.1



### ■ Ablation Study

#### 文内コンテキストを利用することによるBLEUの改善

単語数	1-20	21-40	41-60	61-	全て
ベースライン	49.7	44.8	40.1	31.4	44.7
提案手法	48.8	44.9	42.5	32.5	45.0
w/o context	48.9	43.7	41.3	29.3	43.9