

ニューラル日英翻訳における出力文の態制御

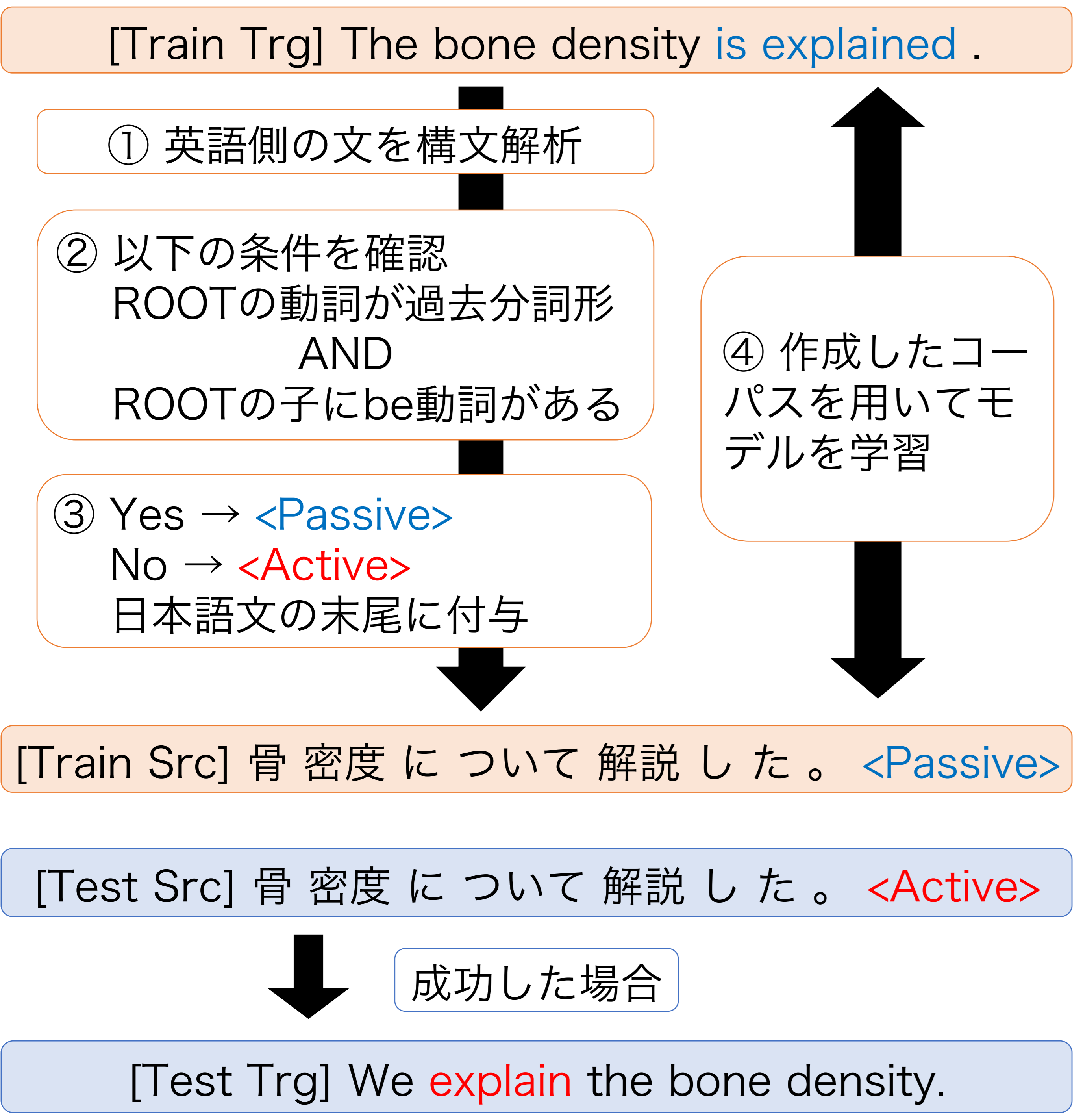
首都大学東京 山岸駿秀 (yamagishi-hayahide@ed.tmu.ac.jp) 叶内農 佐藤貴之 小町守

概要

機械翻訳では、言語間の表現方法の違いを意識しながら翻訳を行う必要がある。例えば、日英翻訳では日本語文と英文で態の使われ方に違いがある。Sennrichら (NAACL, 2016) は、英独翻訳でドイツ語の敬意表現を制御しながら翻訳を行った。本研究は、この先行研究を元に出力の態を制御しながら翻訳することを試みた。

主な高頻度動詞	能動態	受動態	合計
show	27,106	11,082	38,188
describe	16,338	18,043	34,381
develop	9,638	12,368	22,006
find	3,401	16,358	19,759
全動詞	604,158	499,178	1,103,336

実験方法



英語側の態の予測

英語側の態として自然なものを自動で選択したい

以下の6つの素性を1つのベクトルとして結合して用いた

1. 入力文の主語の分散表現
2. 入力文の述語の分散表現
3. 1つ前の入力文の主語の分散表現
4. 1つ前の出力文の目的語の分散表現
5. 3つ前までの出力文の態情報
6. 入力側の述語ごとに出力側の態の出現数をまとめ、多数派となった方の態の情報

2, 5, 6を用いたときが最高精度であり、67.7%であった

実験方法

ASPECのテストデータの先頭から、参照訳で自動詞を用いおらず、かつ態が各100文ずつになるように整形
Baseline: 態ラベルなしのデータで学習とテストを行う

態ラベルの付与方法として、以下の4パターンを実験

- ALL_ACTIVE:** 全文を能動態へ制御
- ALL_PASSIVE:** 全文を受動態へ制御
- REFERENCE:** 各文を参照訳と同じ態へ制御
- PREDICT:** 各文を予測した態へ制御

以下の3つの比較と、BLEUによる比較を行った

- 出力文の態と付与したラベルの表す態との比較
- 参照訳の態と付与したラベルの態との比較
- Pairwiseの人手評価 (BaselineとREFERENCEの2つのみ)

実験	能動態の文数	受動態の文数	その他の文数	参照訳と比較	ラベルと比較	BLEU	Pairwise
参照訳	100	100	-	-	-	-	-
Baseline	31	163	6	60.5%	-	20.60	40.0%
ALL_ACTIVE	147	44	9	57.5%	73.5%	20.22	-
ALL_PASSIVE	6	189	5	51.0%	94.5%	20.18	-
REFERENCE	82	113	5	89.5%	-	22.47	60.0%
PREDICT	74	118	8	64.0%	89.0%	21.06	-

実験設定

- コーパス: ASPEC (科学技術論文の概要をまとめたもの)
- 300万文のうち50単語以上の文を削除した上で、削除のなかった論文のみ収集 → 1,103,336文で学習
 - 論文の平均文数: 3.35

Bahdanauら (ICLR, 2015) のAttention-based NMTを学習させた
 隠れ層、埋め込み層の次元数: 512、語彙数: 30,000
 Optimizer: Adagrad (LR: 0.01)、batch size: 64、15 epoch

考察

1. 受動文の生成は容易に行うことができる。
2. 学習データ内に能動文の例が少ない動詞を能動文に用いる場合は、主語として何を生成すべきかがわからず失敗する。
3. 主語と目的語を入れ替える代わりに、仮主語を用いるなどによって対応していた文もある。
4. 入れ替えの起こらないものも多かった。
5. 人手評価の結果も、自然な態を選択することができれば良くなると言える
6. 予測では過去の文の分散表現は悪影響を与えた。

実験	参照役の態が能動態のときの成功例	参照訳の態が受動態のときの成功例	失敗例	態は制御できたが、主語と目的語の位置が同じ
入力文	最後に、この分野における最近の動向と問題点について述べた。	リサイクルに関する最近の話題を紹介した。	Siドットの密度及び直径分布は生成温度により変化させることができた。	自己組織化構造に分子の形と分子間相互作用が大きく影響する。
参照訳	Finally, this paper describes the recent trend and problems in this field.	Recent topics on recycling are introduced .	the density and diameter distribution of Si dots could be changed through the production temperature.	molecular shape and intermolecular interaction influence self-assembled structures greatly.
能動態へ制御	Finally, this paper describes the recent trends and problems in this field.	This paper introduces recent topics on the recycling.	density and diameter distribution of Si dots could be changed through the production temperature.	the molecular structure and molecular interaction greatly affect the self-organization structure.
受動態へ制御	Finally, recent trend and problems in this field are described .	Recent topics on the recycling are introduced .	density and diameter distribution of Si dots could be changed through the production temperature.	the molecular structure and molecular interaction are greatly affected by the self-organization structure.