

学習者コーパスを用いた日本人英語誤り訂正

水本智也 坂口慶祐 (奈良先端科学技術大学院大学)

背景

- 日本人の英語学習ニーズが増えている
 - グローバリゼーション
 - 小学校での英語導入
 - 英語の社内公用語化
 - iKnow!, Lang-8などの英語学習サービス

Izumi et al., 2003: Automatic error detection in the Japanese learners' English spoken data. *ACL 2003*.
Izumi et al., 2004: SST speech corpus of Japanese learners' English and automatic detection of learners' errors. *International Computer Archive of Modern English Journal*, 28.

- 日本人英語誤り訂正に関する既存研究
 - 冠詞の誤り訂正: Izumi et al., 2003
 - 冠詞の誤り検出: Izumi et al., 2004
 - 名詞(単/複)の誤り検出: Nagata et al., 2006a
 - 冠詞選択: Nagata et al., 2006b

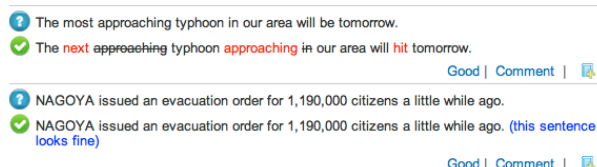
Nagata et al., 2006a: A feedback-augmented method for detecting errors in the writing of learners of English. *ACL2006*.
Nagata et al., 2006b: Recognizing article errors using prepositional information. *Systems and Computers in Japan*, 37.

学習者コーパス

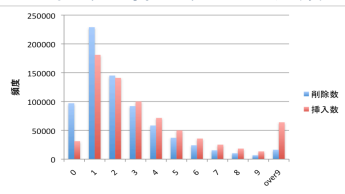
Lang-8*(語学学習SNS)コーパス

*<http://lang8.com>

- 2010年12月までの添削情報付きデータをクローリング
 - 添削付き英語の文数は約73万文(うち約7割の文が日本人学習者)



削除数/挿入数による分類



- 添削だけでなくコメント文も↑

- 削除挿入数の頻度は異なる傾向
 - 日本語の場合と同じ傾向 (Mizumoto et al. 2011)

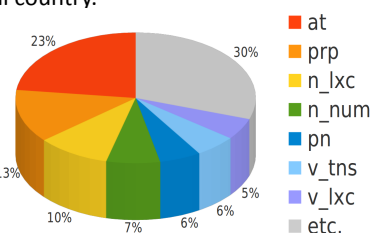
KJ (Konan-JIEM) Learner Corpus

- 日本人大学生によるエッセイ
- 文法誤り*に対するアノテーション
- 分析に使用したデータ*
- 全306文(誤りあり: 164文 53.6%)

誤りの傾向

- 冠詞
 - Australia is */a very beautiful country.
- 前置詞
 - I often go */to/ there.
- 名詞の語彙選択
 - *moving tool/transportation

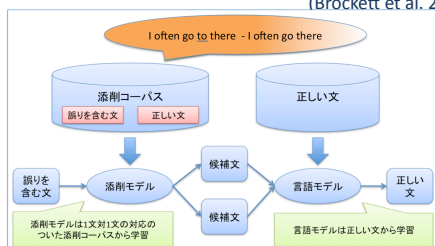
*誤りの中にはスペリングエラーは含まれない。
*分析に使用したKJコーパスは、KJコーパスの一部である



SMTを使った誤り訂正

統計的機械翻訳(SMT)を使った誤り訂正

(Brockett et al. 2006, Mizumoto et al. 2011)

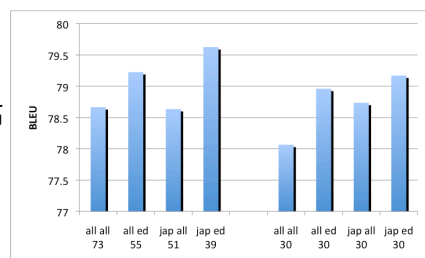


- TM用いるLang-8のデータを8種類に変えて実験
 - 全ての学習者の全ての文 (73万文) [all all 73]
 - 全ての学習者の文を編集距離で足切りしたもの (55万文) [all ed 55]
 - 日本人学習者の全ての文 (51万文) [jap all 51]
 - 日本人学習者の文を編集距離で足切りしたもの (39万文) [jap ed 39]
- 上記4つのデータで使用する文数を30万文に固定したもの
- テストデータにKJコーパス306文を使用
 - デベロップメントに206文 テストに100文使用

結果と考察

評価

- BLEUを使用
- 日本人の文で編集距離を利用が最もBLEU高
- 編集距離の効果有
- 日本人学習者にのみの文に限るのも効果有



考察

- a lot of kinds of game → a lot of kinds of games
- many kind of → many kinds of
- △* for exsample → for example *KJコーパスはスペルエラーを含め
- △* after shcool → after school ていないため
- × I don't success → I don't succeed 正: I don't succeed
- × went to mountain → went to the mountain 正: a mountain
- ・文法誤りよりはスペル誤りに対して訂正可能

まとめ

- 日本人の英語学習のニーズ増
 - 小学生での英語導入
 - Lang-8などのWebサービス

- 日本人学習者の誤りの3割は冠詞、前置詞
- SMTで誤り訂正 → スペル誤りは比較的訂正可能