

Dependency Types in Learner English and Authentic English

Masanori Oya

Mejiro University

m.oya@mejiro.ac.jp

Abstract

This study applies the framework of Dependency Grammar to characterize learner English, with respect to the dependency types found in learner English and authentic English.

Introduction

Sentences randomly chosen from two different corpora (abstracts of academic journals and English essays written by Japanese learners of English) were parsed by Stanford Parser, the numbers of the dependency types in the parse output were counted, and the type per sentence (TPS) for each dependency type is calculated to see how the difference of genre is reflected on the difference of the TPSs of a certain dependency type. The TPSs of some dependency types are divergent across different corpora.

Keywords

Learner English, Dependency Grammar, Corpus, Type per sentence

1 Theoretical background

1.1 Dependency Grammar and Type per sentence (TPS)

In Dependency-Grammar formalism, every word in a sentence depends on another word in the same sentence, and the dependency

relations among words are all given a certain type (e.g., the type "subject" for the dependency between "David" and "is" in a sentence "David is reading a book in his room."). The assumption in this study is that we can find certain differences between corpora of different genres, e.g., learner English and authentic English, in terms of Dependency-Grammar formalism; in particular, certain dependency types may be found more frequently in authentic English than in learner English, or vice versa.

With this assumption in mind, this study introduces a metric called type per sentence (TPS). TPS is defined below; for a given dependency type, TPS is the number of this dependency type in a corpus (DT) divided by the number of sentences in the same corpus (S).

$$TPS = DT / S$$

TPS is expected to show the characteristics of a given corpus in terms of the dependency types used in the sentences that the corpus contains.

2 Research

2.1 Purpose

The purpose of this study is to show how well TPSs can show the characteristics of given corpora by calculating and comparing the TPSs of all the dependency types used in two

different corpora.

2.2 Method

This study uses Stanford Parser (Chen and Manning 2014; de Marneffe, MacCartney and Manning 2006) to obtain the dependency trees for English sentences. Then, the TPS of each dependency type is calculated and compared across the two corpora.

2.3 Data: Corpora used in this study

2.3.1 ICLE

The International Corpus of Learner English ver.2 (Granger et al. 2009) is a corpus of essays written by learners of English from 16 different mother tongues. In this study, 881 sentences were randomly chosen from essays written by Japanese learners of English (henceforth *ICLE*), and were parsed by Stanford Parser.

2.3.2 Abstracts of academic journals

In this study, 881 sentences from the abstracts of seven academic journals (henceforth *Journal*) were parsed by Stanford Parser. The journals used in this study are as follows: *Applied Linguistics*, *International Journal of Corpus Linguistics*, *ELT Journal*, *Journal of Linguistics*, *Language Testing*, *Linguistic Inquiry*, and *Studies in Second Language Acquisition*.

3 Results

The TPSs of some dependency types are found to be larger in *Journal* than those in *ICLE*; for example, the TPS of noun

compound (e.g., the dependency type between *language* and *acquisition* in the phrase *second language acquisition*) in *Journal* is 1.68 and that in *ICLE* is 0.42. The TPSs of other dependency types and their divergence will be shown in the conference.

4. Conclusion

This study introduced the framework of Dependency Grammar and applies it to characterize learner English, with respect to the dependency types found in learner English and authentic English. The metric called TPS was introduced, and some TPSs of certain dependency types are shown divergent across different genres.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 26370500.

Reference

- Chen, D. & Manning, C. (2014). A Fast and Accurate Dependency Parser using Neural Networks. *Proceedings of EMNLP 2014*.
- De Marneffe, M.C., MacCartney, B. & Manning, C. (2006). Generating Typed Dependency Parses from Phrase Structure Parses. *Proceedings of LREC 2006*.
- Granger, S., Dagneaux, E., Meunier, F. & Paquot, M. (2009). *International Corpus of Learner English v2*. Presses universitaires de Louvain, Louvain-la-Neuve.