

Syntactic variations in Chinese–English code-switching

Lin Wang^a, Haitao Liu^{b,*}

^a Centre for the Study of Language and Cognition, Zhejiang University, China

^b School of International Studies, Zhejiang University, China

Received 7 December 2011; received in revised form 3 October 2012; accepted 4 October 2012

Available online 17 November 2012

Abstract

Based on a Chinese–English code-mixed treebank, this paper reports the probable syntactic consequences of code-switching. Compared with monolingual Chinese and English corpora, in the mixed corpus there are syntactic variations: variation in dependency distances and word-order variation in dependency direction. In the mixed corpus, there are two types of dependencies: monolingual and mixed dependencies. Mixed dependencies present longer dependency distances than monolingual ones. Major grammatical relations (subject, object, attribute and adverbial) and certain properties of code-switching (peripherality, flagging and dislocation) contribute to the variability of dependency distances. It is the distributions of major grammatical relations with different dependency directions in monolingual and mixed dependencies that cause the word-order variation.

© 2012 Elsevier B.V. All rights reserved.

Keywords: Code-switching; Syntactic variation; Chinese; English; Treebank; Dependency distance; Word-order

1. Introduction

Code-switching is one of the most studied topics in language contact phenomena (Myers-Scotton, 2006). Code-switching involves both the linguistic behavior of producing and comprehending language. Early in 1952, Jakobson et al. proposed that code-switching would make language processing more difficult: “Obviously such a task of deciphering becoming more difficult in the frequent cases called ‘switching code’ by communication engineers or ‘coexistent phonemic systems’ by linguists” (1952:603–604). Placing ourselves in the variationist framework (Sankoff and Labov, 1979; Poplack, 1993), this paper makes a quantitative study of syntactic variation of Chinese–English code-switching based on naturally occurring corpora and tests the alleged processing cost of code-switching.

Code-switching can be defined as “language use that consists of material from two or more language varieties at any level from the discourse to the clause” (Jake and Myers-Scotton, 2009:207). The phenomenon studied in this paper is intra-clause code-switching or code-switching in a single clause (Myers-Scotton, 2006). In recent decades, bilingual code-switching studies have been done in different parts of the world, including America, Europe, Africa, Asia and the Middle East (Chan, 2009). It has been widely discussed from the perspective of general linguistics, sociolinguistics and psycholinguistics. Using the syntactic approach, researchers have explored some constraints and proposed models for code-switching, such as the Free Morpheme and Equivalence Constraints (Sankoff and Poplack, 1981), the Phrase-structure Congruence Constraint (Woolford, 1983), the Functional Head Constraint and the Word Order Integrity Corollary (Belazi et al., 1994), and the Matrix Language Frame Model (Myers-Scotton, 2002). From the sociolinguistic perspective,

* Corresponding author at: School of International Studies, Zhejiang University, No. 866 Yuhangtang Road, 310058 Hangzhou, China.

E-mail address: lhtzju@gmail.com (H. Liu).

theories and models have been built to explain conversational choices, such as Gumperz's (1982) notion of contextualization cues, Giles and his associates' (1991) Accommodation Theory, and Myers-Scotton's (1983) Markedness Model. With regard to psycholinguistic code-switching studies, using experimental findings, researchers have been interested in a number of issues, including code-switching, the brain and aphasia (Hyltenstam, 1995; Myers-Scotton and Jake, 1995, 2000; Kutas et al., 2009) as well as code-switching and the mental lexicon (Wei, 2009). However, so far, there are very few studies that provide a quantitative syntactic and cognitive analysis and that test the processing cost of code-switching based on authentic language data.

In the grammatical approach, major models and constraints have been proposed to account for the structure of code-switching. However, the long-standing problem is the diversity and variation of code-switching patterns which appear to defy any economical and universal syntactic account (Chan, 2009). As Hudson indicated, "the positive side of the structuralist approach was its focusing of attention on the complex internal structure of language" (2010:107) and its negative side was to encourage linguists "to ignore everything outside language" (2010:107). In the 1980s, linguists including Hudson started to develop theories of language structure building on cognitive psychology and artificial intelligence in the "movement" of Cognitive Linguistics (Geeraerts and Cuyckens, 2010), within which Word Grammar (Hudson, 1995) was proposed (Hudson, 2010). When summarizing Word Grammar, Hudson (2010) indicated that "the best possible theoretical explanation for the organization of language is to show that it's typical of the organization of knowledge in general" (Hudson, 2010:108). As a result, Word Grammar is an effective theory to analyze the language both structurally and cognitively. That is the main reason why our study is based on Word Grammar, because it displays relatively clearly the research approach of finding the human cognitive mechanism by analyzing the language structure. Studies in Word Grammar (Hudson, 1995; Hiranuma, 1998, 1999) and cognitive science (Gibson, 1998, 2000) reveal that both syntactic complexity and language processing cost can be measured by dependency distance or the linear distance between a word and its head in the sentence (hereafter DD) which is an important property of language dependency structure (Temperley, 2007; Liu, 2008; Gildea and Temperley, 2010). In other words, dependency distance, which reflects syntactic complexity, can be a measure for the cognitive difficulty of human language production and comprehension. The mean dependency distance (hereafter MDD) of the sentence can be used as a measure for parsing difficulty as it has been shown for subject-extracted vs. object-extracted structures, center-embedded vs. right-dependent sentences, and Garden Path sentences (Hsiao and Gibson, 2003; Liu, 2008). Apart from sentence parsing complexity, mean dependency distance has been used to test the processing difficulty of different languages. Liu (2008) has found that different languages prefer to have different MDD, and Chinese has the longest MDD in 20 language corpora, which is assumed to make Chinese more difficult to process.

Since MDD is a measure of the complexity of syntactic structure as well as parsing difficulty independent of the language, it is expected that by measuring dependency distances, we can find out whether there exists syntactic variation in dependency distance and test the processing cost of code-switching. Specifically speaking, we answer the following questions: (1) Compared with monolingual corpora, is there syntactic variation in dependency distance in a mixed corpus? (2) If it does occur, in the code-switching corpus, there exist two types of dependencies – monolingual and mixed dependencies: "syntactic relations in which words A and B are from different languages" (Eppler, 2011:146), do mixed ones have longer or shorter MDDs? Regarding the latter question, Eppler (2011) did an insightful study based on a German–English monolingual and code-mixed discourse corpus and proposed the Distance Hypothesis (hereafter DH): "code-switching is more likely in syntactic relations with long dependency distance" (2011:145). In other words, mixed dependencies have longer dependency distances than monolingual ones, and cognitively, DH (Eppler, 2011) can verify Jakobson's prediction. (3) Eppler's study is on two typologically closely related languages, German and English. Can we draw similar conclusions from Chinese and English, two typologically distant languages?

Syntactic variation is also involved in word-order variation in terms of dependency direction. The dependency direction can be defined as "head-initial" when the dependent follows the head and "head-final" when it precedes the head in the linear order of the sentence (Liu, 2010). (4) According to Liu (2010), the dependency direction can be used as a means of word-order typology, by calculating and comparing the distribution of the dependency direction in mixed and monolingual corpora, can we also find whether there is word-order variation? (5) If DH (Eppler, 2011) can be supported, what types of grammatical relations play a major role in causing this variation?

Based on a Chinese–English mixed corpus with 19,766 tokens, the present paper proposes a quantitative empirical method based on a treebank (namely, a corpus with syntactic annotation) and tries to answer the above-mentioned questions. In the next section, we will present the method used in our study, that is, the method of measuring the dependency distance and dependency direction.

2. Method and material

In the present paper, the syntactic analysis is based on a dependency grammar – Word Grammar, one approach of Cognitive Linguistics (Hudson, 2010). This syntactic approach is preferable for our study because "it rests on firmer

Download English Version:

<https://daneshyari.com/en/article/935751>

Download Persian Version:

<https://daneshyari.com/article/935751>

[Daneshyari.com](https://daneshyari.com)