Language Sciences 67 (2018) 33–45

Contents lists available at ScienceDirect

Language Sciences

journal homepage: www.elsevier.com/locate/langsci

What factors are associated with dependency distances to ensure easy comprehension? A case study of *ba* sentences in Mandarin Chinese

Yu Fang [a], Haitao Liu [a,b,c,*]

[a] Department of Linguistics, Zhejiang University, Hangzhou, 310058, China
[b] Ningbo Institute of Technology, Zhejiang University, Ningbo, 315100, China
[c] Centre for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, 510420, China

ARTICLE INFO

Article history:
Received 28 November 2017
Received in revised form 19 March 2018
Accepted 11 April 2018

Keywords:
Givenness
Word frequency
Adverbial length
Long-distance dependency
Comprehension difficulty

ABSTRACT

Dependency distance, referring to the linear distance between two syntactically related words, is often minimized, because long-distance dependencies (LDDs) may cause processing difficulties. However, LDDs can be found in some sentences, for example, Chinese *ba* sentences, because besides LDDs, other factors also influence the comprehension difficulty. Based on *ba* sentences extracted from three genres: interviews, essays and research papers, this study investigated the impact of givenness, word frequencies and adverbial lengths on dependency distances. The results show that: 1) givenness of subjects and NP2s in *ba* sentences affects dependency distances in all three genres, and the effect is consistent among genres; 2) word frequencies of subjects in interviews also affect dependency distances, but such an impact has not been found in essays and research papers; 3) adverbial lengths and NP2 lengths do not counterbalance each other to limit dependency distances. These findings suggest that: 1) comprehension difficulty is influenced by multiple factors; 2) the impact of the three factors on dependency distances is not influenced by genres.

1. Introduction

In Mandarin Chinese, *ba* sentences, with the structure of NP1+*ba*+NP2+VP, are quite conspicuous and intriguing because Chinese is basically an SVO language (Mcdaniel et al., 2015; Xu and Liu, 2015). A typical *ba* sentence is shown in example (1):

(1)	wo	ba	pingguo	chi	le
	I	BA	apple	eat	perfective
	'I have eaten the apple.'				

In this sentence, *wo* (NP1) is the subject, *pingguo* (NP2) is a preposed object marked by the preposition *ba*, and *chi* and *le* (VP) constitute the predicate verb, thus the *ba* sentence takes the SOV order. This sentence can be converted into an SVO

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

https://doi.org/10.1016/j.langsci.2018.04.005
0388-0001/

sentence (*wo chi le pingguo*), which expresses nearly the same meaning. Actually, in most cases, *ba* sentences can be converted into SVO sentences (Sun, 1995; L. L. Wang, 2013); however, the speaker or the writer chooses a *ba* sentence instead of an SVO sentence under some circumstances, which is due to the semantic, syntactic and pragmatic differences between them (P. Liu and Zhao, 2005). For instance, a *ba* sentence can convey a "disposal" meaning but not a SVO sentence (L. Wang, 1948). That is, the NP2 *pingguo* in (1) is affected by the action of *chi* (in most case the apple has been eaten up), effectively excluding the second reading found with the corresponding SVO sentence. Moreover, the focus of a *ba* sentence is on the verb, while the focus of a SVO sentence is on the object (Ho, 1993). Obviously, *ba* and NP2s, as intervening components, increase syntactic distances between subjects and predicate verbs. Moreover, adverbials often separate subjects from predicate verbs (Xu and Liu, 2015), further lengthening their syntactic distances. Such long syntactic distances may cause language comprehension difficulty (Gibson, 1998, 2000; H. Liu, 2008) and are not expected in sentence structuring. However, *ba* sentences, a frequently used structure in Mandarin Chinese, are not always more difficult to understand than SVO sentences. Therefore, it is important to explore what factors are associated with long syntactic distances in *ba* sentences so as to ensure easy comprehension, which is the aim of this study.

Syntactic distance, in the framework of dependency grammar, is known as dependency distance, referring to the distance between a dependent and its governor in a dependency relation. Dependency distance can be measured by the number of intervening words (Gibson, 2000), or by subtracting the position numbers of a governor and a dependent (H. Liu et al., 2009). For the first measurement, the dependency distance between two adjacent words is 0, and the dependency distance between two words separated by one intervening word is 1, etc. In contrast, for the second measurement, the dependency distance between two adjacent words is 1, and the dependency distance between two words separated by one intervening word is 2, etc. In this study, we adopt the second measurement. The dependency structure of example (1) is as follows:

In Fig. 1, arrowed arcs show dependency relations, pointing from a governor to its dependent. For instance, the dependency distance between *wo* (*I*) and *chi* (*eat*) is 4-1 = 3, and the dependency distance between *ba* and *chi* (*eat*) is 4-2 = 2.

Fig. 1. Dependency structure of example (1).

Researchers have proposed dependency distance as a metric of comprehension difficulty: the longer a dependency distance is, the harder this dependency is to establish (H. Liu, 2008; Scontras et al., 2015). Such a claim can be explained by the nature of the language comprehension mechanism. Language comprehension is implemented on the platform of working memory to do syntactic parsing: listeners or readers need to establish the syntactic dependency between a dependent and its governor. To be specific, during the comprehension process, a word coming into the parser looks for its governor and dependents among all words having appeared so far. Again as for the *ba* sentence in Fig. 1, when the word *wo* (*I*) occurs, it looks for its governor, but no dependency relation can be established at that time. Thus this new word must be kept in memory to be retrieved later. Retrieval difficulty of this word increases with the dependency distance due to the limited working memory of human beings (Gibson, 1998, 2000; Hawkins, 1994, 1998; Nicenboim et al., 2015).

The relation between comprehension difficulty and dependency distance is known as a locality effect (Gibson, 2000), which has been observed in many experiments using a variety of language materials (Gibson et al., 2013; Gibson and Wu, 2013; Levy et al., 2013; Rispens and de Amesti, 2017; Safavi et al., 2016). In this sense, for the sake of easy comprehension for listeners or readers, the speaker or the writer prefers local dependencies in sentence structuring (Fedorenko et al., 2013). Such a preference for short distance dependencies is known as *dependency distance minimization* (DDM), which has been claimed to be a linguistic universal of natural languages (H. Liu et al., 2016; Lu et al., 2016). The tendency toward DDM is verified in two large-scale cross-language studies, one including 20 languages (H. Liu, 2008), and the other including 37 languages (Futrell et al., 2015).

Nevertheless, DDM is not the only guiding principle of sentence structuring. Besides easy comprehension, other factors such as special communicative purposes, are also considered in language production (H. Liu et al., 2017; Wray, 2017), which may bring about long-distance dependencies (LDDs) in natural languages (Dąbrowska et al., 2009; Fedorenko et al., 2013; Misyak et al., 2010; Xu and Liu, 2015). Sentences with LDDs are not always difficult to understand; the following *ba* sentence provides an example.

(2)	wo	yao	ba	zhe	baogui	de	dongxi	zengsong	gei	shishang	kelian	de	ren
	I	will	BA	this	valuable	of	stuff	give	to	world	poor	of	people
	'I will give this valuable stuff to the poor in the world.'												

In the example (2), the dependency distance between the subject *wo* (*I*) and the predicate verb *zengsong* (*give*) is 8-1 = 7. This *ba* sentence can be converted into an SVO sentence, then the dependency distance between *wo* and *zengsong* is shortened to 3-1 = 2. Therefore, the dependency distance in the *ba* sentence is much longer than that in the SVO sentence. But