Predictions of Entropy Reduction Theory on Chinese Relative Clauses*

Borui Zhang
University of Minnesota, Twin Cities

Abstract
Processing difficulty in Chinese has been a challenge for theories of sentence parsing due to its lacking morphology and structural ambiguity. I use entropy reduction (ER) theory, based on a minimalist analysis of relative clauses (RCs), and tree-banks-based probabilistic grammar to make predictions on Chinese RCs (with various amounts of temporary ambiguity). The predictions match the results from the previous human experiments on Chinese. This provides supporting evidence to the reliability of ER theory and highlight the ability of the theory to deal with ambiguous sentences with the right grammar and analysis.

Key words
entropy, entropy reduction, Chinese, relative clauses

* I would like to thank my MA advisor Professor Tim Hunter for providing the research materials, his expert advice, and encouragement throughout my study. I would like to thank Professor Zhong Chen and Professor John Hale for sharing their experimental knowledge, and Professor Brian Reese and Professor Sashank Varma for their supports of the study. All errors remain mine.

1. Introduction

Many experiments (Hsiao and Gibson's 2003, Jäger et al. 2015, King and Just 1991, and Traxler et al. 2002) measuring the difficulty of native speakers' perception of reading relative clauses cross-linguistically have been developed. Meanwhile, linguistic research has been suggesting various theories aimed at making the right prediction to match the experimental results. Hsiao and Gibson (2003) and Jäger et al. (2015) have done research on measuring Chinese subject relative clauses (SRC) and object relative clauses (ORC) with different sentence models, and their studies found inverse results: the former study found that ORCs are easier for native speakers to process than SRCs, while the latter one found that SRCs are easier. This is a puzzling phenomenon, presenting difficulties for generalizing a good theory to predict the perception difficulty of Chinese RCs. This may be due to the complex structural ambiguities of Chinese, and the differences in the alternative structures that a parser needs to consider during the proception process. Chinese is also one of the languages in which it is possible to produce ambiguous meanings while forming an RC. I suggest that the ER theory may be able to explain this puzzle, making the right predictions of the perception difficulty on Chinese RCs.

1.1 Relative Clauses

Despite the variations of the details, an RC generally contains a TP and modifies a head noun. More importantly, an RC can have linguistic gaps in it, and there are two kinds of RCs depending on where a gap is located. When a gap occurs at the subject position, it is an SRC, as in the English example shown below in (1a). The DP phrase "the girl" gets moved out of the CP of the RC and leaves a trace (shown with co-indexing) at the subject position where it gets moved from[1]. Similarly, when a gap occurs in the object position in a relative clause (a CP), it is an ORC, as in the example shown in (1b). Here, "the girl" gets moved from the object position in the clause and leaves a trace there. (1a) and (1b) are a minimal pair of English SRC and ORC, which only differ from each other in the positions where the gap occurs while the rest of the sentence remains the same. The sentences also show that the RC comes after the head noun, so English RCs are post-nominal. There are also some languages where RCs precede the head noun, which are so called pre-nominal languages, and Chinese is one of them.

(1) a. The girl$_i$ [CP who t$_i$ met John] left. (SRC)
b. The girl$_i$ [CP who John met t$_i$] left. (ORC)

1.2 Post-nominal RCs and Pre-nominal RCs

The reading-time experiments in King and Just (1991) and eye-movement experiments on SRCs and ORCs in Traxler et al. (2002) suggest that English ORCs require a longer time for native speakers to comprehend than SRCs. In other words, SRCs are easier than ORCs. Therefore, (1a) should take less time to understand than (1b) according to the studies. In addition, similar experiments on other languages which have post-nominal RCs, such as French, German, and Dutch, also show the same results (Yun et al. 2015). So far, in support of these findings, SRCs seem easier to read than ORCs in the post-nominal-RC languages.

To determine whether SRCs are easier than ORCs cross-linguistically, pre-nominal-RC languages also needed to be tested. Chinese, Japanese, and Korean are a few such languages which have been tested. The experiments report that SRCs are easier than ORCs in both Japanese and Korean. But Hsiao and Gibson's study (2003) found that in Chinese ORCs are easier than SRCs. This could potentially falsify the idea that SRCs are easier than ORCs in both post-nominal-RC

[1] There are several movement analyses. This paper uses promotion analysis following Yun et al. (2015)

and pre-nominal-RC languages. However, another experiment by Jäger et al. (2015) using a different test sentence structure from Hsiao and Gibson's shows that Chinese SRCs are easier than ORCs. Their study claims that the sentences they used in the experiment are not temporarily ambiguous, whereas the sentences in Hsiao and Gibson's are, and that this difference could explain the different findings about SRCs and ORCs. "Temporary ambiguity" in this paper refers to the ambiguity between RC and non-RC structures of the words that have already been heard in a sentence. Section 3 will discuss this issue in detail.

2. Current Theories and Predictions

Several theories based on memory burden (Wanner and Maratsos 1978, Gibson 2000, O'grady, 2007) have been raised in an attempt to explain the differences between SRCs and ORCs in terms of reading difficulty cross-linguistically. In this section, I will discuss the following three theories: linear distance theory, hierarchical distance theory, and ER theory, as shown in Table 1. (The labels "SRC" and "ORC" mean that that kind of RC is tested and predicted to be the easier one to process in a language.)

Table 1. Prediction of theories on Relative Clauses

Language	Human Experiment Results		Linear Distance	Hierarchical Distance	Entropy Reduction	
English	SRC		SRC	SRC	SRC	
Japanese	SRC		ORC	SRC	SRC	
Korean	SRC		ORC	SRC	SRC	
Chinese	SRC (Jäger et al. 2015)	ORC (Hsiao & Gibson 2003)	ORC	SRC	SRC (Yun et. al. 2015)	ORC (See details in Section 4)

First, linear distance theory (Wanner and Maratsos 1978, Gibson 2000) predicts that ORCs are harder in general because they require more working memory to comprehend. In other words, the linear distance (the number of words) between the head noun and the trace in an ORC is longer (larger) than that in an SRC. As the example shows in (2) in English ORC (2b) the distance between the head noun "the girl" is longer than that in (2a). Therefore, (2a) is easier than (2b). The theory seems to work for English since the literature suggest that SRCs are easier than ORCs in post-nominal languages.

(2) a. The girl$_i$ [CP who [TP [t$_i$ [met John]]] left. (SRC)
b. The girl$_i$ [CP who [TP [John [met t$_i$]]] left. (ORC)

However, this theory will not work for Japanese or Korean, because the distance between the head noun and the trace in the ORCs are shorter than those in the SRCs. Since the experiment results on Japanese and Korean show that SRCs are easier, this theory will not hold. The theory seems to work in Chinese based on Hsiao and Gibson's (2003) study, but the inverse result found in Jäger et al. (2015) does not support the theory. More discussion about this issue is in Section 3.

The second theory is the hierarchical distance theory (O'grady, 2007), which predicts that SRCs should always be easier than ORCs, because moving an object out of an RC always crosses one more layer, the VP layer, than moving a subject out of an RC. This theory seems to account for most of the language experiments we have discussed so far, but not for Chinese, where Hsiao and Gibson's (2003) find the inverse result.

2.1 Entropy Reduction Theory

The third theory introduced in Table 1, and the one I argue best accounts for the data is entropy reduction theory. In information theory (Shannon 2001, Hale 2016), entropy is a quantitative measurement for information that is gained when one learns the given (incomplete) information. Entropy is calculated with the formula below in (3) with the probabilities given of the expected elements *p(x)*. The entropy is the negative of the sum of the probabilities of possible x's multiply their binary logarithm.

$$(3) \quad H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

ER theory (Hale 2006) calculates the entropy and ER at each point for a given sentence incrementally from the first element to the last element. An initial entropy will be calculated before a sentence is read. The entropy will reduce at a point, if the number of equally possible sentence structures allowed by the grammar becomes less from that point; the entropy will increase if the number of equally possible sentence structures becomes more. The total entropy reduction of every point from the beginning to the end indicates the reading difficulty of a sentence.

If the total ER of sentence A is less than that of sentence B, then A is easier than B. Sentences are also recursive, so the number of possible sentences at one stage are infinitely great. Hale (2006) suggested a way of calculating entropy for the distribution of the infinite sentence probabilities. The initial entropy value of every sentence in one language is the same, and it is determined by the set of probabilistic grammar rules of this language. Figure 1 shows the entropy when the probabilistic grammar rules are assigned with an equal possibility (and sum is equal to 1).

Figure 1. Relationship between Entropy and Uniformed Probabilities

There are also other sensible ways of assigning possibilities. Hale's study (2003) counts the frequencies of the rules used in corpora, instead of assigning an equal probability. For example, as the English sentence *"The horse raced past the barn fell."* in Table 2, an ER is calculated at each step with adding a category and added up to get a numeric measure for the sentence - "total ER".

Table 2. Incremental procedure of ER Theory on English garden path sentence

Words	*the*	*horse*	*raced*	*past*	*the*	*Barn*	*fell*	
ER	0	1	0.123	0	0	0.123	3.82	Total ER: 5.066

This sentence is a garden path sentence, which usually is understood by readers only after reading the last word of the sentence. There are many possible structures the readers might postulate when they have only heard the first few words. Such uncertainty of the structure will be referred to as "temporary ambiguities" in this paper. The following sections will discuss Chinese RCs, which can have these kinds of ambiguities, and will show how ER theory makes better predictions about Chinese RCs than the previous theories.

3. Processing with Two Kinds of Models

3.1 Temporarily Ambiguous Chinese RCs

Temporary ambiguity is a structural ambiguity that appears before the end of the sentence, which means that the ambiguity disappears somewhere along the reading process. For example, the word orders in (4a) and (4b) both contain an RC (a CP clause modifying the matrix subject), and are temporary ambiguous, but not fully ambiguous. The temporary ambiguities occur during the processing within the CP region.

(4) a. [$_{CP}$ t_i Verb Noun de] $Noun_i$ VP. (SRC)
b. [$_{CP}$ Noun Verb t_i de] $Noun_i$ VP. (ORC)

The examples below in (5) and (6) show the alternative structures of the ambiguous partial sentences, respectively to (4a) and (4b). The alternative non-RC structures can be expected before the head nouns of an SRC and an ORC is read in (5) and (6). See Yun et al. (2015) for detailed explanations.

(5) a. pro Verb Noun. (simple pro-drop)
b. proverb [$_{POSS\text{-}NP}$ Noun de Noun]. (pro-drop with poss in object)

(6) a. Noun Verb Noun. (simple sentence)
b. Noun Verb [$_{POSS\text{-}NP}$ Noun de Noun]. (simple with poss in object)

3.2 Non-Temporarily ambiguous Chinese RCs

Jäger et al. (2015) suggested another model comparing to (4), as in the sequences in Table 3, that is considered non-temporarily-ambiguous. This model does not lead other temporal ambiguous readings, as from (4a) to (5), and from (4b) to (6), but only leads to a "pure" SRC versus ORC comparison. Their experiment shows that SRCs are easier than ORCs, which is an inverse result from Hsiao and Gibson's experiment. As discussed above, the two studies tested two different kinds of Chinese RCS (the first one has temporal ambiguities and the second one does not), which may lead to the different experiment results.

Table 3. Incremental procedure of ER Theory for Chinese non-temporarily-ambiguous RCs

SRC	<begin	Det	Cl	Time	[$_{CP}$ t_i	Vt	N	de]	N_i	Vt	N	end>
Entropy	4.62	4.00	4.00	7.45		6.74	5.11	4.86	4.46	3.53	1.4	
ER	none	0.62	0	0		0.71	1.63	0.25	0.40	0.93	2.15	Total ER:6.69

ORC	<begin	Det	Cl	Time	[$_{CP}$ N	Vt	t_i	de]	N_i	Vt	N	end>
Entropy	4.62	4.00	4.00	7.45	5.26	4.52		4.36	3.52	3.41	1.33	
ER	noun	0.62	0	3.45	2.19	0.73		0.16	0.84	0.12	2.08	Total ER: 6.74

As Table 3 shows, Yun et al. (2015) tested the non-temporarily-ambiguous Chinese RC sentences from the Jäger et al. (2015) experiment by applying ER theory from Hale (2006). The total ER of the sentences is calculated by adding up the ER at each stage. The total ER in the SRC sequence is higher than that in the ORC sequence, [6.67 vs 6.74]. Therefore, the theory predicts that SRCs are easier than ORCs. To be more specific, the theory predicts that SRCs are easier in the non-temporarily-ambiguous Chinese RC sentences. In Section 4, I show what the theory predicts for temporarily-ambiguous Chinese RC sentences.

4. Experiments

Yun et al. (2015) show that the ER theory makes the right predictions to match the experimental results from Jäger et al. (2015) for Chinese RCs. The study used corpora and tree banks as the training data to assign the probabilities over the context-free grammar rules converted from the minimalist grammar, then the ER was calculated at each stage of the incremental process by the probabilistic context-free grammar. The stages with more ER are predicted to be the positions where a reader slows down. In this section, I show that ER theory also make predictions that match the experimental result of the temporarily ambiguous Chinese RCs in Hsiao and Gibson's experiment (2003), despite the fact they found opposite results from Jäger et al.'s (2015) study.

4.1 The Evidence for Temporary Ambiguities by ER Theory

Table 4 shows all the possible structures and their probabilities are at the same stage of prefixing "Noun" in the temporarily ambiguous ORC model, and we can see many non-RC sentences co-occurring with the RC ones. The non-RC ones match the possibilities of temporarily ambiguous structures we discussed previously in Section 3.1. The probabilities associated with the structures indicate that the temporarily ambiguous structures affect the probability distribution and further affect the predictions of the theory.

Table 4. Probability Distribution of pre-fix "Noun" with Hsiao and Gibson's (2003) model

Probability	Remainder	Type
0.399	[$_{CP}$ Noun Vt Noun]	Simple sentence (transitive verb)
0.159	[$_{CP}$ Noun Vi]	(intransitive verb)
0.092	[$_{CP}$ [$_{PossP}$ Noun de Noun] Vt Noun]	(possessive subject)
0.092	[$_{CP}$ Noun Vt [$_{PossP}$ Noun de Noun]]	(possessive object)
0.036	[$_{CP}$ [$_{PossP}$ Noun de Noun] Vi]	PossP + intransitive verb
0.029	**[$_{CP}$ Noun Vt [$_{SRC}$ Vt Noun de] Noun]**	**SRC modifying object position**
0.021	[$_{CP}$ [$_{PossP}$ Noun de Noun] Vt [$_{PossP}$ Noun de Noun]]	two possessives
…	…	…
Entropy = 3.622		

4.2 ER Theory with a Modified Grammar

In Jäger et al.'s (2015) study, as the sequences shown in Table 3, their reason for having a temporal adverbial (labeled "Time") is to ensure that the readers will expect that an RC is coming after this word is heard. However, an alternative way to interpret the structure of this sequence with the temporal adverb in this case is to attach from the left of the VP and be a part of the VP, as in (7a), instead of being part of the CP, as the model assumes. Consequently, it is possible for the readers to comprehend an SRC with the same word order as in Table 3, but with a different structural interpretation, as in (7a).

(7) a. Det Cl [$_{CP}$ t_i **Temp** Verb Noun de] $Noun_i$ VP. (SRC)
b. Det Cl [$_{CP}$ Noun **Temp** Verb t_i de] $Noun_i$ VP. (ORC)

This structural ambiguity between SRC in (7a) and Table 3 is caused by the silent subject trace, since readers do not explicitly know where the trace is. I suggest that the natural position for the Temp/Time adverb in an RC sentence is to attach to the VP as the structure in (7a). If this is the grammar that the readers have, the SRC and ORC models in Jäger et al. (2015) (in Table 3) are not a minimal pair, because the ORC sequence in Table 3 may not valid, and (7b) would be the natural sequence for interpreting Temp adverbs.

In addition, again, if (7a) is the right structural analysis for Temp, it will lose its function as a signal of entering an RC region. I propose that instead of using Temp adverbs, complex adverbials (adverbials that contain more than one word) are more likely to attach to CPs rather than VPs, as shown in (8). Complex adverbials in Chinese can be adjunct phrases or adjunct clauses, such as *yinwei* ("because"-clauses) and *dang* ("while"-clauses).

(8) a. Det Cl ComplexADJ [$_{CP}$ t_i Verb Noun de] $Noun_i$ VP. (SRC)
b. Det Cl ComplexADJ [$_{CP}$ Noun Verb t_i de] $Noun_i$ VP. (ORC)

With the updated disambiguated sentence model, I calculate the ER for the SRC and the ORC following the methods in Yun et al. (2015) using a minimalist grammar set with promotion analysis. I modified the relevant grammar rules by deleting the rules that associated with "Temp" and adding the ComplexADJ to the grammar. In assigning probabilities to grammar rules, each grammar rule was looked up in corpora/tree banks, to count its frequency of use. Yun et al. (2015) used Chinese Treebank 7.0, which as about 4.5 million words in it (Xue et al. 2010). The frequency of a grammar rule can be captured in tree structures. Yun et al. used a tool called Tregex to capture all the sentences that had the tree structure of each rule and to return those trees and their total number. I use the same tree bank and the same tool to capture the PP-ADV configuration.

4.3 Results of ER Predictions with the New Grammar

ER predictions of different Chinese RC sentences are calculated using the updated grammar, which allows complex adjuncts to be attached to CP. There are four possible ways of forming a Chinese RC with different prefixing, as shown in Table 5. In the non-temporarily-ambiguous model with the prefix "Det Cl PP-ADV", SRCs are easier to process than ORCs. This suggests that if Jäger et al.'s experiment were re-run with the updated grammar, and if ER theory is the correct analysis, then we would expect SRCs to be easier than ORCs. The new grammar does not lead to different predictions for Yun el al. (2015) in terms of which RC is easier, but as grammar rules are crucial in calculating ER, the new grammar is not only limited to making predictions about RCs but for all other Chinese sentences. A more accurate grammar will make the actual entropy value more meaningful in informing us about sentence perception difficulties.

Table 5. ER results of four types of Chinese RC sentences

Word order	RC	Total ER	ER Predictions
(*non-temporarily-ambiguous*)			
Det Cl PP-ADV Vt Noun de Noun Vt Noun	SRC	6.50	SRC is easier
Det Cl PP-ADV Noun Vt de Noun Vt Noun	ORC	6.56	
(*temporarily-ambiguous)* Hsiao and Gibson 2003			
Vt Noun de Noun Vt Noun	SRC	5.15	ORC is easier
Noun Vt de Noun Vt Noun	ORC	4.40	
(*partial-temporarily-ambiguous*)			
Det Cl Vt Noun de Noun Vt Noun	SRC	6.10	ORC is easier
Det Cl Noun Vt de Noun Vt Noun	ORC	4.20	
(*partial-temporarily-ambiguous*)			
PP-ADV Vt Noun de Noun Vt Noun	SRC	5.10	ORC is easier
PP-ADV Noun Vt de Noun Vt Noun	ORC	4.61	

The other three kinds of RCs in Table 5 are temporarily ambiguous, and the total ER of each pair of RCs predicts that ORCs are easier than SRCs. The model with no prefix was the sentence model in Hsiao and Gibson's (2003). The total ER prediction matches their experimental result that the ORCs were harder for people to understand than the SRCs.

The two kinds of RCs with only one prefix have fewer occurrences of temporary ambiguities, but they are still temporarily ambiguous, and these sentences have not been tested in any human experiment in the current literature. But ER theory predicts that ORCs are easier than SRCs, so if the theory is right, the human experiment should also have the same testing results with either prefix. Interestingly, the sentence with the prefix Det-Cl has the largest total ER difference among all four kinds of RC structures, so again, if the ER theory is correct, it is reasonable to expect the actual reading time of this kind of RC will have the largest time difference in human experiments.

We can now add a column of ER theory predictions to Table 1 for the four kinds of RCs. ER theory uses both syntactic analysis and the probabilistic theory to make predictions, so it is a different approach than the distance theories. ER theory may be the only theory that can make different predictions about different kinds of Chinese RC sentences, which matches both Hsiao and Gibson's (2003) and Jäger et al. (2015) human experiment results. Because of the pre-nominal RC structure and its lack of morphology, Chinese can have different kinds of RCs, and can be temporarily ambiguous during perception process.

4.4 ER Results of Evened-out wh-Movement Rules

In calculating the frequencies of wh-movements in the tree bank, there are two rules of frequencies matter in the grammar: the subject wh-movement rule (assigned a probability of 2552/3695), and the object wh-movement rule (assigned a probability of 1130/3695). One could argue that assigning a higher probability to subject wh-movement leads to the result that SRCs have lower ER and end up being "easier". To check whether this particular probability assigning has an effect on the processing difficulty, I evened out the probability for the two rules, as shown in (9), and re-run the ER theory model.

(9)

1841 / 3695 | (: T -f;: -wh) --> (: +case T -f;: -case -wh)
| t86 --> t73 [0,0][0,1][0,2][0,3]
1841 / 3695 | (: T -f;: -wh) --> (: +case T -f;: -wh;: -case)
| t86 --> t89 [0,4;0,0][0,1][0,2][0,3]

The calculated ER results using the evened-out-RC-rules grammar are shown below in Table 6. Using the evened-out wh-movement rules, ER theory still predicts the same results as I tested previously for the four kinds of RC sentences. In the non-temporarily-ambiguous pair of RCs with both Det-Cl and ComplexADJ prefixing, SRCs are still easier than ORCs.

Table 6. ER results for four types of Chinese RC sentences with the modified grammar and evened out wh-movement rules

Word order	RC	Total ER	ER Predictions
Det Cl PP-ADV Vt Noun de Noun Vt Noun	SRC	6.64	SRC is easier
Det Cl PP-ADV Noun Vt de Noun Vt Noun	ORC	6.70	
Vt Noun de Noun Vt Noun	SRC	5.28	ORC is easier
Noun Vt de Noun Vt Noun	ORC	4.30	
Det Cl Vt Noun de Noun Vt Noun	SRC	6.04	ORC is easier
Det Cl Noun Vt de Noun Vt Noun	ORC	4.04	
PP-ADV Vt Noun de Noun Vt Noun	SRC	5.32	ORC is easier
PP-ADV Noun Vt de Noun Vt Noun	ORC	4.48	

This indicates that how frequently RC-movement rules are used in the corpus does not by itself determine whether ORCs or SRCs are easier in Chinese. Instead, the predictions are more likely to be determined by a combination of rules. In addition, this result also indicates that evening out the wh-movement rules will not affect the final predictions for the non-temporarily-ambiguous RCs, and it also will not change the final predictions for the temporarily ambiguous RCs, even if the actual ER values are changed.

5. Conclusion

In this project, I proposed a modified testing sentence sequence using a complex adjunct adverbial attached to the RC in the experiment in Jäger et al. (2015). With the probabilistic grammar assigned from the Chinese tree bank, I also used ER theory to make predictions about the modified disambiguated RC and three other kinds of RCs which are temporarily ambiguous. The result shows that SRCs are easier in the non-temporarily-ambiguous model, but ORCs are easier in the ambiguous model. These predictions match the opposing human experiments conducted by Hsiao and Gibson's (2003) and Jäger et al. In addition, I examined the four kinds of RC sentences in a hypothesized grammar that has the same frequency of the wh-movement rules for SRC and ORC, and the result did not change. This suggests that the higher frequency of the RC wh-movement rule in the grammar does not determine that type of RC sentence must have smaller total ER or that the sentence must be easier than the other type of RC. I also discussed the relationship between temporary ambiguities and ER predictions. I conclude that the probability distribution is the fundamental factor for determining ER predictions. So far, ER has correctly predicted all the sentences matching the human experiment results. It successfully predicted the result for both non-temporarily-ambiguous and the temporarily ambiguous sentences in Chinese. For the future, studies of other kinds of RC sentences in Chinese need to be tested. Also, more languages which have temporary ambiguities in RCs should also be examined to ensure that ER theory works cross-linguistically.

References

Gibson, Edward. (2000) The Dependency Locality Theory: A Distance-Based Theory of Linguistic Complexity. *Image, Language, Brain.* 95-126.

Hale, John. (2016) Information-Theoretical Complexity Metrics.

Hale, John. (2003) The Information Conveyed by Words in Sentences. *Journal of Psycholinguistic Research* 32, no. 2 101-123.

Hale, John. (2006) Uncertainty About the Rest of the Sentence. *Cognitive Science* 30, no. 4 643-672.

Hsiao, Franny, and Edward Gibson. (2003) Processing Relative Clauses in Chinese. *Cognition* 90, no. 1 3-27.

Jäger, Lena, Zhong Chen, Qiang Li, Chien-Jer Charles Lin, and Shravan Vasishth. (2015) The Subject-Relative Advantage in Chinese: Evidence for Expectation-Based Processing. *Journal of Memory and Language* 79 97-120.

King, Jonathan, and Marcel Adam Just. (1991) Individual Differences in Syntactic Processing: The Role of Working Memory. *Journal of Memory and Language* 30, no. 5 580-602.

O'grady, William. (2007) *Syntactic Development*. University of Chicago Press.
Shannon, Claude Elwood. (2001) A Mathematical Theory of Communication. *ACM SIGMOBILE Mobile Computing and Communications Review* 5, no. 1 3-55.

Stabler, Edward. (1996) Derivational Minimalism. In *International Conference on Logical Aspects of Computational Linguistics*, pp. 68-95. Springer Berlin Heidelberg.

Stabler, Edward P. (2011) Computational Perspectives on Minimalism. *Oxford Handbook of Linguistic Minimalism* 617-643.

Traxler, Matthew J., Robin K. Morris, and Rachel E. Seely. (2002) Processing Subject and Object Relative Clauses: Evidence from Eye Movements. *Journal of Memory and Language* 47, no. 1 69-90.

Wanner, Eric. (1978) An ATN Approach to Comprehension. *Linguistic Theory and Psychological Reality.*

Xue, Nianwen, Zixin Jiang, Xiuhong Zhong, Martha Palmer, Fei Xia, Fu-Dong Chiou, and Meiyu Chang. (2010) Chinese Treebank 7.0." *Linguistic Data Consortium, Philadelphia*.

Yun, Jiwon, Zhong Chen, Tim Hunter, John Whitman, and John Hale. (2015) Uncertainty in processing relative clauses across East Asian languages. *Journal of East Asian Linguistics* 24, no. 2 113-148.