The role of information theory in gap-filler dependencies

1 Introduction

Filler-gap dependencies are computationally expensive, motivating formally richer operations than constituency formation. Many studies investigate the nature of online sentence processing when the filler is encountered before the gap. Here the difficulty is where a gap should be posited. Comparatively few studies investigate the reverse situation, where the gap is encountered before the filler. This is presumably due to the fact that this is not a natural class of dependencies in English, as it arises only in cases of remnant movement, or rightward movement, the analysis of which is shakier and more theory laden than the converse. In languages with \textit{wh}-in-situ constructions, like Chinese, the gap-filler construction is systematic, and natural. Sentences 1 and 2 are declarative and embedded \textit{wh}-questions respectively in Mandarin Chinese.

1. LiuBei zhidao CaoCao ai LuBu
   LiuBei know CaoCao love LuBu
   ‘LiuBei knows that CaoCao loves LuBu.’

2. LiuBei zhidao CaoCao ai shei
   LiuBei know CaoCao love who
   ‘Who does LiuBei know that CaoCao love?’

Gap-filler constructions raise different problems than do filler-gap ones. In the latter, an item is encountered, which needs to satisfy other (to-be-encountered) dependencies to be licensed. There is no uncertainty that a gap must be postulated, only \textit{where} it should be postulated. In gap-filler constructions, a dependency is postulated before the item entering into it appears. In contrast to the filler-gap dependency type, gap-filler dependencies do not require more formal power from the syntax; they can (given a finite upper bound on their number) be analyzed with GPSG-style slash-feature percolation and are thus context-free. In systems with (covert) syntactic movement, the \textit{wh}-mover is predictably silent, and could be optimized away (into the context-free backbone of the derivation tree). The motivation for the postulation of a syntactic dependency is to streamline the account of sentence processing; while a purely semantic scope taking account could be implemented (e.g. using continuations), the role and resolution of semantic information during parsing is not as well understood.

Our goal is to understand the role that information theoretic complexity metrics [2] can play in the analysis of Chinese-like \textit{wh}-in-situ constructions. In particular, whether humans’ use of probabilistic cues about the presence of a gap can be modeled using the metrics of surprisal and/or entropy reduction. To this end, we identified a sentence processing data set where such cues were manipulated, wrote a Chinese grammar fragment deriving the stimuli, estimated probabilities from the Penn Chinese Tree Bank, and calculated (using the Cornell Conditional Probability Calculator, [1]) surprisal and entropy reduction values at each word.

2 Grammatical analysis

We use the minimalist grammar (MG) formalism [3] to frame our analysis. This formalism allows for the straightforward and transparent encoding of prominent linguistic ideas into a formal system. As the CCPC does not (to our knowledge) support covert movement, we adopted a ‘feature movement’ analysis of the Chinese \textit{wh}-in-situ construction, whereby a \textit{wh}-word is the result of combining a ‘pre-\textit{wh}-word’ with a silent \textit{wh}-moving item ( [ ] : [ w, \text{-w} ] ). This analysis allows us to implement the observation that \textit{wh}-words in Chinese can be used as well as indefinites, by deriving (syntactically) the \textit{wh}-word from the indefinite, although this did not play a role in our analysis.

3 Frequency Estimation

The CCPC works by translating MGs to equivalent MCFGs, and then parsing using the MCFG. When multiple rules expand the same non-terminal, we need to assign a (non-unit) weight to these rules.
As there is currently no MG (or MCFG) TreeBank for Chinese, we were forced to estimate weights of rules by reasoning about the structures in the treebank. As an example, there were two MCFG rules targeting the nonterminal :T; :–w, which is the category of an expression of category TP with a wh-moving subexpression. The two rules derive the TP containing a wh-word by

1. checking the case of a +WH subject DP
2. checking the case of a -WH subject DP, and having the TP contain another wh-word

What we counted in the Treebank is the relative frequency of TP/Ss which contain active wh-words where this wh-word is the matrix subject, vs a non-matrix subject.

4 The data set

We used a data set from an existing eye-tracking reading experiment (Experiment 1 in [4]). The original experiment consisted of 8 different conditions, which were largely designed to create different scoping possibilities for the in-situ wh-word. We implemented the structural properties of these conditions into our grammatical analysis in section 2, such that every condition could be derived by our grammar. In the original experiment, participants read sentences silently on a computer screen, and their eye-movements were recorded. The data set consisted of data from fifty native Mandarin speakers, each read 48 critical trials based on the 8 experimental conditions.

5 Results and discussion

We focused on 4 different eye-movement measures. First pass duration is the sum of all fixations in a region from the eyes first entering the region until leaving it either to the left or to the right. Go-past time is the sum of all fixations from first entering a region until leaving the region to the right, including fixations made during regression to earlier parts of the sentence. Second pass duration is the sum of all fixations in a region following the initial first-pass fixations. Total time is the overall reading time (all fixations) in a given region. For each eye-movement measure, we computed average reading time (RT), collapsing over participants and trials, for each word region under each condition. Next using the CCPC software, the grammar analysis in section 2 and the frequency estimation in section 3, we generated the entropy reduction (ER) and surprisal predictions for each word region under each condition. We then performed four linear regressions, using ER and surprisal as predictors and the four eye-movement measures as dependent variables.

Neither ER or surprisal are significant predictors for the first pass duration (ps>.5). For the go-past time, surprisal is not significant (p>.2), but ER is (p<.05). However, the model with ER as a predictor accounted for very little of the overall variance in the data (adjusted $R^2=0.04$). For second-pass and total time RTs, both ER and surprisal are significant (ps for ER <.01; ps for surprisal <.001). When both predictors are considered together, $R^2=0.23$ for the second pass measure and $R^2=0.32$ for the total time measure. When the two predictors are considered separately, surprisal accounted for more variance in the data than ER ($R^2=0.17$ surprisal vs. 0.05 ER for the total time; 0.13 vs. 0.03 for the second pass time).

If we consider the four eye-movement measures first pass, go past, second pass and total time form a scale to measure effects from the earlier stages of processing to the later stages, we observe that for the current data set information-theoretic complexity metrics such as ER and surprisal seem to mostly explain later measures but not the early ones. Furthermore, even with the second pass and total time measures, ER and surprisal only accounted for a relatively small amount of variance in the data, with surprisal having a better performance than ER.

References

[1] Zhong Chen, Tim Hunter, Jiwon Yun, and John Hale. Modeling sentence processing difficulty with a conditional probability calculator. In Proceedings of the Annual Meeting of the Cognitive Science Society, volume 36, 2014.

[2] John T. Hale. Information theoretical complexity metrics. Language and Linguistics Compass, 10(9):397–412, 2016.

[3] Edward Stabler. Derivational minimalism. In International Conference on Logical Aspects of Computational Linguistics, pages 68–95. Springer, 1996.

[4] Ming Xiang and Suiping Wang. Locality and expectation in mandarin wh-in-situ dependencies. manuscript to be submitted. 2019.