Uncertainty and Prediction in Relativized Structures across East Asian Languages

Zhong Chen, Jiwon Yun, John Whitman, John Hale
Department of Linguistics, Cornell University

Introduction

Entropy Reduction (Hale, 2006) is a complexity metric that quantifies the amount of information a word contributes towards reducing structural uncertainty. This certainty level depends on weighted, predictive syntactic analyses that are "still in play" at a given point. This paper uses Entropy Reduction to derive reported processing contrasts in Korean, Chinese and Japanese relativized structures.

Modeling procedure

Minimalist Grammar (Stabler, 1997)

Weighted Constructional Grammar

Weighted Context-Free Grammar

goes each prefix in the sentence

Weighted, predictive syntactic analyses

Uncertainty at this word

Korean

Experimental Observations:

SBJ Relatives > OBJ Relatives (Kwon, 2010)

ER Modeling:

Subject Relatives (SR)

Object Relatives (OR)

Analysis

Comprehension difficulty prediction

Our modeling confirms the SR preference in Korean reported by Kwon et al. (2010) and further shows that this effect could emerge as early as the accusative/nominative marker. This reflects, among other factors, a greater entropy reduction brought by sentence-initial nominative noun phrases.

Chinese

Experimental Observations:

SBJ Relatives > OBJ Relatives (Lin & Bever, 2006; Wu, 2009; Chen et al., CUNY 2012)

SBJ Relatives < OBJ Relatives (Hsiao & Gibson, 2003; Gibson & Wu, in press)

ER Modeling:

Subject Relatives (SR)

Object Relatives (OR)

Analysis

Comprehension difficulty prediction

Our modeling derives an SR advantage at the head noun in line with structural frequencies (SR 55%/OR 45%). It also implicates headless RCs as a grammatical alternative whose existence makes processing easier at the head noun in SRs. A corpus study reveals that 14% of SRs have a null head whereas 31% of ORs are headless. This asymmetry suggests that an overt head is more predictable in SRs and less work needs to be done.

Japanese

Experimental Observation:

Subject Clefts < Object Clefts (Kahraman et al., 2011)

ER Modeling:

Subject Clefts (SC)

Object Clefts (OC)

Analysis

Comprehension difficulty prediction

Our modeling derives a pattern consistent with the empirical finding in Kahraman et al. (2011) that at the "me-wo" marked embedded void, subject clefts are read more slowly than object clefts. Upon reaching the topic marker "me-wo", complement clauses with SBJ pro are still in play in case of the SC prefix, which causes more amount of uncertainties reduced around that point. On the other hand, the OC prefix is less ambiguous because complement clauses with object-pro are extremely rare.

Conclusion

Grammatical phenomena such as case-marking, head-omission, and object-drop create inferential problems that must be solved by any parsing mechanism. The Entropy Reductions brought about by "solving" these problems — moving towards more concentrated derivations — correspond with observed processing difficulty.

Selected References