

Uncertainty and Prediction in Relativized Structures across East Asian Languages

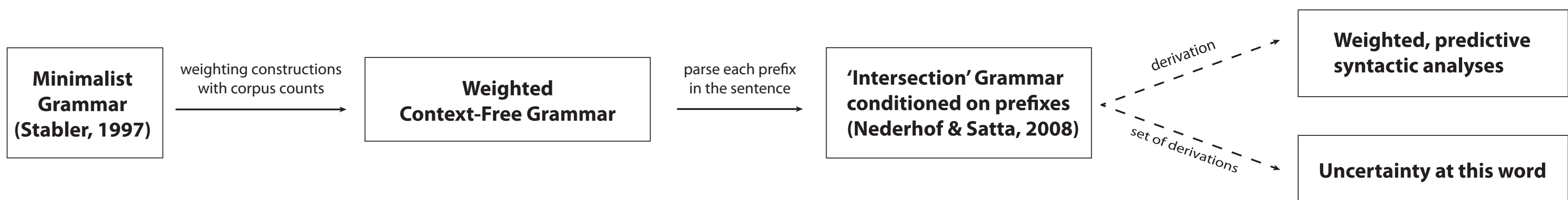
Zhong Chen, Jiwon Yun, John Whitman, John Hale

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 poster uses Entropy Reduction to derive reported processing contrasts in Korean, Chinese and Japanese relativized structures.

Modeling procedure



Korean

Experimental Observation:

SBJ Relatives > OBJ Relatives (Kwon, 2010)

ER Modeling:

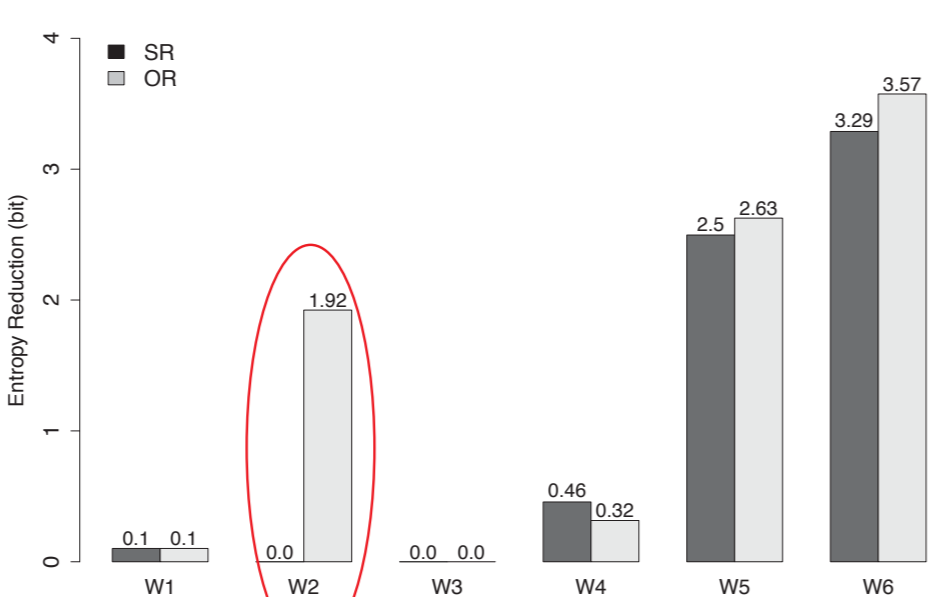
Subject Relatives (SR)

W1 W2 W3 W4 W5 W6
e 기자 를 협박한 의원 이 유명해졌다.
(SBJ) reporter ACC threaten-ADN senator NOM became famous
'The senator who threatened the reporter became famous.'

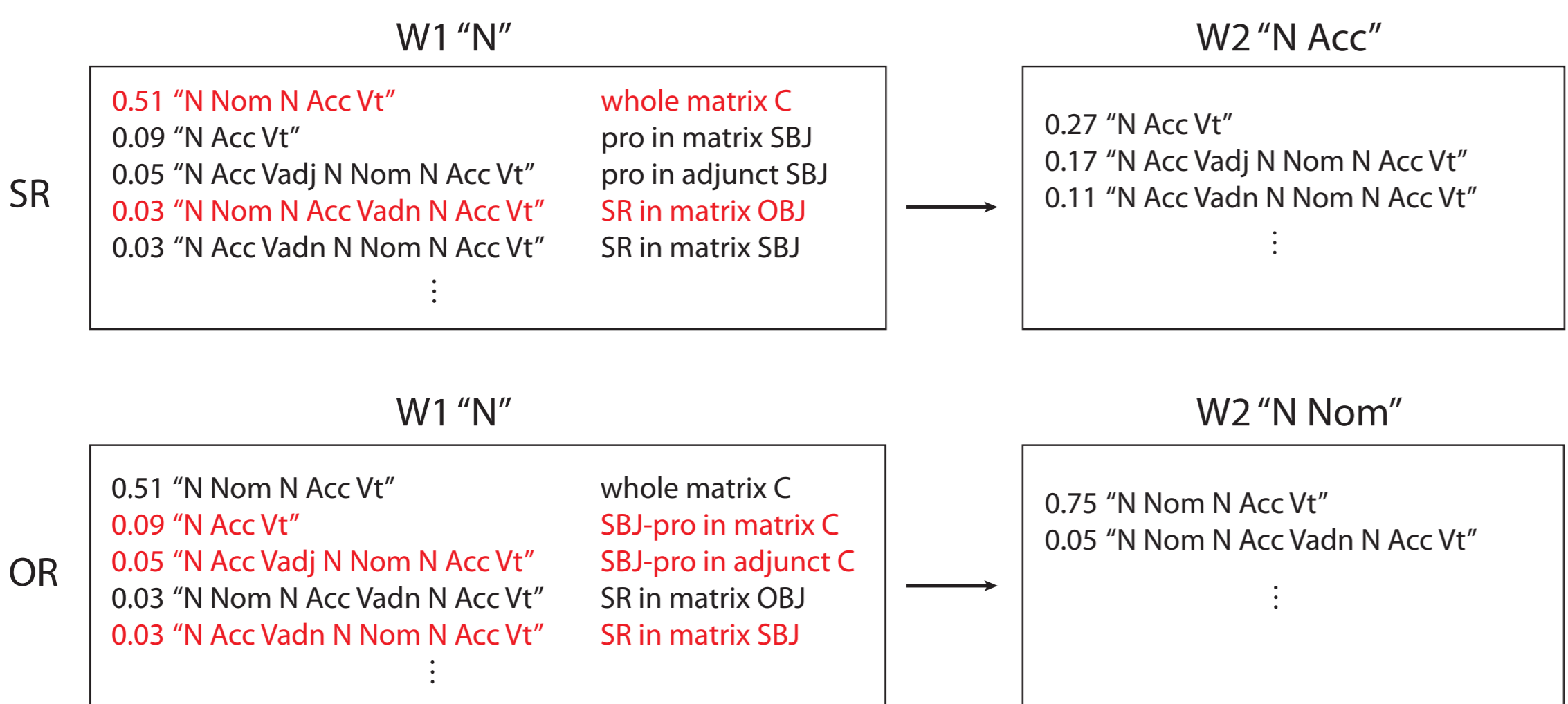
Object Relatives (OR)

W1 W2 W3 W4 W5 W6
기자 가 *e* 협박한 의원 이 유명해졌다.
reporter NOM (OBJ) threaten-ADN senator NOM became famous
'The senator who the reporter threatened became famous.'

Comprehension difficulty prediction



Analysis



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:

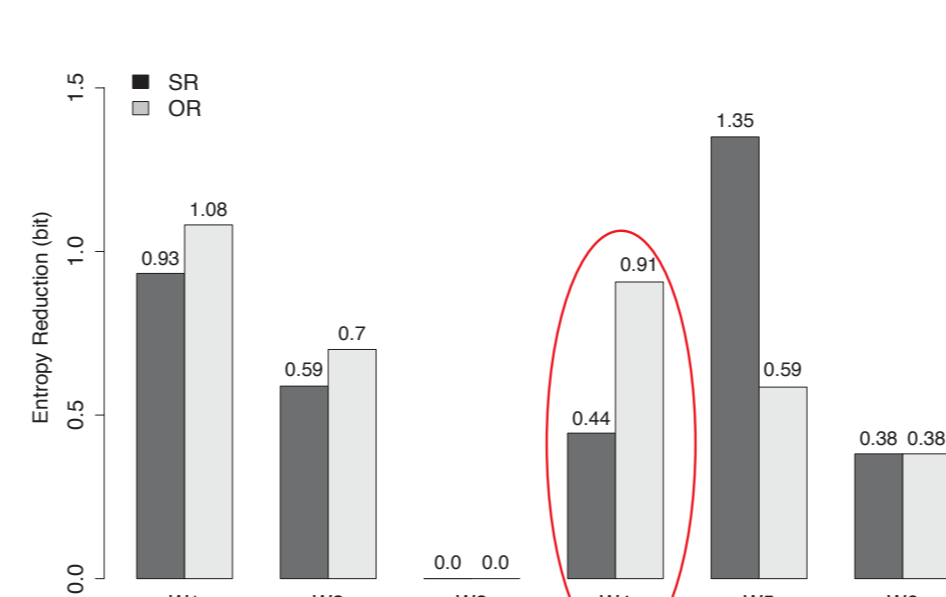
Subject Relatives (SR)

W1 W2 W3 W4 W5 W6
e 邀请 富豪 的 (官员) 打了 记者
SBJ invite tycoon DE official hit reporter
'The official/Someone who invited the tycoon hit the reporter.'

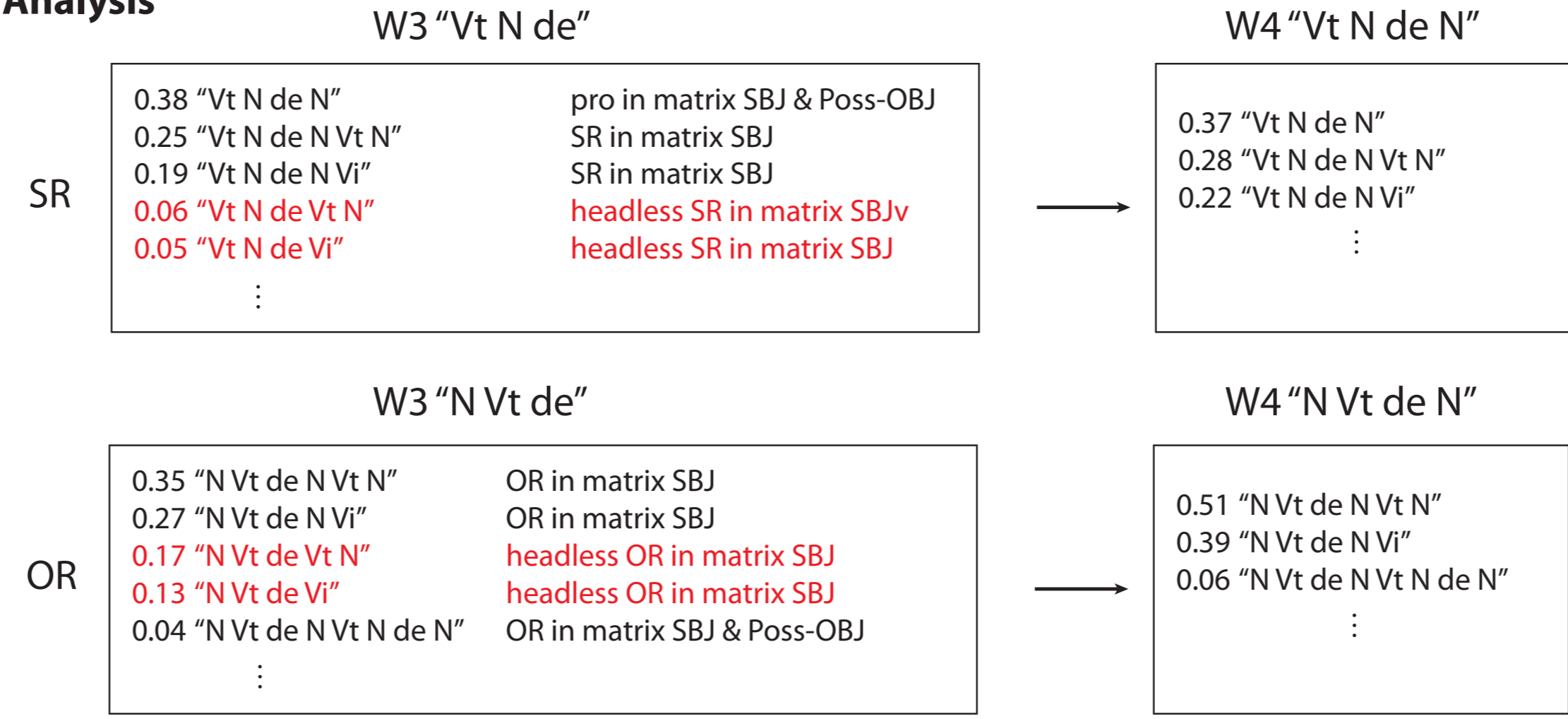
Object Relatives (OR)

W1 W2 W3 W4 W5 W6
富豪 邀请 *e* 的 (官员) 打了 记者
tycoon invite OBJ DE official hit reporter
'The official/Someone who the tycoon invited hit the reporter.'

Comprehension difficulty prediction



Analysis



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:

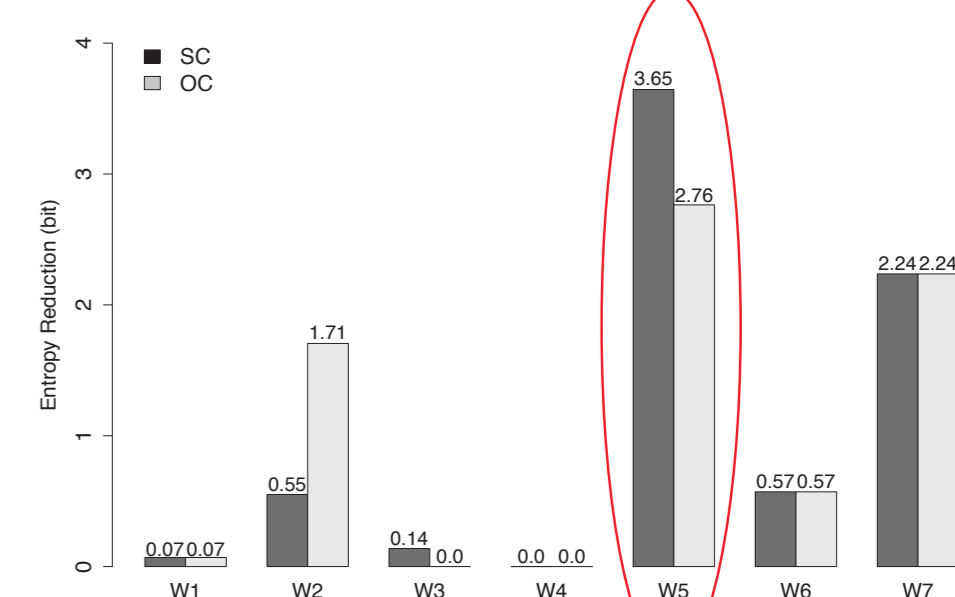
Subject Clefts (SC)

W1 W2 W3 W4 W5 W6 W7
e 祖母 を 介抱した の は 親戚 だ。
(SBJ) grandma ACC nursed NO WA relative COP
'It was the relative who nursed the grandmother.'

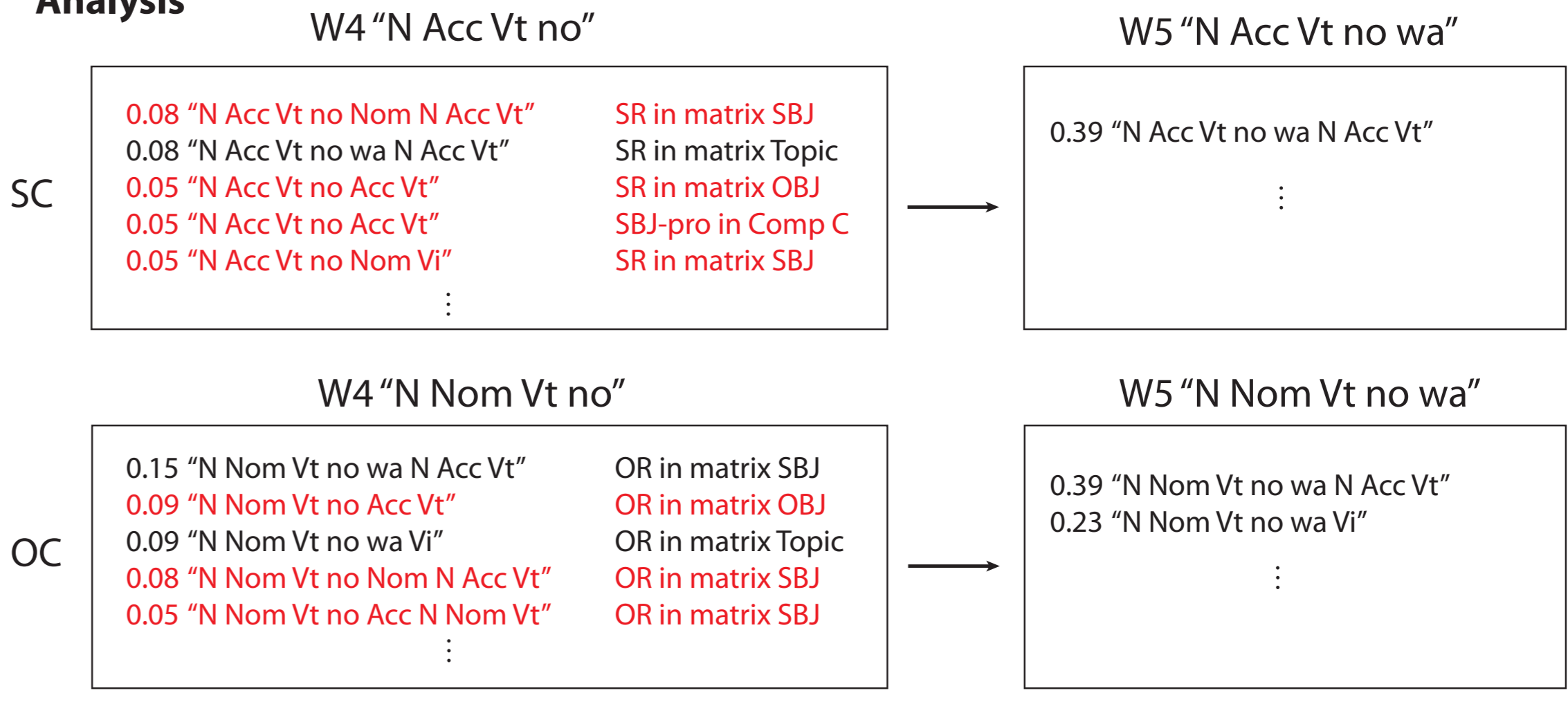
Object Clefts (OC)

W1 W2 W3 W4 W5 W6 W7
祖母 が *e* 介抱した の は 親戚 だ。
grandma NOM (OBJ) nursed NO WA relative COP
'It was the relative who the grandmother nursed.'

Comprehension difficulty prediction



Analysis



Our modeling derives a pattern consistent with the empirical finding in Kahraman et al. (2011) that at the "no-wa" marked embedded verb, subject clefts are read more slowly than object clefts. Upon reaching the topic marker "-wa", 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 distributions on derivations -- correspond with observed processing difficulty.

Selected References

- Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, 30(4).
- Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, 90, 3–27.
- Kahraman, B., Sato, A., Ono, H. & Sakai, H. (2011). Incremental processing of gap-filler dependencies: Evidence from the processing of subject and object clefts in Japanese. In *Proceedings of the 12th Tokyo Conference on Psycholinguistics*.
- Kwon, N., Lee, Y., Gordon, P. C., Kluender, R., & Polinsky, M. (2010). Cognitive and linguistic factors affecting subject/object asymmetry: An eye-tracking study of pre-nominal relative clauses in Korean. *Language*, 86(3), 546–582.
- Lin, C.-J. C., & Bever, T. G. (2006). Subject preference in the processing of relative clauses in Chinese. In *Proceedings of the 25th WCCFL* (p. 254-260).
- Nederhof, M.-J. & Satta, G. (2008). Computation of distances for regular and context-free probabilistic languages. *Theoretical Computer Science*, 395, 235–254.
- Stabler, E. (1997). Derivational minimalism. In C. Retore (Ed.), *Logical aspects of computational linguistics*. Springer-Verlag.