The Role of Inflectional Entropy in VP Ellipsis Resolution
Alex B. Fine, Christina S. Kim, Jeffrey T. Runner
University of Rochester
Contact: afine@bes.rochester.edu

Background
An important question for psycholinguistic accounts of ellipsis phenomena is exactly what kind of information is reconstructed when comprehenders encounter elided or deleted material.

In this study, we focus on the question of how detailed the morphological representation of the reconstructed verb form is.

If the comprehension system reconstructs a rich representation of the morphological features of an elided verb, the processing of VP ellipsis may be sensitive to probabilistic information about the elided verb’s morphological paradigm.

Goals of this study
In (1), we expect a processing cost associated with reconstructing the appropriate form of the verb at the ellipsis site when that form does not match the antecedent form. Further, we expect this cost to vary with the entropy of the distribution of the verb’s inflectional forms (Moscoso del Prado Martin et al. 2004).

(1) John scored a goal, and Susan will score a goal too.

Intuitively, we can think of an inflectional paradigm’s entropy increasing as the probability distribution over the members of the paradigm becomes more balanced. For instance, the verb “learn” has a higher inflectional entropy than the verb “spare”.

Hypothesis: higher inflectional entropy will lead to greater difficulty in reconstructing non-identical morphological forms during ellipsis resolution.

On the other hand, higher inflectional entropy actually makes identifying a word easier (Moscoso del Prado Martin et al. ibid, Baayen et al. 2006). Therefore if the two forms are morphologically identical, entropy may have a facilitatory effect in comprehension.

Methods
Verbs were selected whose inflectional entropy measures ranged in magnitude*. These were used to create sentences manipulated for ellipsis (+/-) and morphological match (+/-), yielding the following conditions:

[No]Ellipsis/Match: John scored a goal, and Susan had [scored a goal] too.
[No]Ellipsis/Mismatch: John scored a goal, and Susan will [score a goal] too.

32 native English speakers participated in a magnitude estimation study, in which subjects provide numerical estimates for how natural a sentence sounds (Bard et al. 1996).

Results
The logs of the normed estimate values were regressed against Ellipsis, Match, and Entropy, with Subject and Item as random effects.

Main effects:
Match (β = .07, SE = .008, p<.01) and Ellipsis (β = .02, SE = .008, p<.05) – sentences with morphological match and with ellipsis were judged as more natural than Mismatch and No Ellipsis sentences, respectively.

Interactions:
2-way Match-Ellipsis interaction (β = .02, SE = .008, p<.01) – the effect of morphological mismatch was greater in Ellipsis than Non-Ellipsis cases (consistent with the hypothesis that there is a cost associated with mismatch).

Two-way Match-Entropy interaction (β = .04, SE = .01, p<.05) and a three-way Entropy-Ellipsis-Match interaction (β = .06, SE = .02, p<.01) show that entropy correlates positively with estimates in the Matched cases, and negatively in the Mismatched cases, and these correlations are only significant in the Ellipsis conditions. This pattern is displayed in the figures on the right.

*All verbs taken from the English database available in the languageR library (Baayen 2008)

Acknowledgements
We would like to thank Florian Jaeger for stats and information theory advice. We also thank Neal Snider and Mike Tanenhaus for helpful discussion.

Summary
This study provides evidence that ellipsis resolution requires reconstructing detailed morphological information about the elided verb, since comprehension depended upon probabilities computed over that information.

Comprehension of ellipsis depends on how difficult it is to identify one form of the verb, given other alternatives. The difficulty of this task was manipulated by choosing verbs whose inflectional paradigms varied in entropy.

This suggests that probabilistic features of constituents involved in anaphoric dependencies can be used to probe the mechanisms underlying their interpretation.

Future Work
The next step is to replicate this finding with an online measure, which we are currently doing with self-paced reading.

Another question is what kinds of information are relevant during the production of ellipsis—are speakers more likely to elide when the material being elided will be easier to reconstruct?

References