

Corpus and Experimental Data as Corroborating Evidence: The Case of Preposition Placement in English Relative Clauses

Thomas Hoffmann

University of Regensburg/Germany

`thomas.hoffmann@sprachlit.uni-regensburg.de`

1 Introduction: Corpus vs. introspection data

As is widely known, linguistic generalisations based on corpus data face two potential problems: 1) just because a phenomenon cannot be found in a corpus doesn't mean that it's ungrammatical (the "negative data" problem), and 2) just because a construction appears in a corpus it doesn't automatically follow that it's grammatical (the "performance" problem). Grammaticality judgements, on the other hand, are not flawed by these problems but the sentence stimuli used in such studies 1) have to be invented by the researcher (the "un-natural data" problem) and 2) thus, unlike corpus data, do not allow the investigation of contextual factors such as e.g. the level of formality (the "context" problem). In this talk, I will demonstrate how the complementary nature of corpus and grammaticality judgement data can be used as corroborating evidence when investigating syntactic variation.

The evidence: Preposition placement in English relative clauses

Now, an interesting area of syntactic variation within the English language is the placement of prepositions. In relative clauses, e.g., the preposition can either precede the WH-relativiser ("preposition pied-piping", cf. 1) or the relativised gap ("preposition stranding", cf. 2).

- (1) the place **[in which]**_i I live _____i
- (2) the place **[which]**_i I live **in** _____i

As I will show, a corpus study on the stranding–pied-piping phenomenon (drawing on data from the British English component of the *International Corpus of English*) reveals many categorical as well as variable tokens (Hoffmann, 2005). For the latter, I will advocate the usefulness of the GOLDVARB-software (cf. Robinson, Lawrence, and

Tagliamonte, 2001) for identifying statistically significant contextual factors determining preposition placement (such as e.g. the level of formality or the restrictiveness of the relative clause). For the former, I will argue that contrasting *wh*- and *that/Ø*-tokens already allows a first distinction between accidental gaps (i.e. grammatical constructions which are just accidentally missing in the corpus) and ungrammatical constructions.

Finally, I will illustrate how the findings of the corpus study can be corroborated by an on-line Magnitude Estimation experiment (cf. Bard, Robertson, and Sorace, 1996). As will be seen, the grammaticality judgements of 36 British English speakers (18 female, 18 male) support the conclusions based on the comparison of the *wh*- and *that/Ø*-corpus tokens. In addition to this, the Magnitude Estimation experiment allows a further sub-classification of ungrammatical constructions into those which violate soft constraints (**the manner which/that/Ø she killed the cat in*) and those which violate hard constraints (**the place in that/in Ø she killed the cat*; cf. Sorace, and Keller, 2005).

Conclusion: Corpora and introspection as corroborating evidence

Having presented the results from the corpus and the magnitude estimation study, I will finally argue that treating corpus and introspection data as corroborating evidence allows a far more detailed analysis of the categorical and variable constraints governing syntactic variation than the two types of data would have allowed individually. Therefore, whenever linguists are investigating syntactic phenomena they should not limit themselves to a single data source, when corroborating evidence to strengthen their case is so readily available.

References

- Bard, E.G., D. Robertson, and A. Sorace (1996). Magnitude estimation of linguistic acceptability. *Language*, 72: 32-68.
- Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgements*. Sage, Thousand Oaks.
- Hoffmann, T. (2005). Variable vs. categorical effects: Preposition pied piping and stranding in British English relative clauses. *Journal of English Linguistics*, 33,3: 257-297.
- Robinson, J. S., H. R. Lawrence, and S. A. Tagliamonte (2001). GOLDVARB 2001: A multivariate analysis application for Windows. <<http://www.york.ac.uk/depts/lang/webstuff/goldvarb/manualOct2001>>.
- Sorace, A. and F. Keller (2005). Gradience in linguistic data. *Lingua*, 115,11: 1497-1524.