CORROBORATING EVIDENCE: CORPUS VS. EXPERIMENTAL DATA
The Case of Preposition Placement in English Relative Clauses

Thomas Hoffmann
University of Regensburg/Germany
thomas.hoffmann@sprachlit.uni-regensburg.de

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 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.

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. 1a) or the relativised gap ("preposition stranding", cf. 1b).

(1) a. I want a data source [on which] I can rely __i
   b. I want a data source [which] I can rely on ___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. For the latter, I will advocate the usefulness of the GOLDVARB-software (cf. Robinson et al. 2001) for identifying statistically significant contextual factors determining preposition placement (such as e.g. 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 et al. 1996). As will be seen, the grammaticality judgements of British English speakers 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 subclassification 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 & Keller 2005).

Thus, the talk will indicate how treating corpus and experimental data as corroborating evidence allows a detailed description of all categorical as well as variable factors influencing preposition placement in English relative clauses.

REFERENCES