

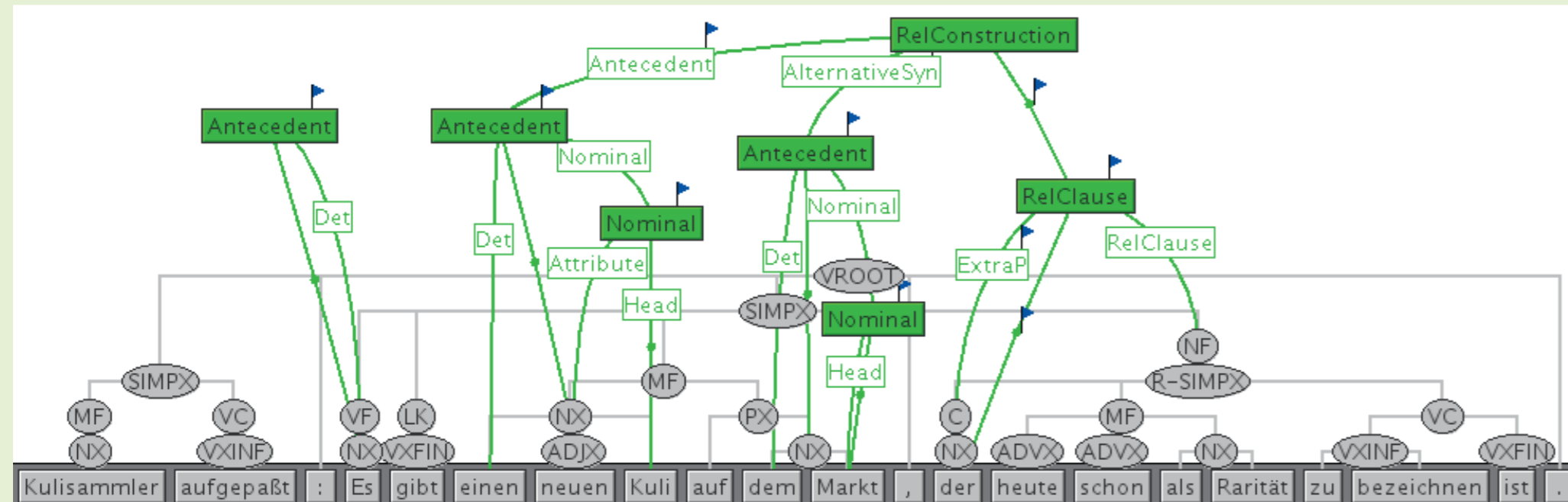
# INVESTIGATING RELATIVE CLAUSE EXTRAPOSITION IN GERMAN USING AN ENRICHED TREEBANK

Jan Strunk  
strunk@linguistics.rub.de

## Relative Clause Extrapolation

- Relative clauses in German can be realized as part of the head noun phrase (*integrated*) or at the end of the matrix clause (*extraposed*)
- Integrated Relative Clause**  
*Ich habe [DP alle diesbezüglichen Threads [RC die ich finden konnte]] gelesen*  
 I have all relevant threads that I find could read  
 "I have read all relevant threads that I could find."
- Extraposed Relative Clause**  
*Ich habe [DP alle Bücher \_\_\_] gelesen [RC die ich finden konnte]*  
 I have all books read that I find could  
 "I have read all books that I could find."

## Enriching the Treebank with Special-Purpose Annotation



- Relative edge construction modeled using frames and frame relations
- Features implemented using SALTO flags

## Basic Corpus

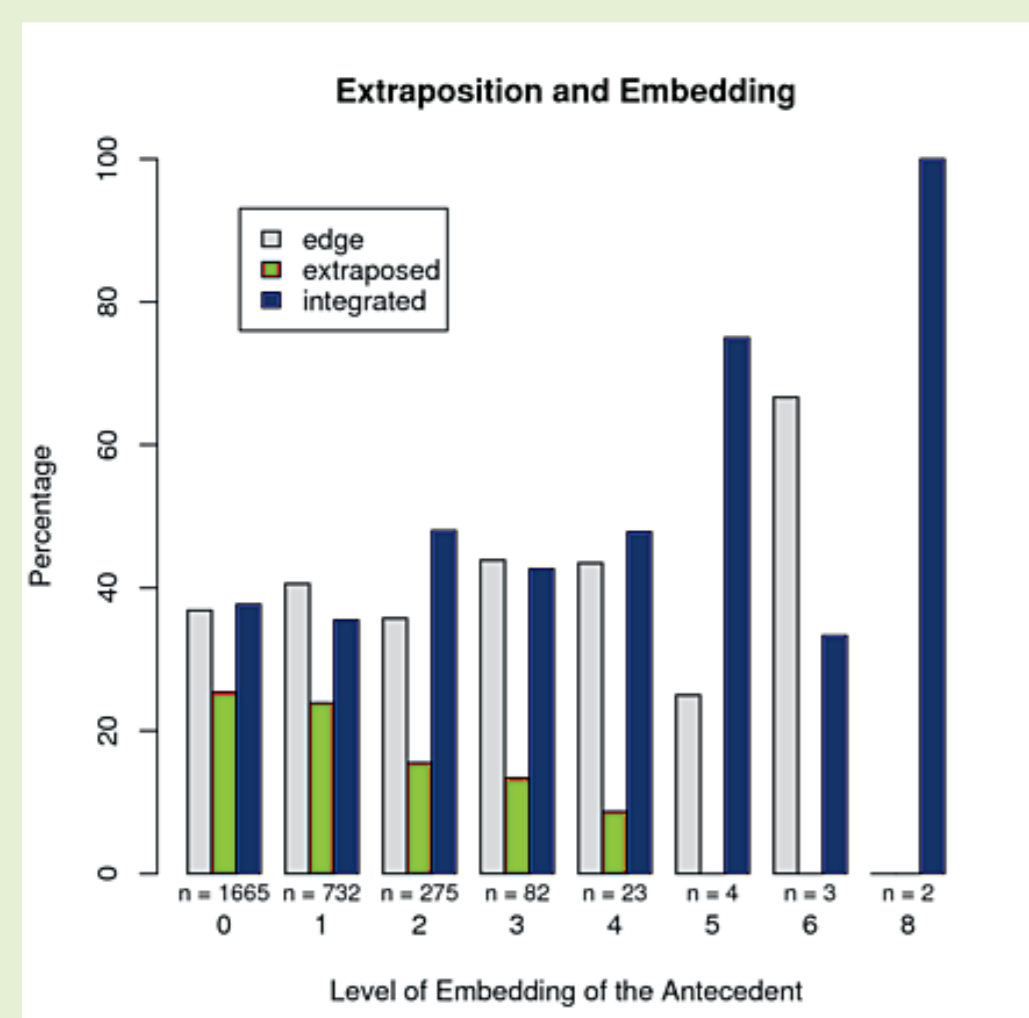
- Tübinger Baubank des Deutschen / Schriftsprache (TüBa-D/Z) (Tübingen Treebank of Written German) (Telljohann et al., 2005)
- Annotated with a relatively flat syntactic structure including topological fields, part-of-speech tags, and morphological features
- Sub-corpus including all sentences that contain a relative clause (R-SIMPX) extracted using TIGERSearch (Lezius, 2002): 2,603 sentences with 2,789 relative clauses

- Enriching the corpus with a second layer of special-purpose annotation using the tool SALTO (Burchardt et al., 2006) (originally intended for the annotation of frame semantic roles)
- Easy automatic processing of TIGER-XML including the additional "frame" annotation
- Convenient for manual checking, correction, and addition of features
- Features automatically deduced from the underlying treebank: *parts of the relative construction, position of the relative clause, depth of embedding, syntactic categories, syntactic functions, person, number, gender, case, definiteness, lengths and distances*
- Features added by hand: *restrictiveness of the relative clause, potential alternative antecedents*
- Planned annotation: *semantic class of antecedent (GermaNet), animacy, givenness, information structure*

## Pilot Studies using the Enriched Treebank

### Locality (Depth of Embedding of the Antecedent)

- Generative theories of locality predict that the antecedent of an extraposed relative clause cannot be embedded arbitrarily deeply
- Chomsky's (1973) *Subjacency* principle rules out extraposition from an NP/DP that is embedded inside another NP/DP
- Baltin's (2006) *Generalized Subjacency* predicts that the extraposed relative clause must be adjoined to the next higher max. projection
- These theories predict a sharp decline in extraposition likelihood for all antecedents that are embedded at least one level deep
- But extraposition likelihood decreases much more gradually



depth	extraposed	integrated	edge
0	423 (25 %)	628 (38 %)	614 (37 %)
1	177 (24 %)	260 (35 %)	297 (41 %)
2	43 (16 %)	133 (48 %)	101 (36 %)
3	11 (13 %)	35 (43 %)	36 (44 %)
4	1 (5 %)	11 (50 %)	10 (45 %)
5	0 (0 %)	3 (75 %)	1 (25 %)
6	0 (0 %)	1 (33 %)	2 (67 %)
8	0 (0 %)	2 (100 %)	0 (0 %)

### Definiteness of the Antecedent

- Guéron & May (1984) connect extraposition to quantifier raising
- This predicts that extraposition should only be possible from indefinite or quantified antecedents but not from definite ones

	extraposed	integrated	edge
<b>definite (n = 1,322)</b>	252 (19 %)	590 (45 %)	480 (36 %)
<b>indefinite (n = 1,122)</b>	335 (30 %)	334 (30 %)	453 (40 %)

### Restrictiveness of the Relative Clause

- Ziv & Cole (1974) claim that appositive relative clauses cannot be extraposed

	extraposed	integrated	edge
<b>restrictive (n = 1,207)</b>	334 (28 %)	450 (37 %)	423 (35 %)
<b>appositive (n = 1,023)</b>	180 (17 %)	457 (45 %)	386 (38 %)

## Conclusion

- Corpus data show that intuitions from the generative literature go in the right direction but go too far by assuming categorical constraints
- Plan to build complex models of relative clause extraposition both from production and perception perspective based on the treebank

Baltin, M. R. (2006). Extraposition. In Everaert, M. & van Riemsdijk, H. C. (eds.), *The Blackwell Companion to Syntax* (Vol. 2). Malden: Blackwell, pp. 237-271.

Burchardt, A., Erk, K., Frank, A., Kowalski, A. & Padó, S. (2006). SALTO - A versatile multi-level annotation tool. In *Proceedings of LREC 2006*, Genoa, Italy.

Chomsky, N. (1973). Conditions on transformations. In Anderson, S. R. & Kiparsky, P. (eds.), *A Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston, pp. 232-286.

Guéron, J. & May, R. (1984). Extraposition and logical form. *Linguistic Inquiry* 15(1): 1-31.

Lezius, W. (2002). Ein Suchwerkzeug für syntaktisch annotierte Textkorpora. PhD thesis, Institut für Maschinelle Sprachverarbeitung, University of Stuttgart.

Telljohann, H., Hinrichs, E. W., Kübler, S. & Zinsmeister, H. (2005). Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z). Seminar für Sprachwissenschaft, University of Tübingen.

Ziv, Y. & Cole, P. (1974). Relative extraposition and the scope of definite descriptions in Hebrew and English. In La Galy, M. W., Fox, R. A. & Bruck, A. (eds.) *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*, April 19-21, 1974. Chicago: Chicago Linguistic Society, pp. 772-786.