

# A Statistical Model of Competing Motivations Affecting Relative Clause Extraposition in German

Jan Strunk

Sprachwissenschaftliches Institut  
Ruhr-Universität Bochum, Germany  
E-mail: [strunk@linguistics.rub.de](mailto:strunk@linguistics.rub.de)

Relative clauses in German and other languages are usually realized *integrated* in the noun phrase that they modify but they can also be separated from their antecedent by intervening material and occur further to the right in *extraposed* position, mostly at the end of the matrix clause. It is usually assumed that the integrated and extraposed variants of relative clauses are semantically equivalent.

Relative clause extraposition has mostly been studied within generative grammar using introspective data (e.g. Baltin, 2006). Although a few corpus studies have also been published (Shannon, 1992; Uszkoreit et al., 1998; Hawkins, 2004), they have mostly concentrated on individual factors and have not tried to account for relative clause extraposition as a syntactic alternation using an integrated (statistical) model – as proposed, for example, for the English dative alternation by Bresnan et al. (2007).

Authors working within a formal generative paradigm like Baltin (2006) have identified different factors affecting (relative clause) extraposition than authors working within a functional and/or corpus-linguistic paradigm such as Hawkins (2004) and Shannon (1992); cf. the lists in (1). Generative linguists have also traditionally regarded constraints as categorical, whereas functionally oriented linguists have tended to regard proposed factors as gradient or probabilistic.

(1) **Generative studies:** syntactic locality, definiteness of the antecedent, restrictiveness of the relative clause

**Functionalist studies:** linear distance between relative clause and antecedent, length of the relative clause, information structure in the matrix clause

In earlier studies (e.g. Strunk 2010), I was able to show using corpus data and univariate statistical methods that the constraints proposed in generative studies based on intuitions do indeed go in the right direction but go too far in assuming categorical constraints.

In my contribution to this conference, I would like to present the results of a more detailed corpus study using an integrated logistic regression model that combines various competing factors and predicts whether a relative clause will be extraposed or not. This model is built on the basis of a treebank of written German that is enriched with an additional annotation level and additional features relevant for relative clause extraposition. The corpus currently contains 2,603 sentences including 2,789 relative clauses in all. In the statistical model, I combine the factors proposed in generative and functionalist studies as well as additional predictors in order to determine which of the different proposed factors are indeed needed to account for the linearization decision and which can be reduced to other factors.

Table 1 gives coefficients and p-values for significant predictors in a preliminary statistical model (prediction accuracy: 85%). These preliminary results already show that both “formal” factors such as syntactic locality (*embedding*) and also functional factors such as *length of the relative clause* and the position of the antecedent (*Nachfeld/Vorfeld*) have a significant impact. The most important predictor in the model is the position of the antecedent within the topological structure of the German clause: Specifically, the likelihood of extraposition decreases dramatically if the antecedent is located in the *Vorfeld* (“prefield”) in front of the finite verb (cf. Shannon, 1992). This is also in accordance with Uszkoreit et al. (1998), who found that the linear distance between the antecedent and the relative clause in words was the strongest factor influencing extraposition (cf. also Hawkins, 2004).

Even though the position of the antecedent within the matrix clause affects relative clause extraposition very strongly, quite a few authentic examples can be found in which a relative clause is extraposed from an antecedent in the *Vorfeld* over a relatively long distance; cf. examples (2) and (3). As an additional topic, I would therefore like to discuss, based on further corpus and experimental evidence, how very strong constraints can sometimes be exceptionally overridden, either by the cumulative force of weaker competing constraints or by special marking strategies (such as cataphoric demonstratives).

Factor	Coeff.	Std. Err.	z value	p value
(Intercept)	-1.59	0.30	-5.23	<0.001
embedding	-0.34	0.14	-2.37	0.018
indefinite antecedent	1.44	0.23	6.28	<0.001
appositive relative clause	-0.63	0.21	-3.03	0.002
length of the relative clause	0.15	0.03	5.70	<0.001
accusative case	1.81	0.26	7.05	<0.001
dative case	1.22	0.28	4.35	<0.001
genitive case	2.16	0.46	4.72	<0.001
complex name	2.11	1.06	1.99	0.046
cataphoric	0.61	0.52	1.18	0.238
Nachfeld	-20.76	554.35	-0.04	0.97013
Vorfeld	-5.13	0.49	-10.55	<0.001

Table 1: Preliminary logistic regression model of relative clause extraposition

- (2) *[PP In [NP welches SKigebiet — ]] kann man über die Osterferien fahren*  
in what skiing region can you over the spring break drive  
*[RC das noch Schneesicher ist] [...]*  
that still snow-sure is

“In what skiing region can you travel over spring break that is guaranteed to have snow?”

(www.bergfex.at/forum/allgemein/?&msgID=1000049637, 2007-02-19)

- (3) *[NP Nur derjenige — ] kann eine Anrechnung einer Maßnahme bei künftigen*  
only DEM can a consideration of a measure during future  
*Eingriffen in Natur und Landschaft verlangen (Ökokonto), [RC der ...]*  
interference in nature and landscape demand (eco-account), who ...

“Only he can demand the consideration of a measure during future interference with nature or landscape (eco-account) who ...”

(Natur und Recht, Volume 28, Number 7, July 2006, pp. 471-472(2))

## References

- Baltin, M. R. (2006). Extraposition. In M. Everaert and H. C. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax*, Volume 2, pp. 237–271. Blackwell, Malden, Massachusetts.
- Bresnan, J., A. Cueni, T. Nikitina, and H. Baayen (2007). Predicting the dative alternation. In G. Bouma, I. Krämer, and J. Zwarts (Eds.), *Cognitive Foundations of Interpretation*, pp. 69–94. Royal Netherlands Academy of Science, Amsterdam, The Netherlands.
- Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford University Press, Oxford, UK.
- Shannon, T. F. (1992). Toward an adequate characterization of relative clause extraposition in modern German. In I. Rauch, G. F. Carr, and R. L. Kyes (Eds.), *On Germanic Linguistics. Issues and Methods*, pp. 253–281. Mouton de Gruyter, Berlin/New York.
- Strunk, J. (2010). Enriching a treebank to investigate relative clause extraposition in German. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC’10), May 19-21, 2010, Valletta, Malta*, pp. 1911–1917.
- Uszkoreit, H., T. Brants, D. Duchier, B. Krenn, L. Konieczny, S. Oepen, and W. Skut (1998). Studien zur performanzorientierten Linguistik: Aspekte der Relativsatzextraposition im Deutschen. *Kognitionswissenschaft* 7(3), 129–133.