Prototypes in the univerbation of German verb-noun units

Roland Schäfer
Deutsche Sprache und Linguistik,
Humboldt-Universität zu Berlin
Dorotheenstraße 24, 10117 Berlin
schaefer@hu-berlin.de

Ulrike Sayatz
Deutsche und niederl. Philologie,
Freie Universität Berlin
Habelschwerdter Allee 45, 14195 Berlin
sayatz@fu-berlin.de

Abstract …

Keywords: univerbation, prototypes, production experiments, corpus data, German

1 The form and history of noun-verb units in German

2 The status of noun-verb units?

3 Corpus-based analysis of the usage of verb-noun units

4 Elicited production of noun-verb units in written language

5 Explaining the process of noun-verb univerbation

Acknowledgments

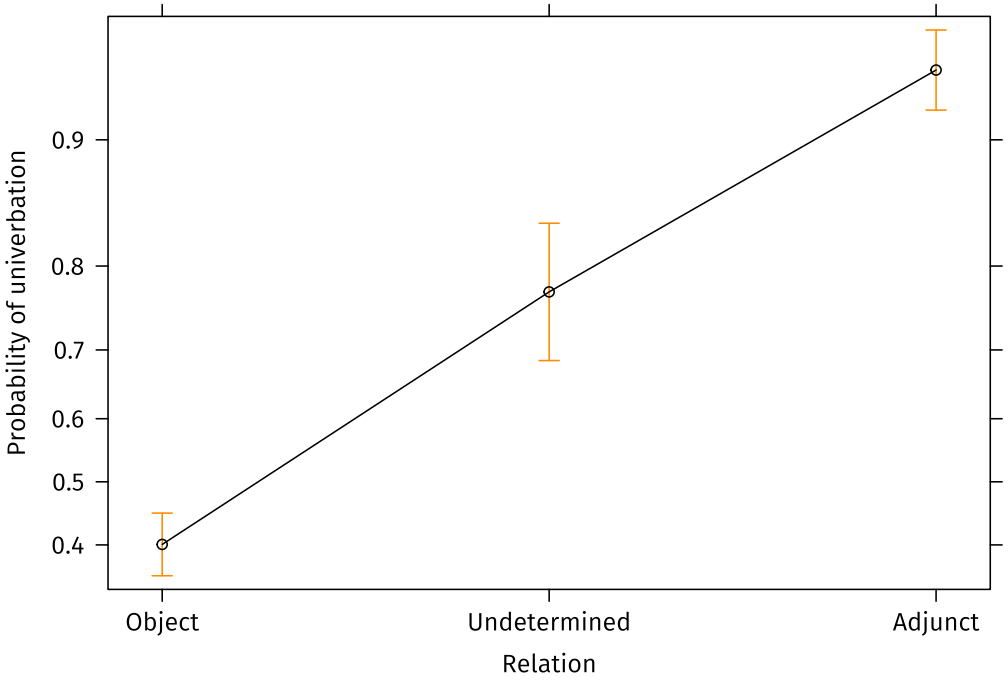

Figure 1: Effect plot for the regressor encoding the syntactic relation within the N+V unit in the GLMM modelling the corpus data.

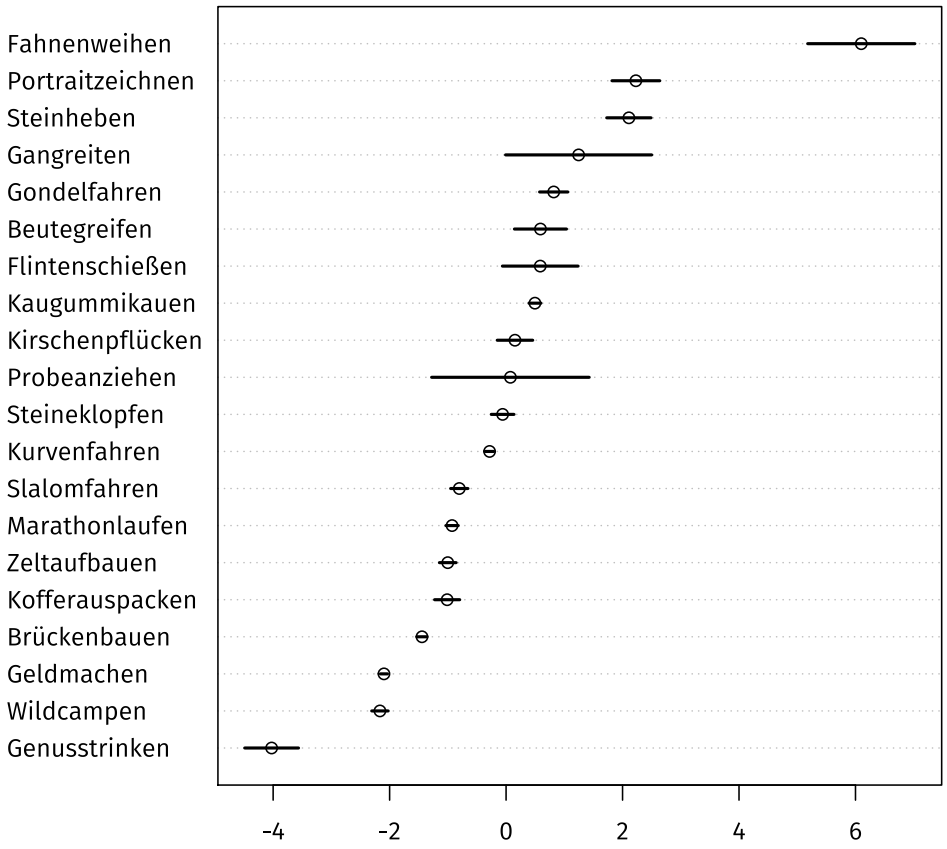

Figure 2: A random selection of conditional modes with 95% prediction intervals for the levels of the random effect in the GLMM modelling the corpus data.

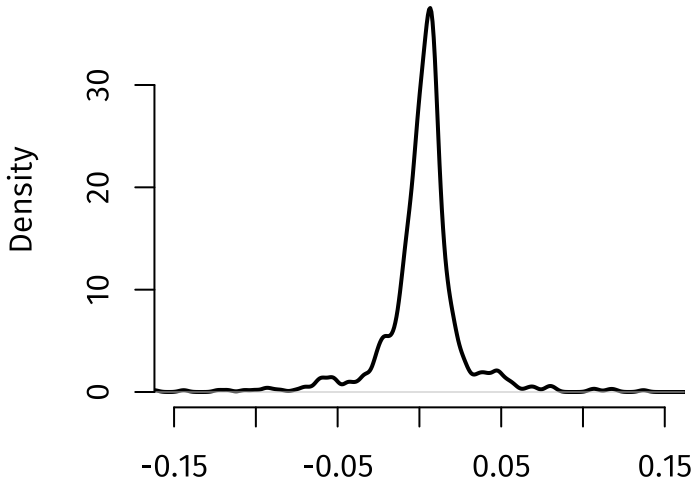

Figure 3: Distribution of the overall association scores (across all morphosyntactic conditions) with $n = 820$.

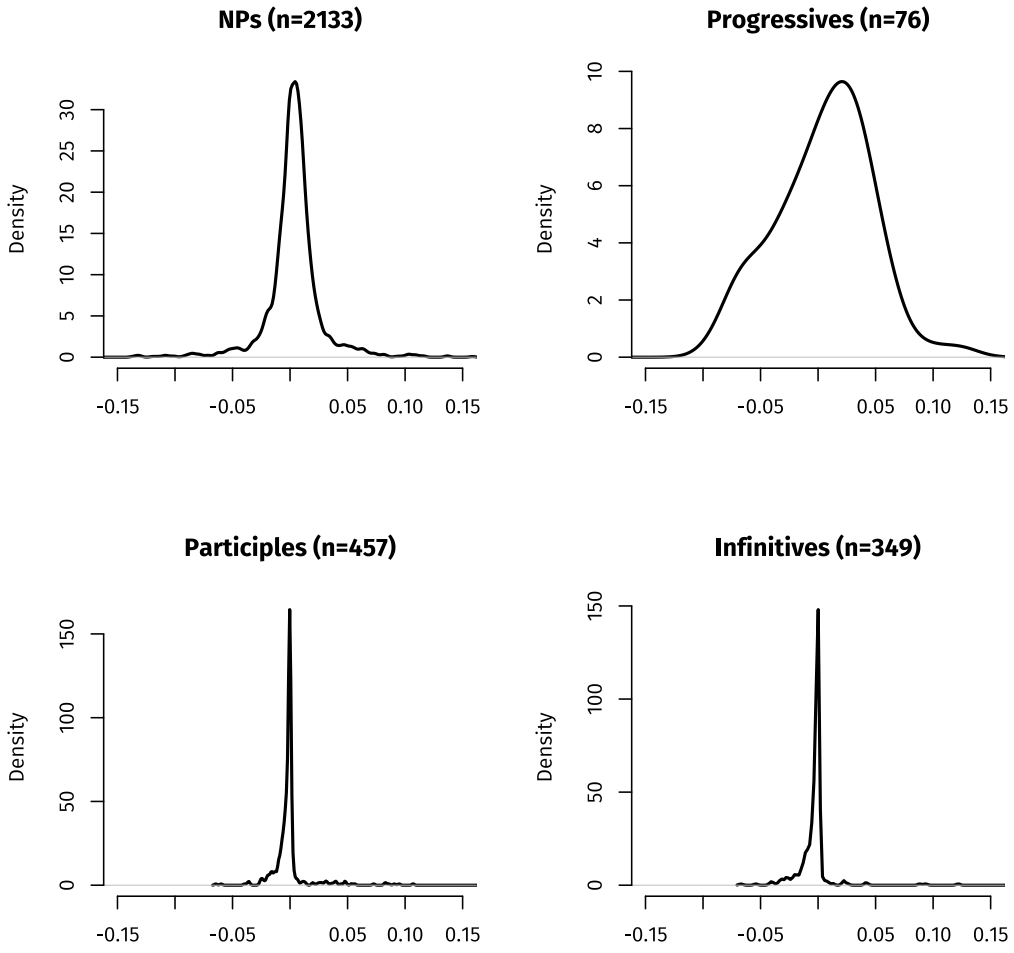

Figure 4: Distribution of the association scores in the specific morphosyntactic conditions; because of some undefined scores the sample sizes n vary.

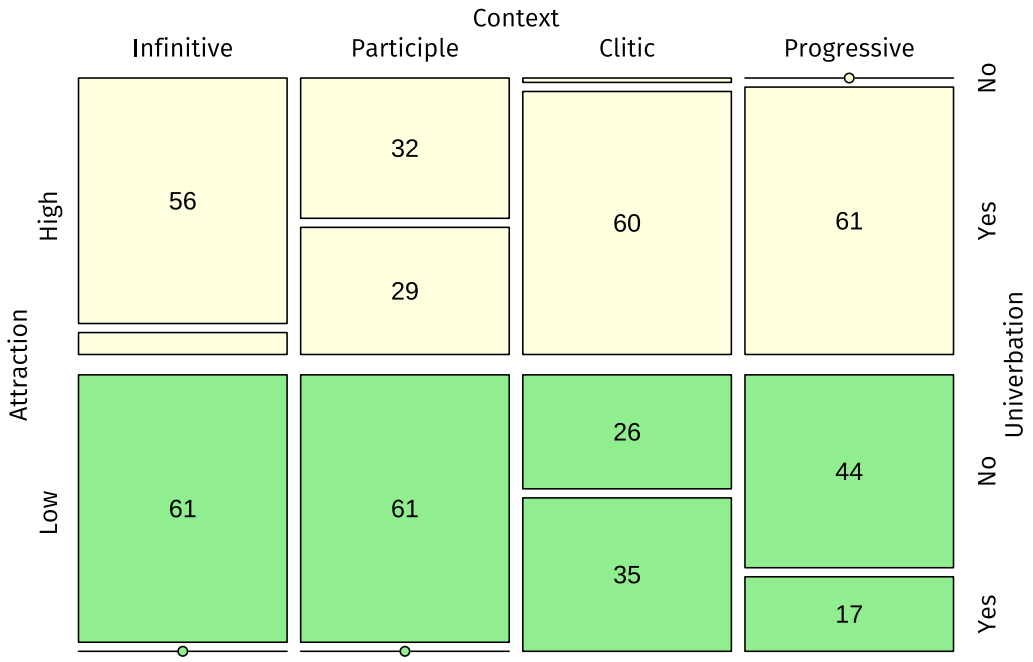

Figure 5: Mosaic plot of the responses in the production experiment (vertical right) grouped by the morphosyntactic context (horizontal) and the binned N+V unit's attraction strength calculated from the corpus (vertical left).

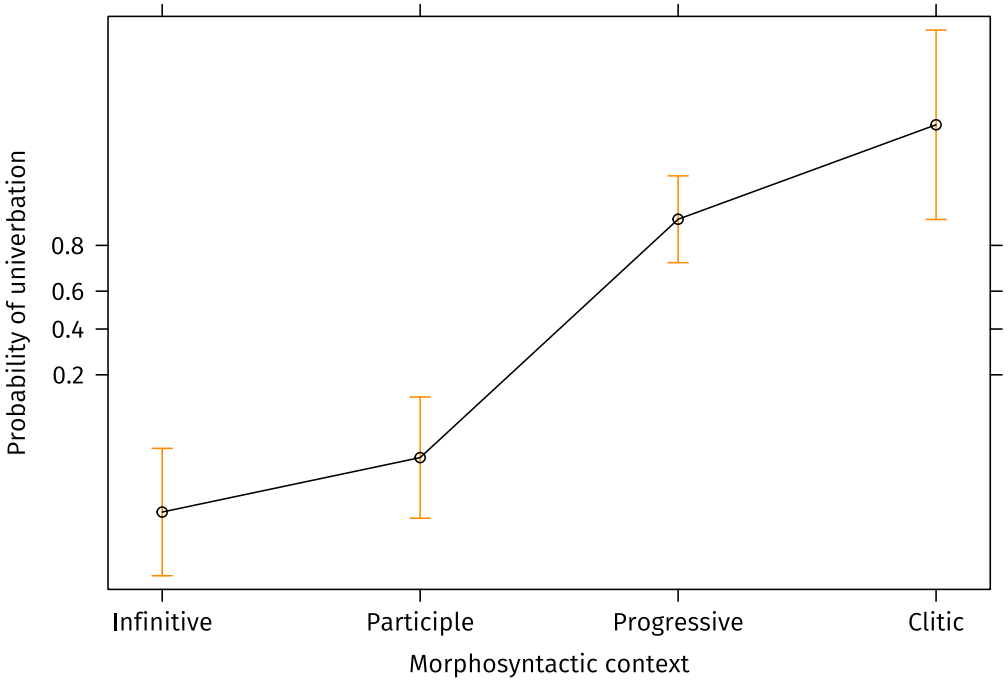

Figure 6: Effect plot for the regressor encoding the morphosyntactic context in the GLMM modelling the experimental data.

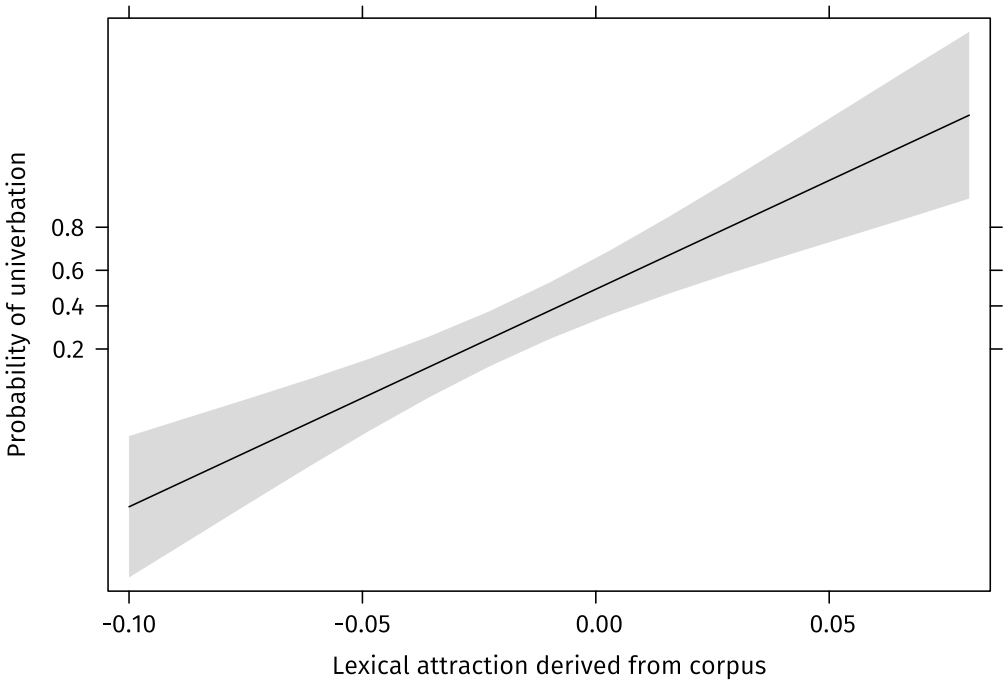

Figure 7: Effect plot for the regressor encoding the N+V unit's corpus-derived association with univerbation in the GLMM modelling the experimental data.

	Estimate	CI low	CI high
(Intercept)	-0.492	-0.716	-0.267
RelationUndetermined	1.623	1.138	2.107
RelationAdjunct	3.048	2.720	3.381
LinkbinaryYes	0.362	0.013	0.717

Table 1: Coefficient table for the corpus GLMM. Nakagawa & Schielzeth's $R^2_m = 0.299$ and $R^2_c = 1.000$. Random effect for V+N lemma: Intercept $= 4.593$, sd $= 2.143$.

V+N Unit	Association	Relation
Teilhaben	0.185	Object
Radfahren	0.180	Undetermined
Computerspielen	0.137	Adjunct
Zeitreisen	0.119	Adjunct
Skifahren	0.116	Adjunct
Autofahren	0.108	Undetermined
Probefahren	0.105	Adjunct
Bogenschießen	0.082	Undetermined
Schifffahren	0.080	Undetermined
Windsurfen	0.080	Adjunct

Table 2: Top ten V+N units with a strong tendency for univerbation.

V+N Unit	Association	Relation
Klavierspielen	0.009	Object
Theaterspielen	0.008	Undetermined
Filmemachen	0.008	Object
Autowaschen	0.007	Object
Zigarettenrauchen	0.006	Object
Haarewaschen	0.003	Object
Notenlesen	0.002	Object
Golfspielen	-0.001	Object
Wasserholen	-0.009	Object
Haareschneiden	-0.009	Object

Table 3: Top ten V+N units without any tendency for or against univerbation.

V+N Unit	Association	Relation
Gedankenmachen	-0.162	Object
Geldverdienen	-0.144	Object
Rechtgeben	-0.123	Object
Spaßhaben	-0.117	Object
Rechthaben	-0.107	Object
Kinderhaben	-0.101	Object
Zeitnehmen	-0.095	Object
Auftraggeben	-0.093	Object
Fehlermachen	-0.089	Object
Urlaubmachen	-0.085	Object

Table 4: Top ten V+N units with a strong tendency against univerbation.

	Estimate	CI low	CI high
(Intercept)	-3.960	-5.460	-2.790
AttractionNum	49.541	35.193	74.789
ContextParticiple	1.167	-0.324	2.614
ContextProgressive	6.273	4.730	8.249
ContextClitic	8.297	6.071	11.720

Table 5: Coefficient table for the experiment GLMM. Nakagawa & Schielzeth's $R^2_m = 0.804$ and $R^2_c = 0.897$. Random effect for participant: Intercept $= 2.967$, sd $= 1.723$.