
The curious case of wandering case morphemes

Stefan Hartmann
University of Bamberg

1 Introduction

This paper investigates non-canonical uses of the German circumposition *um ... willen* as well as the postposition *wegen*¹. Both adpositions usually govern genitive case, although there is quite some variation, especially in the case of *wegen*. (1) illustrates typical uses of both adpositions. However, in non-standard data, we also find numerous cases like (2), where the genitive morpheme *-s* is “relocated” to the postposition.

- (1) *um des Frieden-s willen* / *des Frieden-s wegen*
for the peace-GEN sake / the peace-GEN due-to
‘for the sake of peace’
- (2) *um des Frieden willen-s* / *des Frieden wegen-s*
for the peace sake-GEN / the peace due-to-GEN
‘for the sake of peace’

The aim of the present paper is to test a hypothesis that can account for this phenomenon, based on data from the 20-billion-word webcorpus DECOW16B (Schäfer & Bildhauer, 2012). Specifically, I argue that these variants can be explained by the principle of cleft-formation (see e.g. Ronneberger-Sibold 1997), which Nübling et al. (2017, 117) see as the most important syntax-typological feature of German. On this view, German is characterized by cleft structures both at the phrasal and at the sentence level.² It is often assumed that cleft-formation facilitates language processing by highlighting syntactic structures. This is highly compatible with evidence from psycholinguistic studies, according to which language users continuously predict what their interlocutor is going to say next (“forward modelling”, Pickering & Garrod 2013). Cleft structures offer a special potential for forward modelling as they invite language users to make predictions about the continuation of a phrase or sentence.

The cleft structure of noun phrases in German is particularly salient in genitive noun phrases with masculines or neuters belonging to the strong or mixed declension class, which have *-s* as genitive singular marker. Here, as Zimmer (2018, 67) points out, the rightmost element of the noun phrase is highlighted by the genitive-*s* that agrees with the determiner, e.g. *des neu-en Auto-s* ‘of the new car’. In a way, the *wegens* and *willens* constructions discussed in the present paper mimic this structure: By relocating the case morpheme to the rightmost element of the noun phrase, the cleft structure is made salient. As such, it seems plausible to assume that a principle like “the genitive-*s* occurs at the rightmost element of the cleft” is part of German native speakers’ implicit linguistic knowledge, which would offer an explanation for the phenomenon of the “wandering” genitive-*s* in the *um ... willen(s)* and *wegen(s)* constructions.³

¹*wegen* is also – and much more frequently – used as a preposition in modern German, but I focus on the postpositional use here: Prepositional *wegen* can arguably be considered a different construction, and the non-canonical use discussed here is more characteristic of the postpositional use.

²Ronneberger-Sibold (1997) uses the term “framing” for this phenomenon and “frames” for the cleft structures in question. As these terms are used in a different sense in various other domains, I avoid them here.

³I have borrowed the metaphor of “wandering” case morphemes from Eisenberg (2013, 173), who in turn paraphrases Fourquet (1973).

2 Data, methods, hypotheses and predictions

The DECOW16B corpus was searched exhaustively for instances of *wegens* and for instances of *um ... willens* in a distance of max. 6 words. After manual deletion of false hits and duplicates, 979 instances of *um ... willens* and 351 instances of postnominal *wegens* remained in the data. For comparison, samples of 5000 attestations each for *um ... willen* and *wegen* were extracted from DECOW16B. Again, false hits were deleted, and 4300 instances of *um ... willen* as well as (only) 115 instances of postnominal *wegen* remained in the data. In the remainder of this paper, I will collectively refer to *um ... willens* and *wegens* as the non-canonical constructions and to *um ... willen* and *wegen* as the canonical ones. The data were coded for a number of variables to be discussed in more detail below. The aim is to test two predictions which follow from the hypothesis that the drive towards cleft-formation is the main motivation for the displacement of the *s*-morpheme:

(i) There are significantly more masculine and neuter nouns in the *wegens*- and *um...willens*-data than in comparison datasets with the canonical variants, as the genitive-*s* only occurs in the genitive singular of strong masculines and neuters. (ii) We find a significantly higher proportion of non-canonical *s*-less genitives in the *wegens*- and *um...willens* data than in the comparison datasets: While *s*-omission is quite common especially in the case of low-frequency words, loan words, and proper nouns (Zimmer, 2018), we can predict that it occurs more frequently in combination with the non-canonically *s*-suffixed postpositions.

The hypotheses are tested using logistic regression modeling. For prediction (i), a mixed binomial logistic regression model is fit to the data, operationalizing “gender” (feminine / non-feminine) as a binary response variable. For prediction (ii), the method of CART trees and random forests is used, which is well-applicable to data with relatively few observations but many variables (Tagliamonte & Baayen, 2012, 161). The presence or absence of a genitive-*s* is used as the response variable (only strong and mixed masculine and neuter nouns enter the model, i.e. only nouns where a genitive-*s* is possible). As predictor variables, the factors that have emerged as significant predictors for *s*-lessness in Zimmer’s (2018) multifactorial study were used: a) whether the noun in the NP slot is inherently monoreferent (which is true for proper names but also for common nouns referring to unique entities like *Grundgesetz* ‘(German) constitution’), b) whether it is used as an apposition, e.g. *des Bundeskanzler(s) Adenauer* ‘of-the Federal Chancellor Adenauer’, c) whether it is a proper noun, a non-native word, or a short word (acronym/abbreviation) ending in *-s*. These lexemes often remain uninflected in German, a phenomenon sometimes referred to as *Schonungsbedarf* (roughly: ‘in need of conservation’, i.e. language users tend to “conserve” the phonological structure of the word), which is why Zimmer (2018) refers to this variable as “*Schonungsbedarf*+*s*”. In addition, d) frequency has a significant effect on *s*-lessness: Middle- and low-frequency words drop the genitive-*s* more often. To weigh the impact of these variables against that of the variant, the aforementioned predictors are complemented by the predictor “Variant” in the random forest model used here. Note that only datapoints with masculine or neuter nouns belonging to the strong or mixed declension classes were included in the model (as only they can vary between variants with and without a genitive-*s*). Also, some further datapoints had to be excluded because the respective lemma does not occur in the DECOW16B lemma frequency list, which was used to determine the frequencies of the individual lexemes.

3 Results

Let us discuss the results for each of the two predictions in turn. Table 1 shows the distribution of grammatical gender across the variants *um ... willen(s)* and *wegen(s)*.⁴ In line with prediction (i), the proportion of feminine nouns is much lower in the case of the non-canonical forms (6.6% vs. 37% for *willen(s)* and 6.6% vs. 59.1% for *wegen(s)*). The distribution is quite similar if every lemma type is only counted once. In simple binomial mixed regression models with “Gender” (f vs. m/n) as binary response variable, “Variant” as predictor variable, and “lemma” as random variable, “Variant” emerges as a highly significant predictor both for *um...willen(s)* (Estimate = 2.26, $\Pr(|z|) = 0.005^{**}$) and for *wegen(s)* (Estimate = 23.3, $\Pr(|z|) < 2e-16$). Turning to prediction (ii), the random forest models show that for both constructions, the variant makes a clear difference for the presence or absence of *s*-less genitives (see the CART trees in Figure 1). In the case of *um ... willen(s)*, the variables that proved most influential in Zimmer (2018) emerge as significant predictors of *s*-lessness in the canonical variant. For the non-canonical variant, by contrast, only frequency makes a difference. This is reflected in the measure of conditional permutation variable importance (Strobl et al., 2008): here, “Variant” emerges as the most significant predictor by far. In the case of *wegen(s)*, there are so few canonical uses in the dataset that for the canonical variant, no impact whatsoever of the aforementioned variables can be found⁵ This is also why frequency, rather than variant, emerges as the most significant predictor according to the variable importance measure. Still, the random forest model lends strong support to the hypothesis that the genitives of masculine and neuter nouns behave very differently in the non-canonical than in the canonical constructions.

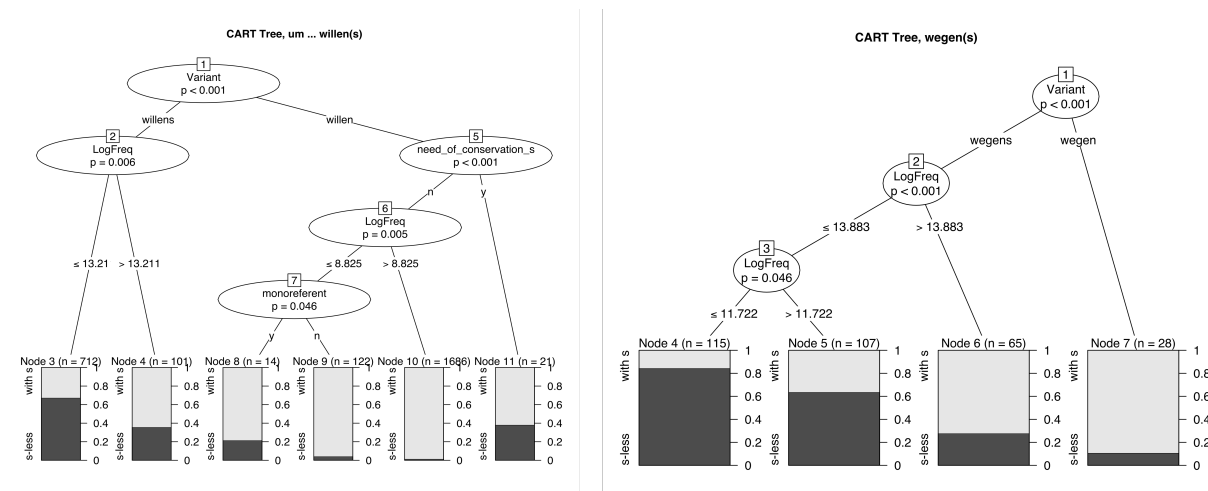


Figure 1: CART Trees for *um ... willen(s)* and *wegen(s)*. Response variable: *s*-lessness of genitives in masculines and neuters of the strong and mixed declension classes.

⁴Note that the numbers for *wegen* are based on a 5000-word sample, of which only the postnominal instances were used. This is why the numbers for *wegens* in the table are higher than those for *wegen*, but in fact the attestations for *wegen*, both pre- and postnominal, vastly outnumber those for *wegens*.

⁵They can, however, be found if one includes prepositional uses of *wegen*.

Variant	Freq in <i>willen</i> (Types)	Freq in <i>willens</i> (Types)	Freq in <i>wegen</i> (Types)	Freq in <i>wegens</i> (Types)
feminine	1344 (570)	61 (55)	68 (46)	23 (22)
masculine	1614 (294)	308 (89)	23 (22)	235 (125)
neuter	679 (208)	561 (304)	24 (10)	177 (177)

Table 1: Gender distribution

4 Conclusion

The principle of cleft-formation offers an adequate explanation for the emergence of the non-canonical variants discussed here. The distribution of grammatical genders in the *um ... willens* and *wegens* constructions support the hypothesis that the -s in the two postpositions actually has its origin in genitive markers that are either fully “relocated” to the rightmost element to the cleft (*um des Frieden willen-s*) or redundantly marked both on the head noun of the NP and the postposition (*um des Frieden-s willen-s*). However, a full account of the constructions discussed here would have to be more complex. For instance, phonological factors as well as the interaction with other, similar constructions seem to play a role as well and should be explored in more detail in future studies. As such, although *um ... willens* and *wegens* are quite infrequent constructions, they offer abundant research possibilities, and they show that small phenomena that can easily be overlooked provide a unique window into native speakers’ linguistic intuition.

References

- Eisenberg, Peter. 2013. *Grundriss der deutschen Grammatik. Bd. 1: Das Wort*. Metzler 4th edn.
- Fourquet, Jean. 1973. *Prolegomena zu einer deutschen Grammatik*. Schwann-Bagel 4th edn.
- Nübling, Damaris, Antje Dammel, Janet Duke & Renata Szczepaniak. 2017. *Historische Sprachwissenschaft des Deutschen: Eine Einführung in die Prinzipien des Sprachwandels*. Narr 5th edn.
- Pickering, Martin J. & Simon Garrod. 2013. An Integrated Theory of Language Production and Comprehension 36(4). 329–347.
- Ronneberger-Sibold, Elke. 1997. Typology and the diachronic evolution of German morphosyntax. In Jacek Fisiak (ed.), *Linguistic reconstruction and typology* (Trends in Linguistics 96), 313–335. Mouton de Gruyter.
- Schäfer, Roland & Felix Bildhauer. 2012. Building Large Corpora from the Web Using a New Efficient Tool Chain. In Cicoletta Calzolari, Khalid Choukri, Terry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of LREC 2012*, 486–493.
- Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional Variable Importance for Random Forests 9(307). <http://www.biomedcentral.com/1471-2105/9/307>.
- Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice 24(02). 135–178. doi:10.1017/S0954394512000129.
- Zimmer, Christian. 2018. *Die Markierung des Genitiv(s) im Deutschen: Empirie und theoretische Implikationen von morphologischer Variation*. De Gruyter.