

## Passives of reflexives: a corpus study

Sina Zarriß, Florian Schäfer, Sabine Schulte im Walde  
Universität Stuttgart

**Introduction** While reflexive (and reciprocal) verbs in most languages do not passivize, German is one of the few languages that forms passives of reflexive verbs, cf. example (1) (Plank, 1993; Vater, 1995; Agel, 1997; Mueller, 1999; Hundt, 2002; Schaefer, to appear).

- (1) Erst wird sich geküsst, dann wird geheiratet.  
First is REFL kissed, then is married.

A number of theoretical approaches predicts that every transitive verb that can be reflexivized and passivized can enter the formation of the Passive of Reflexive (PoR) in German. Others have claimed that the construction is restricted to inherently and naturally reflexive verbs. Based on the sdeWaC, a large web corpus for German that has been automatically annotated with syntactic relations, we identified 134 verbs that enter the formation of the PoR. From the same corpus, we extracted proportions of reflexive uses for transitive verbs in general. We find that occurrences of PoR verbs are strongly correlated with a large proportion of reflexive uses in the corpus. This result favors the hypothesis that the PoR is available only for inherently and naturally reflexive verbs.

**Background** From a theoretical perspective, the PoR raises questions about the status of Principle A of the Binding Theory (Chomsky, 1981). Given that the external argument is absorbed under passivization (Baker et al., 1989), how can the reflexive pronoun pass Binding Principle A? It has been proposed that the implicit argument in German passives, although absorbed, can nevertheless act as syntactic antecedent for anaphoric binding (Sternefeld, 2006). This proposal predicts that, in principle every transitive verb that (i) allows reflexivization and (ii) passivization can appear in the construction. However, it has been claimed that the construction is restricted and hosts mainly *inherently* and *naturally* reflexive verbs (Agel, 1997; Schaefer, to appear). If this restriction is empirically correct, then the above proposal that the implicit argument of passives acts as the syntactic antecedent of the anaphor is arguably wrong because Binding Principle A is not restricted by verb classes. Instead, such a restriction would suggest that a reflexive pronoun in PoRs is exempt from Binding Theory in some way that relates to the lexical semantics of the verbs involved (Agel, 1997; Schaefer, to appear).

**Defining natural reflexivity** According to Kemmer (1993), naturally reflexive verbs carry "... inherent in their meaning [...] the lack of expectation that the two semantic roles they make reference to will refer to distinct entities ...". Kemmer provides a set of semantic verb classes that fulfill this characterization (e.g. "grooming verbs" like *wash*; verbs of social or affectionate events like *meet*, *kiss*; verbs of antagonistic events like *fight*). Recently, Bouma and Spenader (2009) have shown for Dutch reflexive verbs that frequency measures can provide independent evidence that an individual verb is actually naturally reflexive or not. They tested the hypothesis that the verbs selecting the simple reflexive pronoun 'zich' in (2a) are naturally reflexive while verbs selecting the heavy reflexive pronoun 'zichzelf' in (2b) are not. The comparison was done by calculating the ratio of transitive disjoint vs. reflexive uses of these verbs. While they correlated the ratio to the use of a specific reflexive pronoun, we correlate the ratio to the use of the verb in the PoR.

- (2) a. Jan waste zich/??zichzelf/Mary  
John washed REFL/REFL-SELF/Mary  
'John washed/Mary.'  
b. Jan haat ??zich/zichzelf/Mary  
Jan hates REFL/REFL-SELF  
'John hates himself/Mary.'

**Corpus study** Our hypothesis is as follows. If all verbs that allow for reflexivization (and passivization) can potentially form the PoR, we expect that the proportion of reflexive uses of those verbs that actually occur in the PoR in a corpus is similar to the proportion of reflexive uses of verbs that we do not find in a PoR construction. The study relies on a 880 million word web corpus of German, the *sdeWaC* corpus (<http://wacky.sslmit.unibo.it/>) that has been automatically annotated with syntactic dependencies. From the corpus, we extracted (a) occurrences of passives of reflexives, (b) all active occurrences of all transitive verbs and their objects. We use (a) to obtain a sample of verbs that occur in the PoR, and (b) to calculate the ratio of reflexive uses over all transitive verbs. We limit our study to transitive verbs included in the verb lexicon of the German LFG grammar (Rohrer and Forst, 2006), to distinguish different readings of a verb lemma when analyzing its reflexivization properties (note that Bouma and Spender (2009) made the same observation for Dutch). As an example, consider the two readings of 'kümmern' illustrated in Example (3). Whereas the (3a) reading could be hardly used with a reflexive, (3b) is inherently reflexive.

- (3) a. Das Problem kümmert ihn nicht.  
The problem worries him not.  
'The problem does not worry him.'
- b. Jan kümmert sich um das Problem.  
Jan takes care REFL of the problem.  
'Jan takes care of the problem.'

To control for the problem of lexical ambiguity, we make the simplifying assumption that different readings of a given verb are reflected in different subcategorization frames. E.g. the inherently reflexive reading of *kümmern* in (3b) selects for an additional *um*-PP, in contrast to (3a).

**Data collection:** First, we automatically collected all instances of PoRs by searching for parse trees that have a reflexive pronoun with a passivized verb as its head. We found 901 candidate PoR sentences and 371 PoR verb lemmas. We manually filtered these instances to exclude sentences that had been extracted due to an incorrect syntactic annotation. We also excluded instances with a dative reflexive and with a passivized verb was a light verb part of an idiomatic expression (e.g. 'sich in den Weg stellen'). In addition to the filtering, we annotated for each verb in a PoR whether it occurred in a frame that selects additional PP arguments. From the remaining sentences, we obtained a list of 134 verbs forming the PoR. In a second step, we counted all occurrences of the transitive verbs in our lexicon, as well as their subset of reflexive uses. We followed previous research and only considered sentences with a personal pronoun as the subject. Finally, we applied a frequency cutoff and excluded verbs that occurred less than 10 times with a reflexive. The final list consists of 1,340 verbs that allow for reflexivization, with a subset of 126 PoR verbs.

**Analysis:** The graph in Figure 1 displays the final set of verbs sorted according to their reflexive ratio (we use the logarithmic ratio due to a skewed distribution). The points on the line correspond to the ratios of the PoR verbs. If our hypothesis was true, that all verbs that allow for reflexivization can also undergo the PoR, we would expect that the PoR verbs were evenly distributed over the graph. However, the PoR verbs clearly scatter at the top of the graph, demonstrating that in general PoR verbs have a higher ratio than expected from random sampling of the set of verbs. An additional correlation test established that there is indeed a significant correlation between the a verb being in the PoR subset and its ratio; the correlation coefficient is 0.28. Looking at the outliers of the PoR verbs which have a relatively low ratio, we noticed that most of the outliers seem to be naturally reciprocal verbs, such as *begrüßen*. A follow-up study will also take the idiosyncratic properties of the reciprocal verbs into account (e.g., by looking at plural subjects only). In addition to the strong tendencies we found for reflexivization, we observed an interesting effect concerning the tendency for passivization. In parallel to the ratio of reflexive uses, we computed ratios of passive uses for our set of 1,340 verbs: The average ratio of passive uses is significantly lower for the PoR verbs than for the entire set of verbs. This effect is shown in Figure 2. A follow-up study will investigate possible interpretations of this effect.

In sum, we reject our initial hypothesis that all verbs which undergo reflexivization can undergo the PoR. Instead, the construction seems to be restricted to verbs that have a high ratio of reflexive uses, i.e. they are naturally or inherently reflexive. Our analysis of empirical data thereby confirms a specific hypothesis about PoRs in the theoretical literature (Agel 1987, Schäfer to appear).

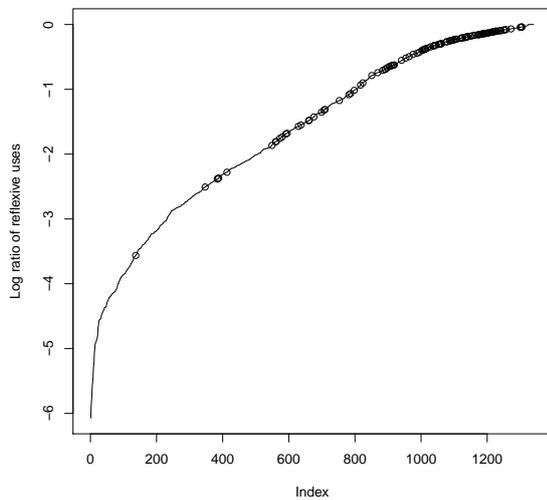


Figure 1: Ratios of reflexive uses

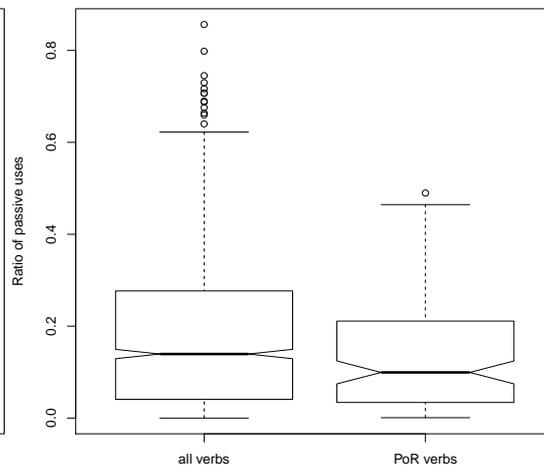


Figure 2: Ratios of passive uses

Agel, V. (1997). Reflexiv-Passiv, das (im Deutschen) keines ist. In Duerscheid, C., Ramers, K. H., and Schwarz, M., editors, *Sprache im Fokus*, pages 147–187. Niemeyer.

Baker, M., Johnson, K., and Roberts, I. (1989). Passive arguments raised. *Linguistic Inquiry*, 20:219–252.

Bouma, G. and Spender, J. (2009). Strong and weak reflexives in Dutch. In *Proceedings of the 7th TLT 2009*.

Chomsky, N. (1981). *Lectures on Government and Binding*. Foris.

Hundt, M. (2002). Formen und Funktionen des Reflexivpassivs im Deutschen. *Deutsche Sprache*, 30:124–166.

Kemmer, S. (1993). *The middle voice*. John Benjamins.

Mueller, S. (1999). *Deutsche Syntax deklarativ - Head Driven Phrase Structure Grammar fuer das Deutsche*. Niemeyer.

Plank, F. (1993). Peculiarities of Passives of Reflexives in German. *Studies in Language*, 17:135–167.

Rohrer, C. and Forst, M. (2006). Improving coverage and parsing quality of a large-scale LFG for German. In *Proceedings of LREC-2006*.

Schaefer, F. (to appear). The passive of reflexive verbs and its implications for theories of binding and case. *Journal of Comparative Germanic Linguistics*.

Sternefeld, W. (2006). *Syntax. Eine morphologisch motivierte generative Beschreibung des Deutschen*. Stauffenberg.

Vater, H. (1995). Zum Reflexiv-Passiv im Deutschen. In Popp, H., editor, *In Deutsch als Fremdsprache: an den Quellen eines Faches. Festschrift für Gerhard Helbig zum 65. Geburtstag*, pages 185–192. Iudicium.