A Treebank-based Investigation of IPP-triggering Verbs in Dutch and German.

Liesbeth Augustinus  Frank Van Eynde
liesbeth@ccl.kuleuven.be  frank@ccl.kuleuven.be

Centre for Computational Linguistics
KU Leuven

Abstract

Infinitivus Pro Particio (IPP), also known as Ersatzinfinitiv, refers to a syntactic construction in which an infinitival verb form occurs instead of the expected past participle. The IPP construction appears in a subset of the West-Germanic languages, such as Dutch (1) and German (2).

(1) NL Peter heeft het boek *kunnen/gekund(1) lezen.
   Peter has the book *can-Inf/could-PastP read
   ‘Peter was able to read the book.’

(2) DE Peter hat das Buch lesen *könennen/gekonnt.
   Peter has the book read *can-Inf/could-PastP
   ‘Peter was able to read the book.’

The examples indicate that the Dutch verb *kunnen ‘can’ and its German equivalent *könennen both trigger IPP. Moreover, the use of the infinitive is obligatory in those constructions. Although several similarities can be observed, both languages differ with respect to IPP in a number of ways. For example, Dutch perception verbs (e.g. *zien ‘see’, *horen ‘hear’) always trigger IPP (3), whereas the corresponding German verbs (*sehen, *hören) can take on the form of a past participle as well (4).

(3) NL Marie heeft haar *horen/gehoord zingen.
   Marie has her hear-Inf/heard-PastP sing
   ‘Marie has heard her sing.’

(4) DE Marie hat sie singen *hören/gehört.
   Marie has her sing hear-Inf/heard-PastP
   ‘Marie has heard her sing.’

Such variation makes IPP an interesting case for a comparative study between those languages. For both languages, the literature provides several lists of verbs that can take on the IPP form; see amongst others Haeseryn et al. (1997), Dudenredaktion (2006), Rutten (1991), and Schmid (2005). Although the set of IPP-triggering verbs is assumed to form a closed class (which might differ across languages and/or dialects), those lists are often incomplete. Moreover, the judgments on the grammaticality of the constructions under consideration are commonly based on the intuition of a handful of informants. Therefore, a corpus-based investigation of the IPP-triggering verbs should lead to a better understanding of this phenomenon.
In order to get a more exhaustive list of verbs occurring as IPP-triggers in the two languages, we have explored four treebanks: LASSY [van Noord et al., in press] and CGN [van der Wouden et al., 2003] for Dutch, and TüBa-D/Z and TüBa-D/S for German. Using treebanks instead of ‘flat’ corpora is particularly interesting for this topic, since it is hard to retrieve IPP constructions by means of a string search, as in main clauses the finite verb and the final verb cluster are often not adjacent. Furthermore, those treebanks are not only suitable for a comparative analysis of Dutch and German in general, since it is also possible to compare spoken (CGN and TüBa-D/S) and written data (LASSY and TüBa-D/Z).

Based on a general division into subject raising, subject control, and Accusativus cum Infinitivo verbs [Sag et al., 2003], we made a more fine-grained distinction along syntactic and semantic lines to account for both intralingual and interlingual differences between IPP-triggers and verbs that do not occur in IPP constructions. Those findings are supplemented with quantitative information to provide a general idea on the frequency of IPP-triggering verb patterns on the one hand, as well as a more detailed account on verbs which optionally trigger IPP on the other hand.

References


TüBa-D/S. Treebank of Spoken German. (http://www.sfs.uni-tuebingen.de/en/tuebads.shtml)

TüBa-D/Z. Treebank of Written German. (http://www.sfs.uni-tuebingen.de/en/tuebadz.shtml)