

Selectional Preferences for (Non)finite Structures as Indicator of Control Relations: A Cross-Germanic Corpus Study

Jutta M. Hartmann, Anne Mucha, Beata Trawiński, Angelika Wöllstein
(IDS Mannheim, Germany)

Background: This paper is concerned with the correlation between syntactic selectional preferences of clause-embedding verbs and their control behavior in a cross-Germanic perspective with special focus on German, Swedish and Dutch. It is well-known that, while pronominal subjects of finite complement clauses can easily refer to referents outside of the superordinate matrix clause, covert pronominal subjects of infinitival complements are usually dependent in their interpretation on arguments in the superordinate clause.

In German, finite complements are usually CPs whereas infinitival clauses can be structurally smaller (depending on the coherence/restructuring properties of the embedding predicate). Based on these premises, Wöllstein (2015) and Brandt et al. (2016) present corpus-based evidence (supported by experimental evidence and a comparison with Polish) showing that a certain class of clause-embedding verbs (e.g., *absegnen* 'to approve', *gutheißen* 'to endorse', or *anordnen* 'order' in (1)) systematically resist co-reference of a covert infinitival subject argument with the matrix subject (Anti-(Subject)-Control, henceforth AC).

- (1) Otto_i ordnete an, [_{CP} PRO_{i*/j} die Katze zu streicheln]. (Brandt et al. 2016, 81)
Otto ordered the cat to stroke

The crucial insight provided in these works for German is that the above mentioned AC predicates although they can embed infinitives, i) show a strong significant preference for finite over infinitival complements (cf. Figure 1) and ii) if they select for an infinitival complement, the structure is necessarily incoherent and the embedded infinitival clause is analyzed as a CP. This correlation between AC and CP selection is not specific to German: We provide data that shows that the selectional preferences for finite complements and the CP status of these complements also holds for other Germanic languages such as Swedish and Dutch.

In Swedish, both infinitival (2-a) and finite complements (2-b) are introduced by a complementizer (*att*). In addition, verbs can select for bare infinitives (as in (2-b), marked with italics).

- (2) a. Och glöm inte **att** kompostera kaffesumpen.
And forget.IMP not ATT.INF compost.INFV coffee.grounds
'And don't forget to compost the coffee grounds.'
b. Och det gör inget **att** du glömde *skicka julkort*.
And that do.PRS nothing ATT.FIN you forget.PST send.INFV christmas card
'And it doesn't matter that you forgot to send a Christmas card.'

Adopting from the literature (e.g., Platzack 1986, Kalm 2016) the assumption that *att* occupies the syntactic C-head and thus by its presence identifies finite and infinitival complements as CPs, we conducted a study in the 'Språkbanken' corpora to investigate the selectional preferences of embedding verbs with respect to finite complements with *att*, infinitival complements with *att*, and bare infinitives. As a central result, Swedish equivalents of German AC verbs strike out in that i) they show a remarkably low frequency of selected propositional complements when compared to other verb classes, and ii) their selected propositional complements are predominantly finite (cf. Fig. 3). Moreover, transitive verbs that semantically require subject control (e.g., *försöka* ('try'), *våga* ('dare')) overwhelmingly select for bare infinitives. Hence, under the assumption that bare infinitives do not project a CP, the Swedish corpus data support the idea of associating AC with CP selection and semantic dependence with 'small' structures.

We can also identify a group of AC predicates in Dutch. In parallel to the study on Swedish, we searched for clausal complements of AC verbs as well as other verb classes defined in the literature (specifically Broekhuis & Corver 2015) differentiating bare infinitives, *te*-infinitives, (*om*)-*te*-infinitives and *dat*-clauses using the LASSY Groot corpus (see van Noord et al 2006, 2013). The group of AC verbs i) very rarely occur with clausal complements, and ii) they show a clear preference for finite complements over infinitival ones, which hardly occur (even though they do allow them in principle)(see Fig. 4 AC). In contrast to these AC verbs, control verbs including the obligatory subject control verbs (e.g. *proberen* 'to try', *verzuimen* 'to fail') preferably occur with infinitival complements (both with and without the optional complementizer *om*, see Ctrl-I & -II in Fig.4). Thus, the Dutch data also support the generalization that AC verbs can be identified based on their selectional preferences.

Conclusion: Corpus evidence shows that in sharp contrast to regular control patterns, the complements of AC verbs are overwhelmingly realized as finite structures. This is in line with the prediction in Rapp et al. (2017) (see also Givón 1990, Wöllstein 2015) that infinitival forms are preferably used when the controller is uniquely

determined in the matrix clause. This unique identification is present with verbs that semantically require subject control, but not with AC verbs. Additionally, we observe that AC occurs with infinitivals with CP status, which can either be detected by its incoherent/non-restructuring behaviour or the presence of a complementizer in these infinitivals.

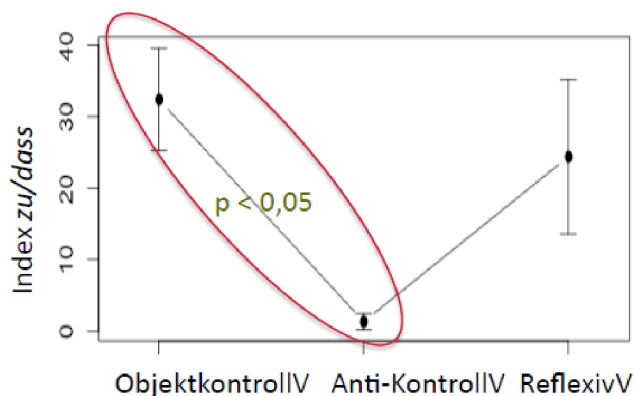


Figure 1: Correlation between the *zu*-index (= ratio of the total frequency of infinite *zu*-structures to the total frequency of finite *dass*-structures) and control type in German

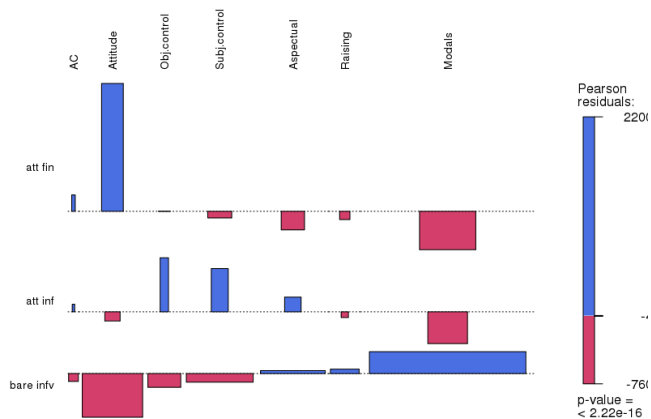


Figure 3: Association plot for the embedded clause type and the control type / verb type in Swedish

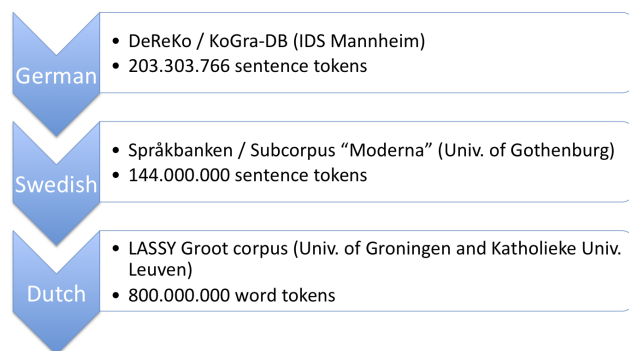


Figure 2: The corpora of German, Swedish and Dutch used in the present study

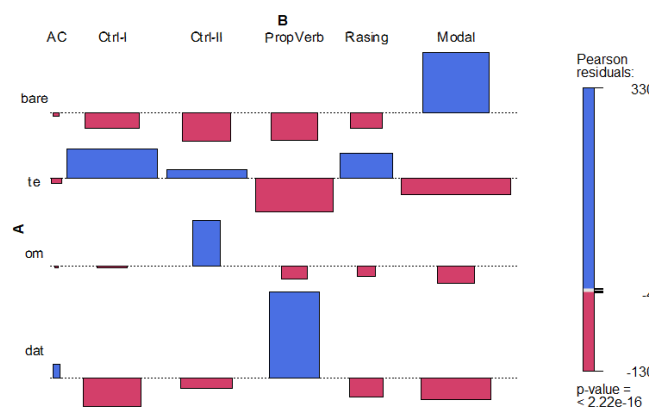


Figure 4: Association plot for the embedded clause type and the control type / verb type in Dutch

References

Brandt, P., B. Trawiński, and A. Wöllstein (2016): (Anti-)Control in German: Evidence from comparative, corpus- and psycholinguistic studies. In I. Reich and A. Speyer (Eds.), *Co-and Subordination in German and other Languages, Linguistische Berichte Sonderheft 21*, pp. 77–98. • **Broekhuis, H. and N. Corver (2015):** *Syntax of Dutch: Verbs and Verb Phrases: Volume 2 Comprehensive Grammar Resources*. Amsterdam: Amsterdam University Press. • **Givón, T. (1990).** *Syntax: A functional typological introduction*. Amsterdam: Benjamins. • **Kalm, M. (2016):** *Satsekvivalenta infinitivfraser i svenskan: En synkron och diakron undersökning* (Doctoral dissertation, Institutionen för nordiska språk, Uppsala universitet). • **van Noord, G., I. Schuurman & V. Vandeghinste (2006):** Syntactic annotation of large corpora in STEVIN. In: Proceedings of LREC'06, Genoa, Italy, 1811–1814. • **van Noord, G., G. Bouma, F. Van Eynde, D. de Kok, J. van der Linde, I. Schuurman, E. Tjong Kim Sang, & V. Vandeghinste (2013):** Large Scale Syntactic Annotation of Written Dutch: Lassy. In: P. Spyns & J. Odiijk (eds.) *Essential Speech and Language Technology for Dutch: resources, tools and applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 147–163. • **Platzack, C. (1986):** COMP, INFL, and Germanic word order. In *Topics in Scandinavian syntax, Studies in Natural Language and Linguistic Theory*, pp. 185–234. Dordrecht: Reidel. • **Rapp, I., K. Laptieva, A. Koplenig & S. Engelberg (2017):** Lexikalisch-semantische Passung und argumentstrukturelle Trägheit - eine korpusbasierte Analyse zur Alternation zwischen dass-Sätzen und zu-Infinitiven in Objektfunktion. *Deutsche Sprache* 45.3, pp. 193-221. • **Wöllstein, A. (2015):** Grammatik-explorativ. In Eichinger, L. (ed.), *Sprachwissenschaft im Fokus, Jahrbuch des Instituts für Deutsche Sprache*, 93–120. Berlin and New York: Mouton de Gruyter.