Reading non-canonical sentences in context: Identity vs. Poset

Yvonne Portele & Markus Bader (Goethe University Frankfurt)

Isolated non-canonical object-before-subject (OS) sentences usually cause comprehension difficulties compared to canonical sentences with a subject-before-object (SO) structure. As theories of discourse structure have shown, non-canonical structures are only licensed under certain discourse conditions. We investigated whether given objects - given either by identity or set-membership - ease the processing of German OS sentences. Results from a self-paced-reading study show that both discourse relations eliminate processing difficulties for non-canonical sentences. Their overall processing pattern does not differ from their canonical counterparts (apart from influences caused by divergent lexical items).

1 Introduction

Non-canonical sentences, such as active object-before-subject (OS) clauses in German, usually cause processing difficulties during online comprehension compared to canonical sentences with an active subject-before-object (SO) structure. This pattern has been detected in several languages for isolated sentences (e.g., Bader & Meng, 1999 for locally ambiguous structures in German; Hyönä & Hujanen, 1997 for unambiguous sentences in Finnish). It is well-known, however, that the occurrence of non-canonical structures must be motivated by discourse-pragmatic conditions. Theories of discourse structure (e.g., Birner & Ward, 1998; Lambrecht, 1996; Prince, 1981) highlight *givenness* – already being introduced via the previous discourse – and *partially ordered set relations* (*poset*) – standing in a poset relation to a referent already mentioned in the previous discourse (see (1) for a set/subset relation) – as some of the central factors in the licensing of non-canonical word order.

(1)   We don't get involved in *all* murders, *but this one we thought we 'd take a look at*.
        ["ABC World News Tonight"; taken from Birner & Ward, 1998, p. 220]

In line with these theories, there is psycho- and neurolinguistic evidence that supportive contexts, when contrasted with unsupportive ones, indeed attenuate processing difficulties for non-canonical sentences (e.g., Burmester, Spalek, & Wartenburger, 2014; Gattei, París, & Shalom, 2021; Kaiser & Trueswell, 2004; Schumacher & Hung, 2012).
For German, Speyer (2005) found that a general preference to fill the sentence-initial position in German with the topic (the backward looking center) is usually overridden when the sentence contains an element standing in a poset relation to the preceding context, because this element then occupies the initial position. Furthermore, Weskott, Hörnig, Fanselow, and Kliegl (2011) found in a self-paced reading study that possible comprehension disadvantages for non-canonical object-subject (OS) in comparison to canonical subject-object (SO) sentences disappear and can even be turned into processing advantages when the object stands in a poset relation to the context. The question whether an identity relation also alleviates possible disadvantages for German OS sentences has received far less attention (but see Experiment 3 in Weskott, 2003 for referential identity in a more complex discourse configuration). In an offline acceptability study, Bader and Portele (2021) found that OS sentences in which the object referent was given and realized as an NP with a demonstrative determiner were rated equally high as their SO counterpart. To our knowledge, corresponding online data are missing.

2 Experiment

The current experiment investigated the question whether appropriate discourse structure, manipulated via contexts, facilitates the online comprehension of non-canonical sentences in Ger-

Table 1: Example stimulus used in the experiment

Identity

Context	Letzte Woche am Donnerstag hat der Sponsor mit einem Torhüter telefoniert. *Thursday last week, the sponsor talked to a goalkeeper on the phone.*
SO target	Er hat diesen Torhüter zu einer Verlängerung seines Vertrages bewegt. *He persuaded this goalkeeper to extend his contract.*
OS target	Diesen Torhüter hat er zu einer Verlängerung seines Vertrages bewegt. *This goalkeeper, he persuaded him to extend his contract.*

Membership

Context	Letzte Woche am Donnerstag hat der Sponsor mit der Fußballmannschaft telefoniert. *Thursday last week, the sponsor talked to the soccer team on the phone.*
SO target	Er hat den Torhüter zu einer Verlängerung seines Vertrages bewegt. *He persuaded the goalkeeper to extend his contract.*
OS target	Den Torhüter hat er zu einer Verlängerung seines Vertrages bewegt. *The goalkeeper, he persuaded him to extend his contract.*
Question:	Hat der Sponsor mit einem Torhüter telefoniert? *Did the sponsor talk to a goalkeeper on the phone?*

man. More specifically, we asked whether an identity relation (not investigated so far) eases the processing of non-canonical sentences in a similar way as poset relations or whether poset relations (found to alleviate and even override disadvantages) constitute licensing conditions that are more helpful for the human parser than an identity relation. A poset relation was established by creating a set-membership constellation, i.e., by including a collective noun (e.g., *soccer team*; *employment agency*) and mentioning a member of this collective (e.g., *goal keeper*; *clerk*).[1] We used this relation instead of the more frequently investigated *part-whole* relation (e.g., Weskott et al., 2011) to extend the investigation of different poset relations.

2.1 Methods

2.1.1 Participants

45 native speakers of German were recruited via Prolific or participated for course credit.

2.1.2 Materials

Twenty experimental item sets consisting of a context sentence and a target sentence were created (see Table 1). We manipulated the two factors *Discourse Relation* (Identity vs. Membership) and *Word Order* (SO vs. OS). In the Identity conditions, a context sentence introduced two male referents. The subject of the context sentence was a definite NP (e.g., *the sponsor*). The object was introduced by using an indefinite NP (e.g., *a goalkeeper*). The subject of the context sentence was rementioned with a personal pronoun (*he*) as subject of the target sentences – either in sentence-initial (SO) or sentence-medial (OS) position. The object was rementioned in the target sentences by using a demonstrative NP including a lexical repetition of the co-referent noun (e.g., *this goalkeeper*) and consequently also appeared in sentence-initial (OS) or sentence-medial (SO) position. In the membership conditions, the same definite NP subject as in the identity conditions was used. Instead of a male character referent, the object was a set-denoting definite NP (e.g., *the soccer team*). The subject of the target sentences was again the personal pronoun (*he*) referring back to the previous subject and also appeared in sentence-initial (SO) or sentence-medial (OS) position. The object of the target sentences

---

[1]Although the set-membership relation is not a poset relation in a technical sense, it is usually subsumed under poset relations in the relevant literature, e.g., IS-A-MEMBER-OF in Ward and Prince (1991). Under the mereological approach to plurals and collectives (Link, 1983), there is a poset relation also in the technical sense in our membership condition.

was a definite NP denoting a member, a human character, of the set introduced in the previous sentence. It could appear either sentence-initially (OS) or in mid-sentence position (SO). Fifty-two filler items were included in the experiment.

2.1.3 Procedure

The 72 items were distributed onto 4 lists according to a Latin Square design. Materials were presented in a self-paced reading experiment (word-by-word moving-window presentation) conducted via Ibex farm (Drummond, Von Der Malsburg, Erlewine, Yoshida, & Vafaie, 2016). Eight of the experimental sentences were followed by a yes-no comprehension question.
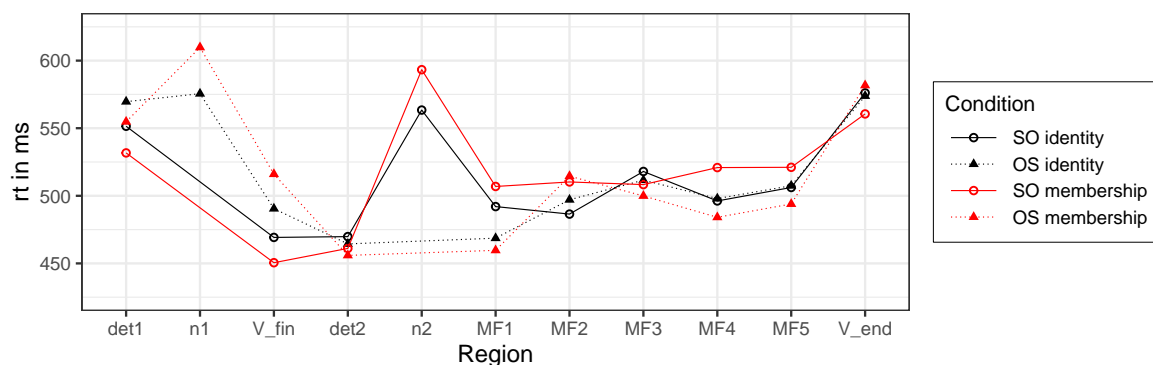
2.2 Results

Raw reading times are shown in Figure 1. Linear mixed-effects models showed significant effects at the positions n1 (first noun), V_fin (finite verb), det2 (second determiner) and MF 1 (first position of the midfield). At n1, there is a significant effect of Discourse Relation. Identity relations were processed faster than membership relations. In the identity relation, this is where the lexical item from the previous sentence is repeated. In the membership relation, this is where the noun denoting a singled-out member is mentioned for the first time. This difference is also visible as a spill-over effect in the form of a significant interaction at the following position V_fin. At det2, there was an effect of Word Order as well as Discourse Relation. OS sentences were processed faster. This difference stems from processing the lexical items for the personal pronoun *he* (er) in OS conditions vs. a determiner in SO conditions. The processing advantage for membership over identity relations at this point reflects the lexical difference between the articles *this* (dieser) in identity relations vs. *the* (den) in membership relations. This difference is still visible at the position of MF1, where there was an effect of Word Order with SO sentences being processed faster than their OS counterpart. No significant differences were detected in later regions within the sentence. Mean accuracy in the question comprehension task was 89.7%. There were no significant differences between conditions.

3 Discussion

We found several effects associated with different lexical items. We see two non-exclusive reasons for the effect for OS sentences found at the position of the first noun (n1), which was not significant for SO sentences at the position of n2. This effect can represent a processing advantage due to the lexical repetition of the noun in identity conditions or a processing disadvantage for having to infer the noun in membership conditions. We are currently running further self-paced reading studies investigating this issue by contrasting a poset relation with an identity relation established without lexical repetition, i.e., by using synonym expressions. We will compare this study with the current results to disentangle shortcomings of the current

Figure 1: Raw reading times in the different conditions. Note: There are no n1 and n2 data points in the respective SO or OS conditions since the subject was the one word personal pronoun *he* (er) whereas the object was a two word NP (determiner + noun).

experiment (such as the confound of discourse relation and article of the noun phrase). Overall, our results suggest that both discourse relations investigated in this study – identity and membership – alleviate potential processing difficulties of non-canonical OS sentences compared to their canonical SO counterpart in German. Self-paced reading patterns did not differ apart from lexical influences. Furthermore, there was no distinct advantage for a poset relation over an identity relation. We leave it to future work to investigate further discourse relations and their effect on the processing of non-canonical sentences. In addition, investigating different manifestations of *givenness* (e.g., situationally inferable vs. standing in a poset relation) will reveal important insight on the role of discourse licensing of non-canonical sentences.

References

Bader, M., & Meng, M. (1999). Subject-object ambiguities in German embedded clauses: an across-the-board comparison. *Journal of Psycholinguistic Research*, *28*, 121-143. doi: 10.1023/A:10232062081

Bader, M., & Portele, Y. (2021). Discourse and form constraints on licensing object-first sentences in German. *Languages*, *6*(2), 82. doi: 10.3390/languages6020082

Birner, B. J., & Ward, G. (1998). *Information status and noncanonical word order in English*. Amsterdam/Philadelphia: John Benjamins.

Burmester, J., Spalek, K., & Wartenburger, I. (2014). Context updating during sentence comprehension: The effect of aboutness topic. *Brain and language*, *137*, 62–76.

Drummond, A., Von Der Malsburg, T., Erlewine, M. Y., Yoshida, F., & Vafaie, M. (2016). *Ibex Farm*. Retrieved from `https://github.com/addrummond/ibex`

Gattei, C. A., París, L. A., & Shalom, D. E. (2021). Information structure and word order canonicity in the comprehension of spanish texts: An eye-tracking study. *Frontiers in Psychology*, *12*.

Hyönä, J., & Hujanen, H. (1997). Effects of case marking and word order on sentence parsing in finnish: An eye fixation analysis. *The Quarterly Journal of Experimental Psychology: Section A*, *50*(4), 841–858.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*, 113–147. doi: 10.1016/j.cognition.2004.01.002

Lambrecht, K. (1996). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press. doi: https://doi.org/10.1017/CBO9780511620607

Link, G. (1983). The logical analysis of plurals and mass terms: a lattice-theoritical approach. In R. Bäuerle, S. Christoph, & A. von Stechow (Eds.), *Meaning, use, and interpretation of language* (p. 302-323). Berlin & New York: Walter de Gruyter.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (p. 223-255). New York etc.: Academic Press.

Schumacher, P. B., & Hung, Y.-C. (2012). Positional influences on information packaging: Insights from topological fields in german. *Journal of Memory and Language*, *67*(2), 295–310.

Speyer, A. (2005). Competing constraints on Vorfeldbesetzung in German. In *Proceedings of the Constraints in Discourse Workshop* (pp. 79–87).

Ward, G. L., & Prince, E. F. (1991). On the topicalization of indefinite NPs. *Journal of Pragmatics*, *16*(2), 167–177.

Weskott, T. (2003). *Information structure as a processing guide: The left periphery of German verb-second sentences and its interpretation in context* (Unpublished doctoral dissertation). Universität Leipzig.

Weskott, T., Hörnig, R., Fanselow, G., & Kliegl, R. (2011). Contextual licensing of marked OVS word order in German. *Linguistische Berichte*, *225*, 3–18.