The interaction of referential form and word order

Yvonne Portele, Markus Bader (Goethe University Frankfurt)
Portele@lingua.uni-frankfurt.de

Research on word-order has shown that a referent's discourse status (e.g., topichood, givenness) is an important determinant of how early within a clause the referent is mentioned. More recently, corpus studies and acceptability experiments investigating the choice between subject-object (SO) and object-subject (OS) word order in German (Bader & Portele, 2019, 2021) indicate that the choice of a particular word order and the choice of a referential expressions are not independent from each other. Based on these findings, we hypothesize that arguments realized as demonstrative NPs preferentially occur in sentence-initial position because they are close to their antecedent in this position. Two experiments tested this hypothesis.

In Experiment 1, 18 participants read 24 contexts consisting of three sentences (see (1)). After each context, participants saw a picture together with a verb (see Figure 1) and had to describe the picture in a single sentence using the given verb. All verbs required an agent and a patient argument. The preceding context established either the agent or the patient as sentence topic. In the second context sentence, the topic was referred to either by a pronoun or a demonstrative NP. All target sentences produced by the participants were transcribed and scored according to word order and the referential expressions used for subject and object. About 95% of all pictures were described with an active SO sentence (e.g., *Er hat ihn untersucht* 'He examined him.'). Table 1 shows the four most frequent combinations of referential expressions for subject and object, depending on whether the agent or the patient was the topic. Unlike the factor Topic, the factor Referential Form (pronoun or demonstrative NP in context sentence 2) had no significant effect. Major findings visible in Table 1 include: (i) Over 95% of all sentences are in accordance with Rule 1 of Centering Theory ('A non-topic may be pronominalized only when the topic is also pronominalized'; Grosz et al., 1995). (ii) While subject and object were often both realized as pronouns with the subject as topic, this happened rarely when the object was the topic. This probably reflects an ambiguity avoidance strategy: having both subject and object pronominalized leads to systematic misinterpretations when the object is the topic. (iii) Whereas the subject was realized quite often as a demonstrative with an object topic, the object was realized as demonstrative only rarely with a subject topic. Given the very strong SO preference, this is in accordance with our hypothesis that demonstratives prefer the sentence initial position.

The few OS sentences that were produced in Experiment 1 all had a demonstrative object. Because of the low number of demonstrative NPs in object function in Experiment 1, we ran a second experiment that used a word-order choice procedure (e.g., Bresnan & Ford, 2010). 27 participants read the context sentences from Experiment 1. After each context, they had to choose between two alternative realizations of the final sentence. This experiment varied two factors, as shown in (2). First, the choice was either between an OS and an SO sentence ((2-a) or (2-b)) or between an OS sentence and a passive sentence ((2-a) or (2-c)). Second, the patient argument was realized as a demonstrative or a definite NP. Figure 2 shows how often the OS variant was chosen in Experiment 2. When SO was the competitor to OS, OS was chosen in about 40% when the patient object was a demonstrative but only in 10% when the patient was a definite NP. When a passive was the competitor to OS, OS was chosen in about 55% when the patient object was a demonstrative but only in 30% when the patient object was a definite NP. In sum, the results show two competing preferences: a default preference for subject-initial sentences and a preference for sentence-initial demonstrative NPs.

Taken together, Experiments 1 and 2 suggest that during language production referential expressions and syntactic structures mutually constrain each other. We will discuss how recent models of reference production and syntactic encoding can be united in order to account for these mutual influences.
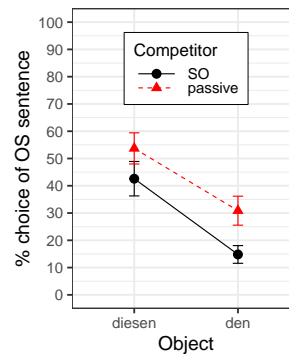
Figure 1: Example picture from Experiment 1

Figure 2: Results of Experiment 2

(1)   a.   **Topic = Agent**
In unserem Stadtviertel gab es einen sehr guten Arzt. Er/Dieser Arzt konnte bei fast allen Problemen helfen. Einmal musste er einen scheinbar schwerhörigen Klavierlehrer behandeln.
'A very good doctor was practicing in our quarter. He/This doctor could help with almost all problems. Once he had to treat a seemingly hearing-impaired piano teacher. '

         b.   **Topic = Patient**
In unserem Stadtviertel gab es einen sehr guten Klavierlehrer. Er/dieser Klavierlehrer hatte eine Zeit lang Probleme beim Hören. Irgendwann suchte er einen sehr angesehenen Ohrenarzt auf.
'A very good piano teacher was living in our quarter. He/This piano teacher was having hearing problems for quite a while. Once he visited a respected ear specialist.'

(2)   In unserem Stadtviertel gab es einen sehr guten Arzt. Dieser Arzt konnte bei fast allen Problemen helfen. einmal musste er einen scheinbar schwerhörigen Klavierlehrer behandeln.
'A very good doctor was practicing in our quarter. This doctor could help with almost all problems. Once he had to treat a seemingly hearing-impaired piano teacher '

         a.   **OS:** Diesen/Den Klavierlehrer hat der Arzt ziemlich lange untersucht.
'This/The piano teacher, the doctor examined for a long time.'

         b.   **SO:** Der Arzt hat diesen/den Klavierlehrer ziemlich lange untersucht.
'The doctor examined this/the piano teacher for a long time.'

         c.   **Passive:** Dieser/Der Klavierlehrer wurde von dem Arzt ziemlich lange untersucht.
'This/The piano teacher was examined by the doctor for a long time.'

Table 1: The four most frequent combinations of referential expressions used in Experiment 1 for subject and object, depending on whether the subject or the object was the topic.

Top=Subj	%	Top=Obj	%
S:pro O:def	31.429	S:dem-pro O:pro	34.807
S:pro O:pro	25.143	S:def O:def	32.044
S:def O:def	21.143	S:def O:pro	17.127
S:pro O:dem-pro	12.571	S:pro O:pro	4.972

References

Bader, Markus & Yvonne Portele. 2019. Givenness and the licensing of object-first order in German: The effect of referential form. In Anja Gattnar, Robin Hörnig, Melanie Störzer & Sam Featherston (eds.), *Proceedings of Linguistic Evidence 2018. Experimental Data Drives Linguistic Theory*, 208–228. University of Tübingen, online publication system.

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

Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1). 168–213. doi:https://doi.org/10.1353/lan.0.0189.

Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21. 203–225.