

*Markus Bader, Vasiliki Koukouloti*

# **When Object-Subject Order is Preferred to Subject-Object Order: The Case of German Main and Relative Clauses<sup>1</sup>**

**Abstract** Overall, subject-before-object (SO) order is preferred in German to object-before-subject order (OS), as reflected in higher acceptability and higher frequency of the former in comparison to the latter. Certain conditions have been identified, however, where OS order is preferred to SO order. First, main clauses in which the object is related to the prior discourse by a partially-ordered set relation, and second, relative clauses with a personal pronoun as subject. In order to explore the circumstances under which OS is preferred to SO order, we present preliminary data from ongoing corpus studies investigating relative clauses and main clauses in which either the subject or the object occupies the prefield. The corpus data confirm prior findings from experimental studies and extend them in several ways. In particular, the corpus data reveal a close connection between referential form and word order, with demonstrative pronouns strongly favoring the use of OS order.

**Keywords** German syntax, word order, prefield, relative clauses, referential form, language production, topic

## 1 Introduction

Although German is considered as a language with relatively free word-order, sentences in which the subject precedes the object(s) occur with a much higher frequency than sentences in which the subject follows one or more objects (Hoberg 1981; Kempen & Harbusch 2005; Bader & Häussler 2010). Not at least for this reason subject-before-object order (SO) is generally considered to be the

1 This work was supported by the Deutsche Forschungsgemeinschaft (Project VER within the Research Unit 1783 “Relative Clauses”). We would like to thank Yvonne Portele, Alice Schäfer and an anonymous reviewer for their helpful comments.

canonical order of subject and object in German. Although clauses deviating from the canonical order – that is, clauses with object-before-subject order (OS) – are much less frequent when considered across the board, under certain circumstances, OS order seems to be preferred to SO order.

Many studies on language processing across a variety of languages have found that sentences with non-canonical argument order are more difficult in comparison to sentences with canonical argument order. Sentences with OS order and passive sentences are acquired later than sentences with SO order (e.g., Friedmann et al. 2009), they are less often produced than sentences with SO order (e.g., Bader & Häussler 2010), they pose severe problems for people with aphasia (e.g., Burchert et al. 2008), and even for adult speakers without any language disturbance, they are often more difficult to comprehend than corresponding SO sentences (e.g., Kaiser & Trueswell 2004).

In some cases, however, the disadvantage for non-canonical sentences vanishes or is even reversed to an advantage. Weskott, Hörnig, Fanselow & Kliegl (2011) coined the terms *weak* and *strong licensing* of the OS order for such cases. Weak licensing refers to the situation where the SO and the OS variant of a sentence are equivalent with regard to measures like acceptability and processing complexity. Strong OS licensing, on the other hand, obtains when the OS order is at an advantage in comparison to the SO variant. In the following, we will focus on corpus frequencies as an indicator of strong or weak OS licensing, but other measures will also be taken into account if available.

A prominent case of strong OS licensing identified by corpus linguistic studies of German word order is illustrated by the two sentences in (1), which differ only with regard to the order of subject and object.

- (1) a. Wahrscheinlich wird der Fehler dem Lehrer entgehen.  
 likely will the teacher the error miss  
 ‘The teacher will probably miss the error.’  
 b. Wahrscheinlich wird dem Lehrer der Fehler entgehen.  
 likely will the teacher the error miss  
 ‘The teacher will probably miss the error.’

The sentences in (1) contain a non-agentive verb, an inanimate subject and an animate object. Given this particular configuration of verb semantics and animacy of the arguments, sentences with OS order occur with higher frequency than sentences with SO order (Hoberg 1981; Bader & Häussler 2010; Verhoeven 2015), and they receive higher ratings in acceptability experiments (Ellsiepen & Bader 2018).

In addition to verb-semantics and animacy, which together comprise the class of lexical-conceptual factors, factors concerning the discourse status of

the individual NPs are known to affect the choice between SO and OS order. In comparison to the lexical-conceptual factors discussed above, discourse-related factors have received less attention in corpus studies on German word order, in particular with regard to the issue of strong and weak OS licensing. However, as will be discussed below, at least two cases have been identified in the experimental literature, one concerning relative clauses and one concerning main clauses with either the subject or the object occupying the prefield.

Because main clauses and relative clauses differ in many ways, they may seem like an odd pair as far as the choice between SO and OS order is concerned. The most important difference in the current context concerns the degree of optionality with regard to the order of subject and object. Since it is obligatory to front relative pronouns in German, OS order can be obligatory for relative clauses in some cases, as in the example in (2).

- (2) Das ist der Lehrer, dem das Buch gefallen hat.  
 this is the teacher who the book pleased has  
 'This is the teacher who the book pleased.'

Given the meaning expressed by sentence (2), the relative clause must occur with OS order. The object relative pronoun must occur clause-initially and a verb like *gefallen* ('to please') cannot be passivized. It is therefore not possible to turn the dative object into a subject, thereby producing a subject-initial relative clause instead of an object-initial one. In this respect, sentences as in (2) contrast with sentences containing a relative clause that allows passivization, as shown in (3). Here, instead of producing an object-initial relative clause with the verb in the active voice, a subject-initial relative clause with a verb in the passive voice can be produced as an alternative.

- (3) a. Das ist der Lehrer, den der Schüler begrüßt hat.  
 this is the teacher who the student greeted has  
 'This is the teacher who the student greeted.'  
 b. Das ist der Lehrer, der von dem Schüler begrüßt wurde.  
 this is the teacher who by the student greeted was  
 'This is the teacher who was greeted by the student.'

In main clauses, in contrast, there is almost always a choice between putting the subject or the object into the prefield. In (4), for example, the same truth-conditional meaning can be expressed either with SO order (4a) or with OS order (4b).

- (4) a. Peter hat diesen Film schon zweimal gesehen.  
 P. has this movie already twice seen  
 ‘Peter has already seen this movie twice.’
- b. Diesen Film hat Peter schon zweimal gesehen.  
 this movie has P. already twice seen  
 ‘Peter has already seen this movie twice.’

To say that the order of subject and object is optional in (4) is not to say that the two orders can be freely exchanged in all contexts. Quite to the contrary, it is a truism that in most cases of word order optionality, the alternative orders are associated with different usage conditions. Starting with the seminal work of Lenerz (1977) and Höhle (1982), the pragmatic conditions that license the use of SO or OS order have been the topic of extensive research. The major insight emanating from this research is that SO sentences are typically (relatively) unrestricted with regard to discourse conditions, allowing uses with both wide and narrow focus, whereas OS sentences typically require narrow focus on one of their constituents. The exact conditions vary depending on whether both subject and object are contained within the middle field or whether one has been moved to the prefield, as in the examples in (4) (see Frey 2004b for the relationship between these two cases). In the following, we will consider only main clauses with either subject or object in the prefield. The question of whether to choose SO or OS order therefore boils down to the question of whether to move the subject or the object to the prefield.

This paper presents data from three ongoing corpus studies, one investigating relative clauses and two investigating main clauses. The reason for investigating these two clause types relates to the issue of strong versus weak OS licensing. Two questions will be pursued in this regard: first, can the experimental findings concerning the conditions that weakly or strongly license the use of OS order be replicated when looking at written language production, and second, can the set of conditions leading to weak or strong OS licensing be extended? Relative clauses are discussed in more detail in the next section. Afterwards, we turn to main clauses. In the final section, we discuss the implications of our findings for future research.

## 2 Non-canonical order in relative clauses

Relative clauses have played a major role in research on language acquisition and language processing, both disturbed and undisturbed. A common finding of this research is that subject-initial relative clauses as in (5a) are easier than object-initial relative clauses as in (5b), for both children and adults.

- (5) a. The gardener  $who_i t_i$  contacted the reporter left early.  
 b. The gardener  $who_i$  the reporter  $t_i$  saw left early.

Starting with Fox and Thompson (1990), it has become clear that object relatives are not in general more difficult than subject relatives. Based on an investigation of naturally occurring relative clauses, Fox and Thompson (1990) showed that object relative clauses occur particularly often with a personal pronoun as subject. Thus, in contrast to relative clauses in which the second NP is a lexical NP, as in (5), object relatives prevail when the second NP is a pronoun, as in the following example.

- (6) a. The gardener  $who_i t_i$  contacted me left early.  
 b. The gardener  $who_i$  I contacted  $t_i$  left early.

Later research has extended this finding to language comprehension (Mak et al. 2008) and to language acquisition (Kidd et al. 2007). For both English and German child language, Kidd et al. (2007) present corpus counts as well as experimental evidence showing that the large majority of object relative clauses produced by children has a pronoun as subject.

What has not been shown so far is whether the same also holds for adult German language production. In an ongoing corpus study of written German relative clauses, we are currently analyzing a set of about 1700 relative clauses randomly drawn from the deWac corpus (Baroni et al. 2009). In this paper, we present selected preliminary results concerning the distribution of the referential form of the second NP in subject and object relative clauses.

644 relative clauses contained both a subject and a direct object with either one being realized as relative pronoun. Of these, 547 relative clauses (85%) were subject-initial and 97 relative clauses (5%) were object-initial. Overall, subject relatives clearly outweigh object relatives. We classified the second argument of each relative clause – that is, the object in subject relatives and the subject in object relatives – with regard to the type of NP, using the same categories as Kidd et al. (2007): first-person pronoun, second-person pronoun, third-person pronoun, proper name, lexical NP, and others. We used one additional category not used by Kidd et al., namely reflexive pronouns. Table 1 shows the distribution of the relative clauses according to the NP type of the second NP, depending on whether the relative clause occurred with SO or OS order.

In subject relative clauses, the second NP is a lexical NP most of the time. Reflexives also occur with some regularity, whereas all other categories are quite rare. For object relatives, in contrast, the second NP is a personal pronoun in the majority of cases, with third- and first- person pronouns occurring most often and with about equal frequency. Lexical NPs also appear as second NP

Table 1: Percentages of relative clauses with different types of the second NP, depending on the syntactic function of the relative pronoun. Raw numbers are given in parentheses.

	Lexical NP	Reflexive	3ps pro	Proper Name	1ps pro	2ps pro	Other
Subj. relative	71.8 (392)	19.8 (107)	3.5 (19)	0.7 (4)	0.4 (2)	0.0 (0)	3.9 (23)
Obj. relative	17.5 (17)	0.0 (0)	28.9 (28)	1.0 (1)	32.0 (31)	5.2 (5)	15.5 (15)

in object relative clauses, but with a strikingly lower frequency than in subject relatives.

For object relatives, our results are similar to those of Kidd et al. (2007) (results for subject relatives are not reported by them). The major difference is that in our study, third- and first-person pronouns occur almost equally often whereas the majority of pronouns in Kidd et al.'s study were first-person pronouns. This difference can be attributed to the fact that we analyzed a corpus sample of written adult language whereas Kidd et al. analyzed a corpus of spoken child language.

Mak et al. (2008) have proposed that the subject in an object relative clause is typically a topic, and that this explains the high proportion of subject pronouns in object relatives. If the subject is a topic even when it is not a pronoun, this makes the prediction that the subject should immediately follow the relative pronoun in most cases because the leftmost position within the middlefield is the default position for topics (Frey 2004a). For subject relatives, it is assumed that the object is typically not a topic. Objects in subject relatives are therefore expected to occur anywhere within the relative clause. To test this hypothesis, we determined the clausal position in which the second NP appears for each relative clause. This is the subject when the relative pronoun is the object and it is the object when the relative pronoun is the subject. The relative pronoun always occurs in position 1. If the second NP occurs directly after the relative pronoun, it appears in position 2. If exactly one phrase intervenes between relative pronoun and second NP, the second NP occurs in position 3, and so on. In our corpus sample, the second NP occurred in one of positions 2–6.

Table 2 shows how often the second NP occurs in each position for both subject and object relatives. Relative clauses in which the second NP was either a personal or a reflexive pronoun were excluded from this analysis because such pronouns obligatorily occur in an early position within the clause. In object relatives, the second NP appears in position 2 in nearly all cases, that is, directly after the relative pronoun. Because the second NP in an object relative clause is the subject, it typically occupies a high position in the syntactic structure, therefore occurring rather early in the clause. However, only specific, topical subjects appear in the highest position below the prefield (Diesing 1992; Frey

Table 2: Percentages of relative clauses in which the second NP (the object in subject relative clauses, the subject in object relative clauses) occurs in clausal positions 2-6, where position 1 is the relative pronoun. Relative clauses where NP2 was either a personal pronoun or a reflexive were excluded from the analysis. Raw numbers are given in parentheses.

	2	3	4	5	6
Subject relative	63.6 (335)	25.3 (134)	8.2 (45)	1.9 (10)	0.9 (1)
Object relative	93.3 (28)	3.3 (1)	3.3 (1)	0 (0)	0 (0)

2004a), and the finding that the subject in object relative clauses occurs in the second position in over 90 % of all cases is therefore suggestive for these subjects being topics. Because the exact position of the subject often remains ambiguous unless there is an adverbial marking the left VP boundary, the evidence is only suggestive and in need of further confirmation. In subject relatives, position 2 is also the most frequent position for the second NP, but here later positions are also observed with some regularity.

In sum, we take the data shown in Table 2 as tentative support of Mak et al.'s claim that the second NP is a topic in an object relative clause but not in a subject relative clause. Note that the frequency counts shown in Table 2 cannot be reduced to the definiteness of the second NP. Definite NPs are known to occur earlier in the clause than indefinite NPs (Lenerz 1977; for corpus evidence, see Bader & Häussler 2010). However, in both subject and object relative clauses, the second NP was a definite NP in the majority of cases, with no significant difference between the two clause types.

A further finding concerning the use of object relatives has been reported by Contemori and Belletti (2014). In an experiment investigating the spoken production of Italian relative clauses, Contemori and Belletti found that adults have a strong preference for producing subject relatives with the verb in the passive voice instead of corresponding object relatives. For children, such a preference became visible only in the oldest age group investigated (between 8 and 9 years).

In order to test whether a similar preference holds in the corpus sample under consideration, we determined the number of relative clauses with a verb in the passive voice and with or without a *von* ('by') PP. Since we are only considering verbs with an accusative object in this paper, these are all subject relative clauses. Table 3 shows the resulting numbers as well as the number of relative clauses with a verb in the active voice, a subject and an accusative object (these are the same subject and object relatives already discussed above).

Table 3 shows that the overall frequency of passive subject relative clauses is higher than the frequency of object relative clauses. However, a clear majority

Table 3: Number of subject and object relative clauses with a verb in the active voice, a subject and an accusative object, and number of subject relative clauses with a verb in the passive voice and with or without a *von* ('by') PP.

	Active	Passive with <i>von</i> PP	Passive without <i>von</i> PP
Subject relative clause	547	18	136
Object relative clause	97	–	–

of passive relative clauses does not contain a *von*-PP. Thus, passive voice seems to be used in relative clauses mainly for the purpose of omitting the underlying subject, and not for the purpose of avoiding OS order. The numbers in Table 3 contrast with the experimental results of Contemori and Belletti (2014). It is an open question as to whether this is related to grammatical differences between Italian and German or to other differences (e.g., spoken versus written language, experimental data versus corpus data).

In sum, the corpus data presented in this section reveal the same pattern as has been found for English child and adult language and for German child language. Whereas subject relatives outnumber object relatives when considering all relative clauses with a subject and a direct object, the reverse relationship is found when we look only at relative clauses in which the second NP is a personal pronoun. Here, object relatives occur with greater frequency than subject relatives, especially for first and second person pronouns. The additional data discussed in this section support Mak et al.'s (2008) hypothesis that the high percentage of subject pronouns in object relatives comes about because the subject in an object relative clause is a topic. On the other hand, we found no evidence that writers revert to the passive voice in order to avoid object relative clauses.

### 3 Non-canonical order in main clauses

We now turn to main clauses in which either the subject or the object occupies the prefield. As before, we restrict our discussion to sentences with an accusative object. For main clauses, the situation is more complex than for relative clauses because there is no constraint restricting the clause-initial phrase in a way similar to the case of relative clauses. With regard to factors favoring OS order, discourse properties of the object can therefore be as relevant as discourse properties of the subject. Before we consider subject and object in turn, the next subsection reviews recent theories of how speakers or writers decide which phrase to put into the prefield.



## The preferred filler of the prefield

The question of whether to put the subject or the object into the prefield is closely related to the question of what the preferred position of the sentence topic is in German main clauses. While older work saw the prefield as the default position for the sentence topic (Gundel 1988), more recent research (Frey 2004a; Rambow 1993; Speyer 2007) suggests that the default position is at the left edge of the middlefield (sometimes called the Wackernagel position). For reasons of space, we consider only the proposal of Speyer (2007, 2009, 2010), who claims that the topic is put into the prefield only if a clause contains no element higher on the prefield hierarchy given in (7).

- (7) scene-setting >> poset >> topic

The hierarchy in (7) contains two kinds of elements that have precedence when it comes to filling the prefield. Scene-setting elements are typically adverbials that locate an event in time and space. A poset element is linked to the prior discourse by a poset (partially ordered set) relation in the sense of Ward and Prince (Ward & Prince 1991). Examples for poset relations are the set-membership relation and the part-of relation.

For purposes of illustration, we consider the experiments of Weskott et al. (2011), which confirm the importance of the poset relation for the purposes of filling the prefield. Weskott et al. (2011) obtained acceptability ratings and reading times for short texts consisting of two sentences, as illustrated in (8).

- (8) Peter hat den           Wagen   gewaschen.  
 Peter has the.ACC   car       washed.  
 'Peter has washed the car'
- a. Er           hat den       Außenspiegel ausgelassen.  
 He.NOM has the.ACC side mirror   left-out
- b. Den       Außenspiegel hat er       ausgelassen.  
 The.ACC side mirror   has he.NOM left-out.  
 'The side mirror, he left out.'

The first sentence of each text introduces two referents. The first of them is taken up again in the second sentence by a subject pronoun. The second referent of the initial sentence is not taken up again in toto, but a part of it is referred to in the second sentence by means of a definite NP serving as the object. Thus, the NP *den Außenspiegel* ('the side mirror') in (8) stands in a poset relation to the NP *den Wagen* ('the car') of the first sentence. As shown in (8), the second sentence

appears with either SO or OS order. Both acceptability ratings and reading times revealed an advantage for OS sentences in comparison to SO sentences. This is therefore a case of strong OS licensing. Following Speyer (2007), Weskott et al. (2011) attribute the strong OS licensing for sentences as in (8) to the poset relation between the object of the second and the object of the first sentence, but they note that taking up the first NP by means of a subject pronoun may also have contributed to the advantage observed for OS order.

In the next two subsections, we present corpus data addressing two questions. The first is whether the referential form of the subject affects the probability of using a sentence with OS order. In particular, does a pronominal subject have a similar effect as seen in relative clauses? The second question concerns the object itself. Given that Weskott et al. (2011) found strong licensing when the object referent stood in a poset relation to a prior referent, the question is whether other relationships between the object referent and the prior discourse strongly or weakly license OS order as well. Here, we will consider the simplest relation, namely the identity relation which holds when the object referent is simply given in the prior discourse.

### Subject properties favoring OS order

In accordance with prior findings on child language and language comprehension, our corpus study of relative clauses found that in the majority of object relative clauses the second NP is a personal pronoun. Whether we should expect a similar finding for main clauses is not straightforward because word order is less optional in relative clauses than in main clauses. That is, whereas declarative main clauses leave a choice as to which element to put into the prefield, there is no choice of word order in relative clauses as far as the initial element is concerned – the relative pronoun always has to come first.

Preliminary evidence on this issue comes from an ongoing corpus study that investigates the conditions governing the choice between personal pronoun and d-pronoun. This is a follow-up study to Portele and Bader's (2016) study, which investigated the choice between personal pronoun and d-pronoun for the case of subject pronouns. Based on a search of about 20% of the deWac Corpus (Baroni et al. 2009), Table 4 shows how the form of the NP immediately following the finite verb in a verb-second clause depends on properties of the phrase that fills the prefield.

When the prefield is filled by a subject pronoun, the percentage of personal pronouns is quite low. When the prefield is filled by an object pronoun, in contrast, personal pronouns are found much more often directly after the finite verb. This is so when the personal pronoun *ihn* fills the prefield, and to an even greater

Table 4: Percentages of personal pronouns and other elements directly following the finite verb in a verb-second clause depending on syntactic function and the pronoun type of the pronoun in the prefield.

Syntactic function	Pronoun type	Word form	Element after C°	
			Personal pronoun	Other
Subject	Personal pronoun	Er	5.7	94.2
	D-pronoun	Der	8.1	91.9
Direct object	Personal pronoun	Ihn	22.3	77.8
	D-pronoun	Den	59.9	40.1

extent when the prefield hosts the d-pronoun *den*. In fact, almost 60% of all sentences starting with the d-pronoun *den* had a personal pronoun as the subject. Sentences with a d-pronoun in the middlefield are contained within the category of other elements after C° in Table 4. A preliminary analysis of this category revealed that d-pronouns occur quite infrequently as subjects within the middlefield, a finding which has also been obtained by Bosch, Katz and Umbach (2007). Thus, sentences in which the object is a d-pronoun and the subject a personal pronoun seem to constitute a further case of strong OS licensing.

In sum, in both relative and main clauses the probability of OS order increases when the subject is a pronoun. In contrast to relative clauses, the evidence for main clauses is only suggestive. Further research is necessary – including research on language acquisition and language comprehension – in order to determine how far the parallels go.

## Object properties favoring OS order

The corpus data presented in this section are from an ongoing corpus study testing the prefield hierarchy given in (7) for the case of sentences with a subject and an accusative object. This study analyzes sentences from a random selection of Wikipedia texts, including 10,000 Wikipedia articles for each letter of the alphabet unless fewer were available. Because the focus is on determining what properties of the object increase or decrease its probability of occurring before or after the topic of the sentence, all sentences analyzed in this study contain the subject pronoun *er* ('he') with the discourse function of topic. Five types of object NPs are analyzed:<sup>2</sup>

2 The corpus search also included the d-pronoun *den* ('the.ACC') but there were too few corpus hits to warrant further analysis.

- definite NPs starting with the definite article *den* ('the.ACC')
- indefinite NPs starting with the indefinite article *einen* ('a.ACC')
- demonstrative NPs starting with the demonstrative determiner *diesen* ('this.ACC')
- the personal pronoun *ihn* ('him.ACC')
- the demonstrative pronoun *diesen* ('this.ACC')

Table 5 shows the number of corpus hits for each of the five types of object NPs listed above. With regard to the proportion of OS sentences, Table 5 shows a clear distinction depending on the type of the object NP. When the object is either a definite NP, an indefinite NP or a personal pronoun, SO order is preferred. This preference is strongest in the case of personal pronouns, which is in accordance with linguistic descriptions according to which object pronouns cannot occupy the prefield except for highly specific discourse conditions, including contrastive stress (see Lenerz 1992). For demonstrative objects, however, a preference for OS order is observed, for both full and pronominal NPs.

Table 5: Number of corpus hits broken down by order and type of the object NP.

	Non-pronominal object			Pronominal object	
	Demonstrative NP	Definite NP	Indefinite NP	Demonstrative pronoun	Personal pronoun
SO	121	3860	1907	35	183
OS	388	841	305	110	4
% OS	76.2	17.9	13.9	75.8	2.1
Ratio	1 : 3.2	4.6 : 1	6.3 : 1	1 : 3.1	45.8 : 1

Definite NPs in particular, but indefinite NPs to some degree too, are known to show a great variety of relationships to the prior discourse. In order to investigate how the relation of non-pronominal objects to the prior discourse affects word order, 100 corpus examples with a non-pronominal object for each of the six combinations of order and object type were randomly selected for a detailed analysis. With regard to the discourse status of the object NP, the examples were annotated using the classification proposed in Birner and Ward (2009), which extends the influential proposal of Prince (1981). According to Birner and Ward (2009), each referent can be classified as given or new in two dimensions. The first dimension concerns the prior discourse: a referent can have been mentioned in the prior discourse or it can have been newly introduced. The second dimension concerns the hearer (or reader): a referent can be old or new relative to the hearer's prior knowledge. Birner and Ward make the further assumption that the two dimensions can be freely combined, giving rise to four categories as shown

Table 6: Discourse status as proposed by Birner &amp; Ward (2009) following Prince (1981).

	Hearer-old	Hearer-new
Discourse-old	Evoked <i>he</i>	Inferable
Discourse-new	Unused <i>Gov. Rod Blagojevich</i>	Brandnew <i>a \$3.6 billion state construction budget</i>

in Table 6. The names for the four categories are from Prince (1981). An example is given in (9). The bold-printed phrases in (9) illustrate three of the four categories shown in Table 6.

- (9) **Gov. Rod Blagojevich**, while scaling back a massive capital program, said Friday **he** would endorse **a \$3.6 billion state construction budget** that includes new money to build schools and millions of dollars for legislative pork-barrel projects.  
(Chicago Tribune, 8/23/03) [from (Birner 2003)]

The referent of the proper name *Gov. Rod Blagojevich* is mentioned for the first time at this point of the discourse and is therefore discourse-new. Since the typical reader of the newspaper where this text is from can be assumed to be familiar with this referent, it is hearer-old. This referent is taken up again by the following personal pronoun *he*, which thus refers to a referent evoked in the preceding clause and is therefore both discourse- and hearer-old. The indefinite NP *a \$3.6 billion state construction budget* introduces a new referent that cannot be assumed to be known by a typical reader. This referent is thus brand-new, that is, discourse- and hearer-new. Example (9) does not contain an instance of the fourth category, the inferables. However, we already saw an example of an inferable referent when discussing the experiments of Weskott et al. (2011). In example (8), the referent of the NP *den Außenspiegel* ('the side mirror') has not been mentioned before, but it can be inferred because it stands in a poset relation (more precisely a part-of relation) to the car mentioned in the preceding sentence.

We next discuss the main findings for each of the three types of non-pronominal NPs included in the present corpus study. For each NP type, a representative example containing a main clause with OS order is provided in Table 7.

- (i) *Demonstrative objects*. Demonstrative objects show a rather uniform relationship to the preceding context. In almost all cases, they refer to a referent evoked in the immediately preceding clause. Most of the time, this is the referent of an NP, as in the example in Table 7, but references to the event

Table 7: Representative examples of OS sentences in which the fronted object was either a demonstrative NP, an indefinite NP, or a definite NP.

Demonstrative NP	Alfred Alexander Taylor (. . .) war ein US-amerikanischer Politiker und der 38. Gouverneur von Tennessee. <u>Diesen Bundesstaat</u> vertrat er außerdem im US-Repräsentantenhaus.
	'Alfred Alexander Taylor (. . .) was an US-American politician and the 38th Governor of Tennessee. He also represented <u>this state</u> in the House of Representatives.'
Indefinite NP	Anschließend promovierte Monar im Jahre 1989 an der Ludwig-Maximilians-Universität München in moderner Geschichte. <u>Einen zweiten Dokortitel</u> erlangte er im Jahre 1991 auf dem Gebiet der Politik- und Sozialwissenschaften am Europäischen Hochschulinstitut in Florenz.
	'Subsequently, Monar graduated in 1989 in modern history from Ludwig-Maximilians-University of Munich. <u>A second doctoral degree</u> he achieved in 1991 in the field of political and social sciences from the University of Florence.'
Definite NP	Loos starb im Sanatorium Kalksburg bei Wien, wo er mit einer Krankenschwester befreundet war, die er dem Vernehmen nach heiraten wollte. Er ruht in einem Grab auf dem Wiener Zentralfriedhof (Gruppe 0, Reihe 1, Nummer 105). <u>Den Grabstein</u> hatte er selbst entworfen.
	'Loos died in the Kalksburg sanatory near Vienna, where he was friends with a nurse, who he wanted to marry, it is said. He rests in a grave at the Vienna Central Cemetery (Group 0, Row 1, Number 105). <u>The gravestone</u> he had designed himself.'

introduced by the VP are also not uncommon (e.g., *Last year, Peter won the German championship. This victory ...*). As shown in Table 5, demonstrative objects occur more often in the prefield than in the middlefield and thus constitute an instance of strong OS licensing. This is not predicted by the prefield hierarchy of Speyer because the relevant discourse relation – identity with a referent evoked in the prior discourse without being a topic – does not appear in the prefield hierarchy.

- (ii) *Indefinite objects.* Indefinite NPs introducing a brand-new referent occur in the middlefield most of the time. When the referent of indefinite NPs stands in a poset relation to a referent in the prior discourse, as in the example in Table 7, the indefinite NP preferentially appears in the prefield. A similar observation has been made for English by Ward and Prince (1991).
- (iii) *Definite objects.* As expected given the linguistic literature, definite NPs showed the most varied behavior in terms of discourse status. A preference for the prefield and thus OS order was only found for definite NPs in a poset relation to the prior discourse. For NPs which were inferable from the situation as a whole, but not from a specific referent in the prior discourse, in contrast, SO order prevailed (see Ward & Prince, 1991, for the

difference between NPs given by a poset relation and NPs that are situationally given). Anaphoric definite NPs, that is, NPs referring to referents that are discourse- and hearer-old, are not uncommon in the prefield, but they are even more common in the middlefield. Definite NPs referring to an unused referent, that is, a referent that is discourse-new but hearer-old, appear most of the time in the middlefield.

#### Toward a prefield hierarchy for SO/OS order

The findings reviewed in this section are mostly compatible with Speyer's prefield hierarchy in (7). Only one discrepancy was found: NPs referring to given referents show a preference for the prefield if the NP is a demonstrative – either an NP or pronoun – or a d-pronoun. For the task of choosing between subject and object as the filler of the prefield, we therefore propose the following prefield hierarchy.

##### (10) SO/OS prefield hierarchy

given(demonstrative, d-pronoun), poset > topic, given(definite) >  
brand-new

Like the more general prefield hierarchy of Speyer, the SO/OS prefield hierarchy in (10) is not meant as a categorical hierarchy but a preference hierarchy which captures preferences in the case of competing orders. Note that the SO/OS prefield hierarchy differs from Speyer's prefield hierarchy not only with regard to the number of elements, but also with regard to the type of information that is referred to. In contrast to Speyer's hierarchy, the hierarchy in (10) refers not only to the discourse status of the various referents but also to the referential form used for making reference. The finding that given referents which are referred to by a demonstrative expression are especially prone to fill the prefield and thus to occur in sentence initial position may possibly be related to the very nature of demonstratives, that is, pointing to an element in the nearby context (see Consten and Averintseva-Klisch 2010).

## 4 General discussion

As noted in the introduction, sentences with non-canonical word order are typically acquired later and are more difficult to process than sentences with canonical word order. However, in some cases, sentences with non-canonical word order are in fact advantageous in comparison to sentences with canonical word

order, as captured in Weskott et al.'s (2011) notion of strong licensing of OS order. The first question asked in this paper was whether reported instances of strong OS licensing can be replicated when looking at written language production in German. The second question was whether additional instances of strong (or weak) OS licensing can be found.

With respect to relative clauses, we found that object relative clauses are produced more frequently in written language when the second NP is a topic, an entity already introduced in the discourse. This claim is based on two findings. First, we found that object relative clauses are more frequent when the second NP is a pronoun, which typically refers to topics. A second finding was that in object relative clauses the subject almost always occurs directly after the relative pronoun, which is the canonical topic position.

These corpus findings are in accordance with previous experimental findings. For Dutch, Mak et al. (2008) report that object relative clauses are easier to process than subject relative clauses when the second NP is a case-ambiguous pronoun. This means that when a relative clause is processed, a pronoun that does not commit the reader to a specific reading is preferably interpreted as the subject of the relative clause and consequently the relative clause is interpreted as an object relative clause. Our findings replicate this comprehension pattern in written language production. Mak et al. (2008) also manipulated the context of subject and object relative clauses. They presented subject and object relative clauses in neutral and topic contexts (which introduced the second NP of the relative clause). They found that when the second NP was introduced and thus the topic, object relative clauses were equally easy (but not easier) to process as subject relative clauses.

The corpus analysis of relative clauses showed that passive voice in subject relative clauses occurs most of the time without a *by*-phrase. This is in contrast to the findings of Contemori and Belletti (2014). However, this pattern resembles the findings of Friedmann et al. (2009), who report that Hebrew-speaking children in some cases produced subject relative clauses with a reflexive verb instead of an object relative clause. In any case, our corpus findings suggest that passive subject relative clauses are not used as an alternative to object active clauses.

All in all, our findings confirm earlier findings that object relative clauses are not less frequent than subject relative clauses across the board, but are in fact preferred under specific conditions related to discourse factors. In particular, the present findings provide further evidence for strong OS licensing when the subject of a relative clause is a topic, and especially so when it is a pronominal topic. This may also explain why in the study of Hirschberg et al. (2014) object relatives occur with a rather high percentage of about 25%. Although subject relatives (which are not differentiated with regard to whether they also contain an object) are still the most frequent type in this study, the percentage of object



relatives is much higher than in our study or in Mak et al. (2002). Hirschberg et al. (2014) investigate a corpus of spoken language and almost all examples of object relatives contain a first-person pronoun as subject. As shown above, this is exactly the condition that strongly favors the production of object relatives.

With respect to main clauses, the situation is more complicated because the order of subject and object is affected by properties of both constituents. First, we found that – similarly to relative clauses – the proportion of OS sentences increases when the subject is a personal pronoun. Since this finding was restricted to sentences in which the object is a pronoun, further corpus research is necessary to determine whether this finding generalizes to other types of object NPs. With respect to the object, we found that word order is affected both by the relation of the object to the prior discourse and by the particular referential expression of the object NP (demonstrative vs. definite vs. indefinite NP vs. personal pronoun).

The corpus data discussed in this paper raise a range of questions in need of further research. First, in contrast to experimental research on relative clauses, few experimental studies exist on strong OS licensing for main clauses. For acceptability ratings and reading times obtained for adult participants, Weskott et al. (2011) have shown strong OS licensing when the object is related by a poset relation to the prior discourse, but for the other cases experimental evidence is lacking (for related work on language acquisition, see Saueremann 2016).

If the cases of strong OS licensing in main clauses can be corroborated, a further question is whether the findings can be accommodated within an overarching account. In particular, main clauses and relative clauses were similar insofar as the object precedes the subject more frequently when the subject is the topic than when the subject is not a topic. Can this similarity be rooted in the discourse function associated with topics, or is this just a superficial similarity between relative clauses and main clauses that has no common source?

A final set of questions concerns the SO/OS prefield hierarchy proposed in (10). One task for future research is to integrate the SO/OS prefield hierarchy, which only applies to the order of subject and object, with Speyer's hierarchy, which applies to all potential fillers of the prefield. To do so, it has to be determined, among others, how scene-setting phrases and demonstrative NPs are ranked relative to each other. In addition, it remains to be seen how lexical-conceptual information (e.g., animacy and thematic roles) interacts with the discourse-based information encoded in the SO/OS prefield hierarchy. As shown by several corpus studies (e.g., Bader & Häussler 2010; Verhoeven 2015), lexical-conceptual information does not only affect the order of arguments within the middlefield but also when one argument occupies the prefield. Addressing this issue will require taking prefield and middlefield into account simultaneously (see Frey 2004b for a theoretical-linguistic proposal.)

## References

- Bader, Markus and Jana Häußler. 2010. Word order in German: A corpus study. *Lingua* 120(3): 717–762.
- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi and Eros Zanchetta. 2009. The WaCky Wide Web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation Journal* 23(3): 209–226.
- Birner, Betty. 2003. Discourse functions at the periphery: Noncanonical word order in English. In B. Shaer, Werner Frey and Claudia Maienborn (eds.), *Proceedings of Dislocated Elements Workshop, ZAS Berlin*, 41–62.
- Birner, Betty J. and Gregory Ward. 2009. Information structure and syntactic structure. *Language and Linguistics Compass* 3(4): 1167–1187.
- Bosch, Peter, Graham Katz and Carla Umbach. 2007. The non-subject bias of German. In Monika Schwarz-Friesel, Manfred Consten and Mareile Knees (eds.), *Anaphors in Text: Cognitive, formal and applied approaches to anaphoric reference*, 145–164. Amsterdam: John Benjamins Publishing.
- Burchert, Frank, Nadine Meißner and Ria d. Bleser. 2008. Production of non-canonical sentences in agrammatic aphasia: Limits in representation or rule application? *Brain and language* 104(2): 170–179.
- Consten, Manfred and Maria Averintseva-Klisch. 2010. ‘Nahe Referenten’-ein integrativer Ansatz zur Funktion demonstrativer Referenz. *Sprachtheorie und germanistische Linguistik* 20(1): 1–34.
- Contemori, Carla and Adriana Belletti. 2014. Relatives and passive object relatives in Italian-speaking children and adults: Intervention in production and comprehension. *Applied Psycholinguistics* 35(06): 1021–1053.
- Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: MIT Press.
- Ellsiepen, Emilia and Markus Bader. 2018. Constraints on argument linearization in German. *Glossa: Glossa: a journal of general linguistics*, 3(1), 6. DOI: <http://doi.org/10.5334/gjgl.258>.
- Fox, Barbara A. and Sandra A. Thompson. 1990. A discourse explanation of the grammar of relative clauses in English discourses. *Language* 66: 297–316.
- Frey, Werner. 2004a. A medial topic position for German. *Linguistische Berichte* 198: 153–190.
- Frey, Werner. 2004b. The grammar-pragmatics interface and the German pre-field. *Sprache & Pragmatik* 52: 1–39.
- Friedmann, Naama, Adriana Belletti and Luigi Rizzi. 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua* 119(1): 67–88.
- Gundel, Jeanette K. 1988. Universals of topic-comment structure. In Michael Hammond, Edith A. Moravcsik and Jessica R. Wirth (eds.), *Studies in syntactic typology*, 209–239. Amsterdam: John Benjamins Publishing.

- Hirschberg, Tim, Carolin Reinert, Anna Roth and Caroline Féry. 2014. Relative Clauses in Colloquial and Literary German: A Contrastive Corpus-Based Study. *Linguistische Berichte* 240: 405–445.
- Hoberg, Ursula. 1981. *Die Wortstellung in der geschriebenen deutschen Gegenwartssprache*. München: Hueber.
- Höhle, Tilman N. 1982. Explikation für „normale Betonung“ und „normale Wortstellung“. In Werner Abraham (ed.), *Satzglieder im Deutschen. Vorschläge zur syntaktischen, semantischen und pragmatischen Fundierung*, 75–153. Tübingen: Narr.
- Kaiser, Elsi and John C. Trueswell. 2004. The role of discourse context in the processing of a flexible word-order language. *Cognition* 94: 113–147.
- Kempen, Gerard and Karin Harbusch. 2005. The relationship between grammaticality ratings and corpus frequencies: A case study into word-order variability in the midfield of German clauses. In Marga Reis and Stephan Kepser (eds.), *Linguistic evidence. Empirical, theoretical and computational perspectives*, 329–349. Berlin: de Gruyter.
- Kidd, Evan, Silke Brandt, Elena Lieven and Michael Tomasello. 2007. Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children’s processing of relative clauses. *Language and cognitive processes* 22(6): 860–897.
- Lenerz, Jürgen. 1977. *Zur Abfolge nominaler Satzglieder im Deutschen*. Tübingen: Narr.
- Lenerz, Jürgen. 1992. Zur Syntax der Pronomina im Deutschen. *Sprache und Pragmatik* 29.
- Mak, Willem, Wietske Vonk and Herbert Schriefers. 2002. The influence of animacy on relative clause processing. *Journal of Memory and Language* 47: 50–68.
- Mak, Willem M., Wietske Vonk and Herbert Schriefers. 2008. Discourse structure and relative clause processing. *Memory & Cognition* 36(1): 170–181.
- Portele, Yvonne and Markus Bader. 2016. Accessibility and referential choice: Personal pronouns and d-pronouns in written German. *Discours. Revue de linguistique, psycholinguistique et informatique* 18: 1–41.
- Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–255. New York etc.: Academic Press.
- Rambow, Owen. 1993. Pragmatic aspects of scrambling and topicalization in German: A Centering Approach. In *IRCS Workshop on Centering in Discourse*.
- Sauermann, Antje. 2016. *Impact of the type of referring expression on the acquisition of word order variation*. Potsdam: Universitätsverlag Potsdam.
- Speyer, Augustin. 2007. Die Bedeutung der Centering Theory für Fragen der Vorfeldbesetzung im Deutschen. *Zeitschrift für Sprachwissenschaft* 26(1): 83–115.

- Speyer, Augustin. 2009. Das Vorfelddrinking und das Vorfeld-es. *Linguistische Berichte* 219: 323–353.
- Speyer, Augustin. 2010. Filling the German vorfeld in written and spoken discourse. In Sanna-Kaisa Tanskanen, Marja-Liisa Helasvuo and Marjut Johansson (eds.), *Discourses in Interaction*, 263–290. Amsterdam: John Benjamins Publishing.
- Verhoeven, Elisabeth. 2015. Thematic asymmetries do matter! A corpus study of German word order. *Journal of Germanic Linguistics* 27(01): 45–104.
- Ward, Gregory L. and Ellen F. Prince. 1991. On the topicalization of indefinite NPs. *Journal of Pragmatics* 16(2): 167–177.
- Weskott, Thomas, Robin Hörnig, Gisbert Fanselow and Reinhold Kliegl. 2011. Contextual licensing of marked OVS word order in German. *Linguistische Berichte* 225: 3–18.