

Word Order in German: A Corpus Study[☆]

Markus Bader*, Jana Häussler

Department of Linguistics, University of Konstanz, 78457 Konstanz, Germany

Abstract

This paper presents a corpus study of the order between subject and object in German main and embedded clauses. Since prior studies have shown that object-subject (OS) sentences are rare in comparison to subject-object (SO) sentences, for both main and embedded clauses two corpora were assembled: One corpus containing both SO and OS sentences, and a second corpus containing only OS sentences. In accordance with prior work, the rate of OS sentences was low, but the construction of a corpus restricted to OS-sentences still allowed a statistical analysis of OS-sentences. For embedded clauses, the main results of the current corpus study are: (i) Accusative objects predominate in SO-sentences but dative objects in OS-sentences. (ii) The use of OS order is tied to lexical-semantic properties of verbs and their arguments. OS order occurs mainly when the subject is inanimate and the object animate. The verbs used in OS-sentences are mainly ditransitive verbs in the passive voice and unaccusative verbs. (iii) Weight had no significant effect on order. For main clauses with either subject or object in the prefield somewhat different results were found. Following the presentation of the corpus results, we discuss how they fit into current conceptions of word-order variation.

Key words: German, word-order, corpus, information-structure, lexical-semantic information, constituent weight

1. Introduction

In the linguistic literature concerned with German, it is widely agreed that the canonical order between a subject and an object is subject-before-object (SO).¹ However, German also allows a wide range of sentences with object-before-subject (OS) word order. Setting special cases like fronting of *wh*-phrases and relative pronouns aside, a survey of the linguistic literature on German reveals three possible reasons why a sentence might have OS order instead of the canonical SO order. The first reason has to do with discourse-related notions like focus-background and topic-comment. For example, a question like *Who will buy this book?* can be answered with an SO sentence like (1a) but also with an OS sentence like (1b).

- (1) Wer wird dieses Buch kaufen? ('Who will buy this book?')
- a. [_{Focus} Der Direktor] wird [_{Topic} dieses Buch] kaufen.
the-NOM director will this-ACC book buy
'The director will buy this book.'
- b. [_{Topic} Dieses Buch] wird [_{Focus} der Direktor] kaufen.
this-ACC book will the-NOM director buy
same meaning as (1a)

[☆]This work was supported by the Deutsche Forschungsgemeinschaft (SFB 471, Project D2). For helpful comments, we would like to thank Josef Bayer, Miriam Butt, Gerard Kempen, and two anonymous reviewers.

*Corresponding author: Tel.: +49 7531 884753; fax +49 7531 884459

Email addresses: markus.bader@uni-konstanz.de (Markus Bader), jana.haeussler@uni-konstanz.de (Jana Häussler)

¹The relevant literature is too extensive to be cited exhaustively. We therefore provide just a list of representative work: Dietrich (1994); Engel (1972); Hoberg (1981); Jacobs (1988); Lenerz (1977, 2001); Lerot (1985); Lötscher (1981); Müller (1999); Primus (1994); Reis (1986); von Stechow and Sternefeld (1988); Uszkoreit (1986, 1987).

By switching from SO to OS word order, both the topic (*dieses Buch*) and the focus (*der Direktor*) are brought into their preferred positions: the topic into the clause initial position and the focus into the position directly preceding the clause-final verb.

A rather different reason for using a sentence with OS word order is related to semantic properties of verbs and their arguments. For example, when a clause contains an agentive verb, the agent is realized as the subject and the sentence has SO order when discourse reasons do not indicate otherwise (cf. (2a)). Certain psych-verbs, in contrast, have the stimulus as the subject and the experiencer as the object. Sentences with these verbs show up with OS word order unless SO is required for discourse reasons (cf. (2b)).

- (2) a. Vermutlich hat der Direktor das Buch gelesen.
presumably has the-NOM director the-ACC book read
'Presumably the director read the book.'
- b. Vermutlich ist dem Direktor das Buch aufgefallen.
presumably is the-DAT director the-NOM book attracted-attention
'Presumably the book attracted the director's attention.'

A third reason for using OS instead of SO word order is related to the notion of constituent weight, as captured in the *Law of Increasing Constituents* ('Gesetz der wachsenden Glieder') proposed by Behagel (1909/1910). More recently, John Hawkins (e.g., Hawkins, 1994, 2004) has proposed a comprehensive theory of the human parsing mechanism which claims that constituent weight is the main determinant of constituent order (see also Wasow, 2002).

The three reasons that have been proposed as governing the choice between subject-object (SO) and object-subject (OS) word order in German coincide with three types of linearization hierarchies that have been identified in cross-linguistic research on word order variation (summarized in Siewierska, 1993): first, hierarchies in terms of discourse-related notions (e.g., "Definite NPs precede indefinite NPs"); second, hierarchies in terms of lexical semantics (e.g., "Animate NPs precede inanimate NPs"); and third, hierarchies in terms of weight (e.g., "Short NPs precede long NPs").

Despite the large interest in word order variation found in the linguistic literature on German, quantitative information about the actual usage of sentences with either SO or OS word order is relatively sparse. Several prior corpus studies which we will review below (e.g., Hoberg, 1981; Kempen and Harbusch, 2005) have provided important information about the frequencies of SO and OS sentences, and about some of the factors governing the choice between the two competing orders. However, as our review will show, these prior studies have several limitations both with regard to the sentence types investigated, and with regard to the factors that were taken into account. In particular, since OS order is much less frequent than SO order, the number of OS sentences considered in these studies was too limited for detailed statistical analyses. To overcome these limitations, we have conducted a more comprehensive corpus study that has investigated the order between subject and object in both main and embedded clauses of German. In contrast to prior studies, we not only assembled sentence sets containing both SO and OS sentences—with the same low rate of OS sentences as in the prior studies—but also sentence sets containing only OS sentences. In this way, it becomes possible to base claims about the nature of OS sentences on a much larger data base. The main objective of this corpus study was to determine to what extent the various linearization hierarchies that have been proposed in the linguistic literature are reflected in actual corpus frequencies.

The organization of this paper is as follows. In section 2, we provide linguistic background information on word order variation in German and review previous corpus studies. The method and basic results of the current corpus study are presented in section 3. Section 4 investigates which of the linearization hierarchies that are introduced in the next section are necessary to account for the results presented in section 3. Consequences of the results presented in this paper for grammatical models of word order variation will be discussed in section 5. The final section 6 provides a summary and a general discussion of our findings.

2. The Order between Subject and Object: Prior Work

Limitations of space prevent us from even trying to provide a comprehensive review of the pertinent literature concerned with word order variation in German (cf. footnote 1 for a representative list of references). A short overview of the major positions will be given when we discuss grammatical models of word order variation in light

of our corpus results in section 5. At this point, we confine ourselves to a summary of the word order hierarchies that are discussed by Siewierska (1993) in her review article of word order in the languages of the world (incidentally, by taking Siewierska, 1993, as the starting point for the following corpus analysis, we indirectly take into account much of the word order literature concerned with German because this literature is among Siewierska’s sources). In contrast to the amount of theoretical work on word order in German, the number of corpus studies is rather small. We give a short summary of these studies in the second part of this section.

For the discussion of German grammar, the so-called *topological model* has proven to be especially useful because it provides an insightful but theory-neutral way to refer to the different parts of a German clause (cf. Drach, 1937; Engel, 1972; Höhle, 1986). The terminology provided by this model is established in (3). The corresponding modern terms are also given in (3).

| (3) Clause type | Prefield (SpecCP) | Left Bracket (C°) | Middlefield | Right Bracket (Verb cluster) |
|-----------------|----------------------|----------------------|----------------|---------------------------------|
| Main | Peter | hat | ein Buch | gelesen. |
| Embedded | | dass | Peter ein Buch | gelesen hat. |
| | P. | has/that | P. a book | read has |

‘(that) Peter has read a book.’

The topological model captures several important aspects of German clause structure.

- The skeleton of the sentence is formed by the left and right bracket. In both main and embedded clauses, the right bracket consists of all non-finite verbs. In embedded clauses, the right bracket also contains the finite verb, and the complementizer forms the left bracket. Main clauses do not contain a complementizer, and the left bracket instead contains the finite verb; this is what makes German a verb-second language. Note that the right bracket in main clauses can be empty if a sentence contains only a single verb (which has to be finite and is thus located in the left bracket).
- Declarative main clauses contain exactly one constituent in front of the left bracket (the prefield). In declarative embedded clauses, this position must remain empty.

According to an assumption shared by traditionally oriented work (e.g., Engel, 1972) and work within generative grammar (cf. Bierwisch, 1963, Thiersch, 1978, and subsequent work summarized in, e.g., von Stechow and Sternefeld, 1988), one of the constituents which normally occupy the middlefield is fronted to the prefield in declarative main clauses. Which constituent is chosen for this purpose is usually considered to be a matter of discourse factors, in particular involving the notion of topic. It is therefore to be expected that word order in the middlefield and word order involving the prefield exhibit at least partially different properties. For this reason, we will present separate corpus analyses of word order internal to the middlefield and word order involving the prefield.

2.1. Word Order Hierarchies

In her typological review of the determinants of word order, Siewierska (1993) lists a number of hierarchies that can affect the linearization of arguments. Following Allan (1987), she classifies these hierarchies into *formal*, *dominance*, and *familiarity hierarchies*. She explains these notions as follows: “The formal hierarchies encompass matters of length and internal syntactic complexity. The dominance hierarchies cover factors connected with perception of salience as reflected in the way humans experience the world. And the familiarity hierarchies deal with the speaker’s individual interests as manifested in discourse via parameters such as topicality, givenness, definiteness etc.” (Siewierska, 1993: 830)

The full list of hierarchies discussed by Siewierska (1993) is given in (4)-(6), broken down according to the three hierarchy types just mentioned. Since in our opinion the labels for the three types of hierarchies are not as telling as they could be, and in order to enhance compatibility with the later discussion of our results with respect to current conceptions of word order variation, we will use a different terminology instead (Siewierska’s original labels are

shown in brackets). Hierarchies which are investigated in the following corpus study are typeset in italics.²

- (4) Weight (Formal) Hierarchies
 - a. *structurally simpler* > *structurally complex*
 - b. *short* > *long*
- (5) Lexical-Semantic (Dominance) Hierarchies
 - a. *The Personal Hierarchy:*
1stp. > *2ndp.* > *3rdp.* *human* > *higher animals* > *other organisms* > *inorganic matter* > *abstracts*
 - b. *The Semantic Role Hierarchy:*
agent > *patient* > *recipient* > *benefactive* > *instrumental* > *spatial* > *temporal*
- (6) Discourse-Related (Familiarity) Hierarchies
 - a. more familiar > less familiar topic > comment
 - b. given > new
 - c. *definite* > *indefinite*
 - d. referential > nonreferential

In accordance with common usage (cf. Wasow, 2002), we will use the term *weight* hierarchies for Siewierska's formal hierarchies. Instead of dominance hierarchies we will talk of *lexical-semantic hierarchies*, because the two hierarchies in this rubric are both defined in terms of semantic properties of lexical items—the Semantic Role Hierarchy in terms of semantic properties of verbs, and the Personal Hierarchy in terms of semantic properties of a verb's arguments. For the familiarity hierarchies, finally, we will use the term *discourse-related hierarchies* because the hierarchies in this category are all related to properties of phrases in unfolding discourses.

We defer a more detailed discussion of the individual hierarchies to the subsections where they become relevant for our corpus analysis. The only question that we would like to address at this point is whether such a multitude of hierarchies is indeed necessary. A particularly strong view with regard to this question has been endorsed by Hawkins (1994). According to Hawkins, the lexical-semantic and the discourse-related hierarchies are in effect epiphenomena of an overarching ordering principle falling under the broader notion of weight. The way the notion of weight is defined in Hawkins's theory is, however, much more refined than the definition of weight used by the weight hierarchies in (4). We consider Hawkins's work in some detail in section 4.4 after the analysis of our own corpus data with regard to the two weight hierarchies in (4).

Hawkins's (1994) claim that syntactic weight makes recourse to the lexical-semantic and the discourse-related hierarchies superfluous has been put to test in several recent studies. These studies converge at the conclusion that word order patterns cannot be reduced to a single factor of weight, but that instead discourse structure and lexical-semantic content affect word order independently of weight. The relationship between syntactic weight and discourse status has been addressed by Arnold et al. (2000) in an investigation of three English constructions (heavy NP shift, the dative alternation, and the verb-particle construction). Evidence from a corpus analysis and a sentence elicitation experiment shows that weight and discourse status contribute independently to the determination of word order. A similar conclusion has been reached by Rosenbach (2005) in her investigation of the role of weight and animacy for the position of possessive phrases within English NPs. Typically, an animate possessor precedes the head noun but an inanimate possessor follows it (*Peter's success* versus *the success of the book*). Rosenbach (2005) shows that, although animate NPs are on average shorter than inanimate NPs, animacy is an indispensable factor for determining the position of possessive NPs.

In summary, cross-linguistic work on the factors that determine word order has identified three families of hierarchies: *weight* hierarchies, *lexical-semantic hierarchies*, and *discourse-related hierarchies*. Despite attempts to reduce the different hierarchies to a single one (see in particular Hawkins, 1994), the available evidence indicates that none of the hierarchies is dispensable. Since the evidence concerning the independence of the various hierarchies is closely tied to only a few constructions—mainly from the English language—, we cannot assume without further justification that the hierarchies are also independent of each other for the linguistic domain considered here. With regard to our own corpus study, this has the following consequence. On the one hand, we will provide separate analyses of a subset

²The labels for the two dominance hierarchies are from Siewierska (1993); the other hierarchies are not labeled there.

Table 1: Corpus results for middlefield sentences from Hoberg (1981) and Kempen and Harbusch (2005).

| | Hoberg (1981) | | Kempen & Harbusch (2005) | |
|----|---------------|------------|--------------------------|-------------|
| | ACC | DAT | ACC | DAT |
| SO | 378 (86.1%) | 45 (10.3%) | 1476 (86.1%) | 182 (10.6%) |
| OS | 2 (0.5%) | 14 (3.2%) | 4 (0.2%) | 53 (3.1%) |

of each of the three types of hierarchies. On the other hand, we will also address the issue of factor confounding as far as this is possible given our corpus data.

2.2. Prior Corpus Studies of Word Order Variation in German

The seminal corpus study of word order variation in German is Hoberg (1981). Hoberg investigated a selection of 20 texts from the “Mannheimer Corpus” of contemporary written German texts since 1945, including newspapers, magazines, novels and popular science books. This selection contains approximately 85.000 words in about 11.000 clauses. The main results of Hoberg’s study with respect to the order between subject and object within the middlefield are shown on the left side of Table 1.³

Table 1 reveals two striking features. First, the overall rate of OS order is very low: OS sentences make up less than 4% of all sentences. Second, there is a striking interaction between order and case. Sentences with an accusative object occur with SO order in over 99% of all cases; for sentences with a dative object, in contrast, SO order still predominates, but the rate of OS order rises to almost 25%. Based on a more detailed analysis of her corpus data, Hoberg (1981, p. 62) proposes the basic order template in (7) (N = nominative; A = accusative; D = dative).

$$(7) (N-A-D)_{\text{pronominal}} - ((N-A-D)_{\text{animate}} - (N-A-D)_{\text{-animate}})_{\text{non-pronominal}} - (N/A/D)_{\text{FN}}$$

The template in (7) embodies two important claims. First, arguments within the middlefield fall into three groups. The first group— $(N-A-D)_{\text{pronom}}$ —contains pronouns; these are located at the beginning of the middlefield. The last group— $(N/A/D)_{\text{FN}}$ —contains arguments in so-called ‘Funktionsverbgefügen (FN)’ (light verb constructions and idiomatically combining expressions in the sense of Nunberg et al., 1994). Arguments which are neither pronominal nor light-verb arguments in the sense just defined form the middle group. This group is further subdivided, with animate arguments preceding inanimate arguments. Second, with the exception of the light-verb group, order within a group is subject before accusative object before dative object. Since the light-verb group can contain only a single element, no order is defined for this group. Deviations from the template in (7) are possible under certain conditions, which is why this template is said to capture the basic argument order within the middlefield. Among the conditions licensing reorderings of the elements in (7) are conditions defined in information-structural terms.

Results very similar to those found by Hoberg (1981) have been reported by Kempen and Harbusch (2005). Kempen and Harbusch (2005) analyzed two German treebanks, namely the NEGRA-II corpus (Skut et al., 2002) and the TIGER corpus (Brants et al., 2002). These are newspaper corpora comprising a total of 60,000 sentences. The order between non-pronominal subjects and objects is summarized on the right side of Table 1 (taken from Table 1 on page 332 of Kempen and Harbusch, 2005). As in Hoberg’s study, counts include both sentences with only one object and sentences with two objects. Despite the larger numbers, the percentages obtained by Kempen and Harbusch (2005) are almost exactly the same as those obtained by Hoberg (1981).

Based on their findings, Kempen and Harbusch (2005, p. 334) present the word order rule in (8) for the German middlefield (for ease of exposition, the template in (8) is presented somewhat differently than in Kempen and Harbusch (2005); this does not affect the content of the rule, however).

$$(8) \begin{array}{c} (N - A - D)_{\text{pronom}} - N - D - A \\ \text{Transitive} \quad \text{Intransitive} \quad \text{Ditransitive} \end{array}$$

³For sentences with either subject or object in the prefield, neither Hoberg (1981) nor Kempen and Harbusch (2005) (discussed below) provide the numbers necessary to compute a comparable table.

The template in (8) is to be read as follows: the basic order within the middlefield is pronominal arguments before non-pronominal arguments, with the internal order as indicated; the arrows show certain reorderings which are licensed by information-structural and lexical-semantic properties (see also Kempen and Harbusch, 2004).

Overall, the corpus study by Hoberg (1981) and the corpus study by Kempen and Harbusch (2005) show a high degree of similarity both with respect to the data and with respect to the word-order rules derived from them. In addition to these two corpus studies, a number of more specialized studies have investigated individual aspects of the order between subject and object in German (e.g., Fanselow, 2000; Hawkins, 1992; Heylen, 2005; Kempen and Harbusch, 2004; Kurz, 2000; Uszkoreit et al., 1998; Weber and Müller, 2004). These will be discussed when they become relevant in the context of the present corpus analysis.

2.3. Open Questions

In summary, prior research has led to several important findings concerning word order in German. However, these studies also have several shortcomings which justify conducting a new corpus study of the factors determining the order between subject and object in German. First and foremost, as shown in Table 1, the overall numbers of OS sentences analyzed in these studies was quite small, making it difficult to come up with reliable claims about the properties of such sentences. Second, these studies looked only at few of the factors which might be relevant for choosing between SO and OS order. Finally, neither of the two studies discussed above compared middlefield-internal word order with word order involving the prefield.

3. Corpus Study: Method and Basic Results

In order to address the shortcomings listed at the end of the last section, we have performed a new corpus study investigating the order between subject and object in German clauses. We first describe how the corpus was prepared for analysis and then present the basic results.

3.1. Method

The current corpus study is part of a larger project jointly investigating grammaticality judgments, language comprehension and corpus frequencies. As with much current research into human parsing, the focus of the comprehension studies is the interplay between lexical and syntactic ambiguity resolution (for overviews, cf. Mitchell, 1994; Tanenhaus and Trueswell, 1995). For this reason, we were particularly interested in sentences containing an object introduced by the definite article *den*. Depending on the accompanying noun, NPs introduced by *den* are ambiguous between accusative/singular and dative/plural and can thus be used to investigate the syntactic ambiguity between accusative and dative object. This is illustrated in (9). For reasons of space, the comprehension experiments that are based on the present corpus study must be reported in a separate article.

- (9) a. Wir haben den Studenten unterstützt.
 we have the student-ACC/Sing supported
 'We supported the student.'
- b. Wir haben den Studenten geholfen.
 we have the students-DAT/Plur helped
 'We helped the students.'

Four sets of sentences containing a *den*-object were prepared. The four sets were defined in a quasi-factorial way, the first factor being defined by the opposition between main and embedded clauses. Since it was known from the prior literature that the rate of OS-sentences is low in comparison to the rate of SO-sentences, the second factor was whether the position of the *den*-object was unrestricted or restricted to clause-initial position. The unrestricted set is later used to investigate the relative rates of SO- and OS-clauses; the restricted set allows a detailed investigation of OS-sentences. The definition of the four sentence sets, as well as their labels for later reference, are summarized in (10).

| | Embedded clause | Main clause |
|---------------------|--------------------------|-----------------------|
| Order unconstrained | <i>middlefield SO/OS</i> | <i>prefield SO/OS</i> |
| Object-initial only | <i>middlefield OS</i> | <i>prefield OS</i> |

Table 2: Overview of Corpora

| | Middlefield Corpora | | Prefield Corpora | |
|--------------------------|---------------------|----------------|------------------|----------------|
| | Unconstrained | Object-initial | Unconstrained | Object-initial |
| <i>Total</i> | 1176 | 930 | 668 | 827 |
| <i>Subj = Nominal</i> | 828 | 837 | 517 | 598 |
| <i>Subj = Pronominal</i> | 339 | 0 | 147 | 202 |
| <i>No Subj</i> | 9 | 93 | 4 | 27 |
| <i>1 NP Arg</i> | 9 | 93 | 4 | 27 |
| <i>2 NP Args</i> | 1015 | 825 | 559 | 719 |
| <i>3 NP Args</i> | 152 | 12 | 104 | 80 |
| <i>4 NP Args</i> | 0 | 0 | 1 | 1 |
| <i>Acc : Dat</i> | 75 : 25 | 6 : 94 | 81 : 19 | 73 : 27 |

To build these sets, sentences were downloaded from the newspaper corpus made available by the Institute for German Language (IDS) in Mannheim/Germany (<http://www.ids-mannheim.de>). Since this corpus is neither parsed nor tagged, we had to query the system with a relatively coarse-grained pattern. In order to keep the following corpus preparation to a manageable size, the initial query was restricted by the random-sampling function of the COSMAS system which is provided by the IDS as a web-based interface to its corpora. The resulting sentence sets were then further processed to include only sentences of interest. For example, all sentences in which *den* was not a verbal argument but contained within a PP were subsequently removed.

Two of the four sentence sets that were downloaded were restricted to sentences which contained an embedded clause introduced by the complementizer *dass* ('that') and a *den*-object within this embedded clause. One sentence set with embedded clauses was unrestricted with respect to the position of the *den*-object (*corpus middlefield SO/OS*) whereas the other set was restricted to *dass*-clauses in which the *den*-object immediately followed the complementizer, thus guaranteeing that the set contained only object-initial *den*-clauses (*corpus middlefield OS*). The two remaining sentence sets contained main clauses with a *den*-object. As with embedded clauses, one main-clause set was unrestricted with regard to the position of the *den*-object (*corpus prefield SO/OS*) whereas the other was constrained to have the *den*-object in clause initial position (*corpus prefield OS*).

The sentences in each set were annotated for a range of properties that were expected to be relevant for determining word order given prior work on word order variation. The following properties were coded: (i) number of NP-arguments, (ii) case of the *den*-phrase, (iii) pronominality of subject, (iv) animacy of subject and object(s), (v) definiteness of subject,⁴ (vi) length of subject and object, and (vii) verb-related properties including voice. Details about each property and its coding will be given when we discuss the relevant factors.

Table 2 shows the total number of sentences in each of the four sentence sets, as well as summary information about the pronominality of the subject, the number of NP arguments, and the overall rate of accusative and dative case.

The factor pronominality of the subject is included in Table 2 because one of the rare constraints about argument ordering in German that are considered to hold categorically concerns the position of subject pronouns: Nothing must intervene between C° and an unstressed subject pronoun. This is demonstrated in (11) for the case of embedded clauses where a complementizer occupies C° .

- (11) a. Peter hat gesagt, **dass er** *das Buch* erhalten hat.
 P. has said that he the book received has
 'Peter said that he has received the book.'
 b. *Peter hat gesagt, **dass** *das Buch* **er** erhalten hat.

The very same pattern is found in main-clauses with the finite verb in C° . The special behavior of subject pronouns is captured in the *Constraint on Subject Pronouns* in (12).

⁴Remember that the object was always definite because all objects were introduced by the definite determiner *den*.

(12) *Constraint on Subject Pronouns*
 *C° XP pronoun-NOM

This constraint states that a subject pronoun must immediately follow the left bracket (C°). For a subject pronoun in embedded clauses, this is in fact the only possible position, whereas in main clauses a subject pronoun can also occupy the prefield. In our data, all subject pronouns obeyed this constraint.

In embedded clauses in which the order between subject and object was unconstrained, all 335 clauses with a pronominal subject had the subject adjacent to C°. In embedded clauses confined to OS order, the subject was never a pronoun. In main clauses, the situation is slightly more complicated because there are three ways to obey the constraint on subject pronouns, as shown in (13).

- (13) a. Gestern HAT **er** den Opa besucht.
 yesterday has he the grandpa visited
 ‘He has visited grandpa yesterday.’
 b. **Er** HAT gestern den Opa besucht.
 he has yesterday the grandpa visited
 (same meaning as (13a))
 c. Den Opa HAT **er** gestern besucht.
 the grandpa has he yesterday visited
 (same meaning as (13a))

Subject pronoun and object can occur together in the middlefield with SO order (XP_{prefield} VSO; cf. (13a)), the subject pronoun can be in the prefield and the object in the middlefield (S_{prefield} VO; cf. (13b)), or the object can be in the prefield and the subject pronoun in the middlefield (O_{prefield} VS; cf. (13c)). The number of sentences in each of these categories is given in (14) separately for sentences with accusative and dative object; percentages are across columns. The category “XP_{prefield} VOS_{pron}” is missing from this table because, as said above, there were no sentences in which a subject pronoun followed an object in the middlefield.

(14)

| | Corpus prefield-SO/OS | | | Corpus prefield-OS |
|-----|---|-------------------------------|--|--|
| | XP _{prefield} VS _{pron} O | S _{prefield/pron} VO | O _{prefield} VS _{pron} | O _{prefield} VS _{pron} |
| Acc | 58 (85%) | 49 (82%) | 16 (84%) | 150 (74%) |
| Dat | 10 (15%) | 11 (18%) | 3 (16%) | 52 (26%) |

As shown in (14), all of the three configurations illustrated in (13) actually occur in our corpus. When the subject precedes the object, it occupies the prefield and middlefield with about equal frequency. As expected, sentences with an object in the prefield and a subject pronoun in the middlefield (OS-sentences) were less frequent than sentences with SO order. Finally, the ratio between accusative and dative sentences visible in the different conditions in (14) is approximately the same as the overall ratio in the complete prefield SO/OS corpus (rate of accusative and dative of 81% and 19%, respectively). Because subject pronouns have a special status, all sentences with subject pronouns were removed from the upcoming analyses; a corpus analysis dedicated to the position of pronouns in German can be found in Heylen (2005).

We next turn to the number of NP arguments which is also shown in Table 2. In all four sentence sets the majority of clauses have two NP arguments. There are also a number of sentences with a single NP argument. Since our sentence sets were constructed in such a way that all sentences contained an object with the determiner *den*, this object was the single argument in sentences with only one argument. Such sentences are therefore sentences without a subject in either pre- or middlefield. Because a clausal argument can also occur in extraposed position, that is, after the clause-final verb(s), there are three types of sentences without a subject in pre-/middlefield, as shown in (15) for the case of sentences with dative object.

- (15) a. ... dass den Besuchern geholfen wurde.
 that the-DAT visitors helped was
 ‘... that the visitors were helped.’

- b. ... dass den Besuchern erzählt wurde, dass das Museum geschlossen ist.
 that the-DAT visitors told was that the museum closed is
 ‘... that the visitors were told that the museum is closed.’
- c. ... dass den-DAT Besuchern half, was Maria zubereitet hatte.
 that the visitors helped what Maria prepared had
 ‘... that what Maria had prepared helped the visitors.’

In contrast to languages like English which require a subject in every finite clause, sentence (15a) without any subject at all is a grammatical option in German. In sentence (15b) the subject is an embedded declarative clause in extraposed position. Sentence (15c) finally has a free relative clause as subject, also in extraposed position. In principle, the three sentence types in (15) are possible with dative and with accusative objects. However, for accusative objects the two options (15a) and (15b) are severely restricted, in particular because of the lack of accusative objects in passive clauses.

The table in (16) shows the distribution of the three construction types in the middlefield OS and the prefield OS corpus (for the other two corpora, numbers of one argument sentences are too few for sensible analysis).

| (16) | Corpus middlefield OS | | Corpus prefield OS | |
|----------------|-----------------------|--------|--------------------|--------|
| | Accusative | Dative | Accusative | Dative |
| Subjectless | 0 | 60 | 0 | 17 |
| Subject clause | 0 | 30 | 1 | 9 |
| Free relative | 2 | 0 | 0 | 0 |

As shown by (16), the vast majority of sentences with no subject in the prefield/middlefield occur with a dative object although some rare examples with accusative object were also found. Most sentences are sentences without any subject at all, followed by sentences with a subject clause. Sentences with a free relative clause also occurred but only with very low frequency. On the one hand, these data show that subjectless sentences are indeed used in German texts; on the other hand, when the numbers of sentences with more than one NP argument in Table 2 are also taken into account, our data show that a clause starting with an object is most often a sentence with OS word order and not a sentence without a subject (and thus with an object only).

The two example sentences with four NP arguments contained a complex verb cluster containing the permissive verb *lassen* (‘to let’). They were excluded from the following analyses because of their small number. Sentences with only a single NP argument were also excluded because they lack an NP subject and thus the question of what determines the order between subject and object does not arise.

An additional restriction is only relevant for the prefield SO/OS corpus. In this corpus, 58.5% of the sentences had the subject in the prefield, 12.4% the *den*-object, and 29.1% some other phrase like adverbial phrases and PPs of various sorts. The table in (17) shows the distribution of the sentences in this corpus according to word order, case, and position of subject and object.

| (17) | S or O in prefield | | S and O in middlefield | |
|------|--------------------|------------|------------------------|------------|
| | ACC | DAT | ACC | DAT |
| SO | 256 (69%) | 46 (12.8%) | 116 (77%) | 18 (12.0%) |
| OS | 43 (11.7%) | 21 (5.7%) | 3 (2.0%) | 13 (8.7%) |

Since one of the questions addressed in our study was whether the order between subject and object is affected by the field(s) involved, we consider only sentences with either S or O in the prefield from the prefield SO/OS corpus. Word order in the middlefield is investigated on the basis of the two corpora containing embedded clauses. Based on the literature on word order in German, generalizations concerning the middlefield should hold equally for main and embedded clauses. A thorough investigation of this prediction is beyond the scope of this paper due to the small number of main clauses with both subject and object in the middlefield. Here, we only note that for sentences with both subject and object in the middlefield of a main clause, the table in (17) shows an interaction between order and case. SO sentences most of the time contain an accusative object whereas dative objects predominate in OS sentences. This is the same interaction that we already saw in discussing the corpus results obtained by Hoberg (1981) and Kempen and Harbusch (2005) for word order in the middlefield (cf. Table 1). Our own analysis of word order in the middlefield of embedded clauses (cf. Table 3 below) shows this interaction too. The data presented in (17) are thus

Table 3: Number of sentences by order, case and number of arguments for the two middlefield corpora and the two prefield corpora as defined in the text.

| Middlefield: SO/OS corpus (n = 829) | | | | |
|--|-------------|-------------|-------------|------------|
| | 2 Arguments | | 3 Arguments | |
| | Acc | Dat | Acc | Dat |
| SO | 546 (74.6%) | 75 (10.3%) | 38 (39.6%) | 54 (56.3%) |
| OS | 6 (0.8%) | 105 (14.3%) | 1 (1.0%) | 3 (3.1%) |
| Middlefield: OS corpus (n = 838) | | | | |
| | 2 Arguments | | 3 Arguments | |
| | Acc | Dat | Acc | Dat |
| OS | 54 (6.5%) | 771 (93.5%) | 0 (0%) | 12 (100%) |
| Prefield: SO/OS corpus (n = 366) | | | | |
| | 2 Arguments | | 3 Arguments | |
| | Acc | Dat | Acc | Dat |
| SO | 231 (74.3%) | 23 (7.4%) | 24 (43.6%) | 24 (43.6%) |
| OS | 38 (12.2%) | 19 (6.1%) | 5 (9.1%) | 2 (3.6%) |
| Prefield: OS corpus (n = 597) | | | | |
| | 2 Arguments | | 3 Arguments | |
| | Acc | Dat | Acc | Dat |
| OS | 409 (76%) | 130 (24%) | 39 (67%) | 19 (33%) |

compatible with the hypothesis that the same word order constraints hold for the middlefield in main and embedded clauses.

3.2. The Order Between Subject and Object: Basic Results

The basic results of the current corpus study are shown in Table 3. As detailed above, these results do not include any sentences with pronominal subjects or sentences with either one or four NP arguments. Furthermore, prefield sentences are sentences in which either the subject or the object is in the prefield.

As already shown in Table 2, sentences with two NP arguments are much more frequent than sentences with three NP arguments. The most striking finding with regard to sentences with three NP arguments is that these sentences are almost always SO-sentences, independently of whether the *den*-object is a dative or an accusative object. Given the low numbers of sentences with three NP arguments, we will not analyze them any further and concentrate on sentences with two NP arguments in the following. We suggest a possible reason for the high rate of SO order in sentences with three NP arguments in the final discussion; a detailed study of sentences with both an accusative and a dative object based on experimental results and on corpus data can be found in Häussler and Bader (accepted).

Two alternative views of the results for sentences with two NP arguments—that is, sentences with a subject and an object—are provided in Figure 1. The diagram on the left side of figure 1 shows the percentages of SO-sentences broken down by object case separately for middlefield and prefield sentences. This diagram reveals a striking effect of case: Sentences with an accusative object almost always occur with SO order whereas dative sentences do not exhibit a clear preference for either SO or OS order. For accusative sentences, sentence type has a small but significant effect on the rate of SO. In middlefield sentences, the rate of SO reaches almost 99%, whereas in prefield sentences the rate of SO is significantly lower although it is at a still high 86% (99% vs. 86%; $\chi^2 = 58, p < .01$). For dative sentences, the difference between middlefield and prefield sentences is not significant (42% vs. 54%; $\chi^2 = 1.4, n.s.$).

An alternative view of the current results is provided by the diagram on the right side of Figure 1. This diagram shows the distribution of object case expressed as percentages of accusative broken down by word order, again separately for middlefield and prefield sentences. This diagram shows that sentences with SO order predominantly occur with an accusative object, for both middlefield and prefield sentences (88% vs. 91%; $\chi^2 = 2, n.s.$). Sentences with

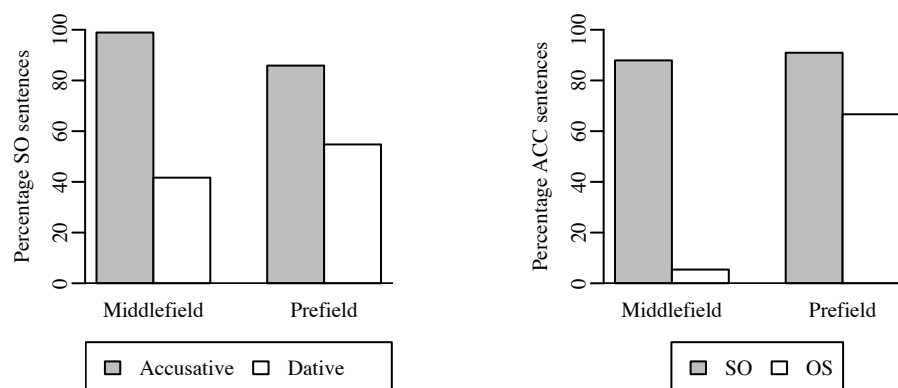


Figure 1: Left side: Percentages of SO-sentences broken down by case and position within the sentence. Right side: Percentages of accusative sentences broken down by word order and position within the sentence.

OS order, in contrast, usually have a dative object when the object is in the middlefield whereas objects in the prefield show a preference for accusative case, which, however, is not as strong as the dative preference in the middlefield (5% vs. 67%; $\chi^2 = 72, p < .001$).

In summary, the results presented in this section show that the order between subject and object in German clauses is dependent on the case of the object, the number of NP arguments, and the placement of the arguments in either the middlefield alone or in both prefield and middlefield. In the next section, we discuss the representativeness of our results.⁵

3.3. Are the Current Corpora Representative?

Before we can proceed with a detailed discussion of our corpus results, we have to address the issue of how the restriction to sentences which contain an object introduced by the article *den* affects the results that are presented in this paper. This issue must be considered in two respects. First, does this restriction lead to a distorted picture of the true relationships which hold for a corpus free of any peculiar restrictions? Second, what limitations does this restriction imply with regard to the conclusions that we can draw from our particular corpora? In this section, we present evidence arguing that the corpora that we have assembled are representative with regard to the issues at hand. At the same time, we also note certain necessary limitations that result from analyzing only sentences with *den*-objects.

Restricting our analysis to sentences containing an object with the article *den* raises two different issues. First, *den* is a form of the definite article. Thus, whereas the subject can freely vary in terms of definiteness, the object is always a definite NP. Second, considering only objects with the article *den* leads to a confounding of case and number in our corpora. Accusative objects are always singular NPs whereas dative objects are always plural NPs. How might these two properties of *den*-objects affect the results of the current study?

A first way of assessing the representativeness of our data is provided by Figure 2. This figure shows the interaction between order and case in three different corpus studies of word order in the middlefield: the prior studies by Hoberg (1981) and Kempen and Harbusch (2005) (cf. Table 1), and our own study. The three graphs in Figure 2 show a converging pattern: First, SO-sentences are much more frequent than OS-sentences. Second, there is a strong interaction between order and case. In SO-sentences, the majority of objects occur with accusative case; in OS-sentences, in contrast, dative objects outnumber accusative objects by far.

With regard to the relationship between order and case, our own data exhibit a pattern similar to the two prior corpus studies in which the object was not restricted in any way, although the exact numbers differ somewhat. In

⁵Sentences introduced by a *den*-object also figured prominently in a recent dispute between Bornkessel et al. (2002, 2003) and Kempen and Harbusch (2002). Bornkessel et al. (2002) claim that the brain differentiates between accusative and dative first OS-sentences whereas corpus data show that accusative and dative first sentences occur with equal frequencies. This latter claim has been disputed by Kempen and Harbusch (2002). Besides other corpus evidence, they analyzed 300 sentences beginning with a *den*-object randomly sampled from the same corpus that we used for initial sampling and found a rate of 95% dative initial sentences. Our own data closely agree with these results.

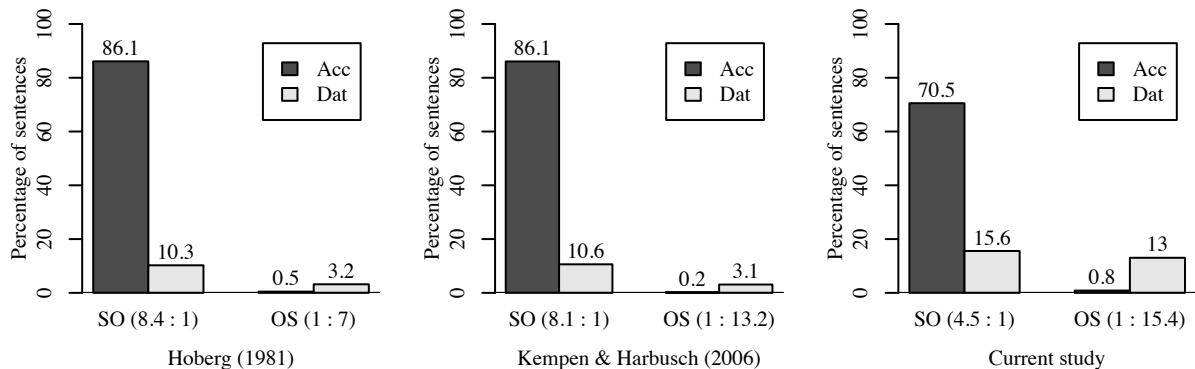


Figure 2: The interaction of order and case in three different corpus studies of word order in the middle field.

order to exclude the possibility that these differences might be the result of the confounding of number and case, we obtained additional frequency counts from the TIGER Treebank (Version 2) created by the Universities of Stuttgart, Saarbrücken and Potsdam. This treebank includes approximately 50.000 sentences of German newspaper text, annotated with morphosyntactic information including number.⁶ For both singular and plural objects, the table in (18) shows how word order depends on the case of the object.

| | Singular Object | | Plural Object | |
|--------|-----------------|------------|---------------|------------|
| | Acc | Dat | Acc | Dat |
| SO-Acc | 86.9% (760) | 10.7% (94) | 83.4% (281) | 14.5% (49) |
| OS-Acc | 0.1% (1) | 2.3% (20) | 0% (0) | 2.1% (7) |

The table in (18) reveals quite similar results for sentences with a singular object and sentences with a plural object. The overall ratio between SO and OS sentences is almost identical for sentences with singular objects and sentences with plural objects (97.6% SO versus 97.9% SO; $\chi^2 = 0.02, n.s.$). For SO-sentences, the bias towards accusative is somewhat more pronounced when the object is a singular NP (760 accusative versus 94 dative sentences) than when the object is a plural NP (281 accusative versus 49 dative sentences). However, this difference did not reach full significance ($\chi^2 = 2.96, p = 0.086$). Furthermore, even if this difference had turned out to be fully significant, its magnitude is quite small—89% versus 85% accusative SO sentences—which makes it unlikely that it might have any relevant repercussions of the upcoming results. The possible interaction between case and number cannot be tested statistically for OS sentences due to their low numbers. Note finally that the corpus study by Bresnan et al. (2004) included number as a possible predictor for the English dative alternation but found that number had no significant effect.

The data presented in this section show that the interaction between order and case which is visible in our data is not a mere side effect of the correlation between case and number which results from considering only objects introduced by the article *den*. Instead, this interaction must be due to a genuine case effect. We thus conclude that it is highly improbable that the correlation between number and case inherent in our corpora affects the representativeness of our results.

With regard to main clauses, we do not know of existing corpus studies providing information about the interaction of order and case. However, we can make two other comparisons. First, Table 8 on page 162 of Hoberg (1981) gives corpus counts of the elements occupying the prefield. As revealed by this table, in the corpus investigated by Hoberg (1981) (see section 2.2 for details) 63.15% of all main clauses had a subject in the prefield, 10.15% another argument of the verb, and 25.06% some kind of adverbial. In our unrestricted main-clause corpus, the prefield SO/OS corpus, 58.5% of the sentences had the subject in the prefield, 12.4% the *den*-object, and 29.1% all other kinds of phrases.

⁶Further information about the Tiger Treebank can be found at <http://www.ims.uni-stuttgart.de/projekte/TIGER>. A comprehensive analysis of the TIGER corpus (Version 1) with regard to the order between syntactic functions can be found in Kempen and Harbusch (2005).

We thus see quite similar rates of subject-initial main clauses. The remaining categories overlap only partially, so the similar numbers for them must be taken with caution. The second comparison is based on the corpus study of Weber and Müller (2004). This study investigated the order between subject and object in main clauses with either subject or object in the prefield; it is thus similar to our corpus analysis. In the corpus analyzed by Weber and Müller (2004), a rate of 18.4% sentences with OS order was observed. This is almost identical to the rate of 17.5% sentences with OS order that we have found in our prefield SO/OS corpus (see Table 3). Taken together, the two comparisons involving main clauses indicate that our main clause corpora are also representative, despite being restricted to objects introduced by *den*.

Of course, considering only sentences with a *den*-object necessarily limits the range of conclusions that we can draw. The most important limitation concerns the influence of semantic-pragmatic properties tied to determiner choice. Due to this limitation, it is not possible to provide a rigorous evaluation of the Familiarity Hierarchies in (6), although—as will be demonstrated in section 4.3—a partial evaluation is still possible. Furthermore, we will not be able to look at certain more specialized word order patterns, like the one in (19) involving the relationship between order and scope.

- (19) a. Ich weiß, dass [jeder Student] [mindestens ein Buch] gelesen hat.
 I know that every student at-least one book read has
 ‘I know that every student read at least one book.’
- b. Ich weiß, dass [mindestens ein Buch] [jeder Student] gelesen hat.
 I know that at-least one book every student read has
 ‘I know that at least one book every student read.’
- c. Ich weiß, dass [mindestens ein Buch] [von jedem Studenten] gelesen wurde.
 I know that at-least one book by every student read was
 ‘I know that at least one book was read by every student.’

In (19a), the subject precedes and takes scope over the object. In order to give the object scope over the subject, the order between the two has to be reversed, but this can be achieved in two ways. As in (19b), the object can simply be put in front of the subject, leaving all else equal.⁷ Alternatively, the sentence can be passivized, as shown in (19c). Since this not only affects order, but also the assignment of syntactic functions, sentence (19c) is not an OS sentence but a sentence with a subject followed by a prepositional phrase. Our corpus analysis cannot answer the question of whether OS order is used for reasons of scope, or whether scope relations are signaled by alternative syntactic means, like passivization. While this clearly is a limitation of the current analysis, we suspect that questions of this kind can only be addressed by compiling corpora tailored to the particular questions at hand. Otherwise, there will simply be too few examples for a thorough analysis. For example, the 50000 sentences of the Tiger Treebank discussed above contain 2181 accusative objects introduced by the determiner *den* but just 19 accusative objects introduced by the universal quantifier *jeden* (‘every’).

The preceding remarks are also relevant when we turn again to Figure 3. A closer inspection of this figure reveals that the results obtained by Hoberg (1981) and Kempen and Harbusch (2005) are more similar to each other than to our results. The main difference concerns the rate of OS sentences. In the data of Hoberg (1981) and Kempen and Harbusch (2005), this rate is about 3.5% whereas it is about 14% in our data. We attribute this difference to the fact that in the sentence sets that form the basis of the current study, the object always started with the definite determiner *den*. Given the preference for definite NPs to precede indefinite NPs (cf. section 2.1 above and section 4.3 below), it is to be expected that in a corpus that contains only sentences with definite objects the rate of OS should be higher than in a corpus where the object is unconstrained with respect to definiteness.

4. What Determines the Order Between Subject and Object?

In this section we explore several of the word order hierarchies discussed in Siewierska (1993) (see the complete list in (4), (5), and (6)). Only sentences with two NP arguments will be considered for the reasons discussed above. As mentioned in section 2.2, the corpus study by Hoberg (1981) found animacy to be the main predictor of the order

⁷Sentences like (19b) are often considered as scopally ambiguous; cf. Frey (1993) and Pafel (2006).

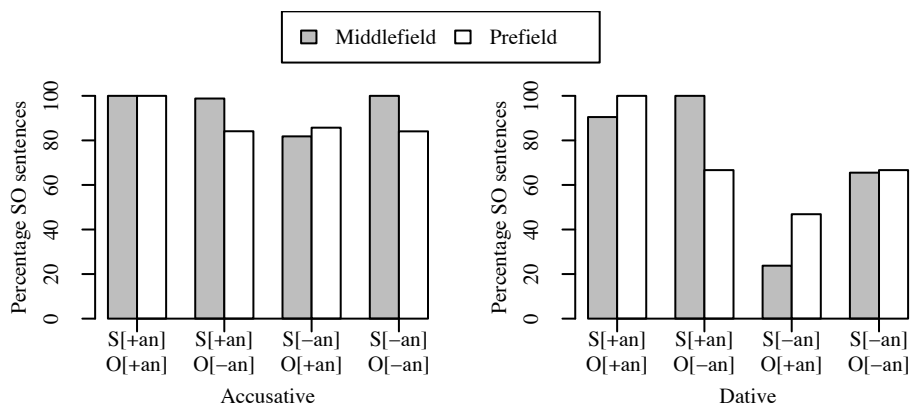


Figure 3: Percentages of SO-sentences broken down by position (middlefield versus prefield) and animacy properties of subject and object ([+an] = animate; [-an] = inanimate). Results for sentences with accusative object are shown on the left and results for sentences with dative object on the right.

between non-pronominal arguments, with animate arguments preceding inanimate arguments. We therefore begin with the role played by animacy.

4.1. The Animacy Constraint

Among the word order hierarchies reviewed by Siewierska (1993), animacy is part of the personal hierarchy which is one of the two lexical-semantic hierarchies. The personal hierarchy is repeated in (20).

(20) The Personal Hierarchy:

1stp. > 2ndp. > 3rdp. human > higher animals > other organisms > inorganic matter > abstracts

Although the number of OS sentences included in the current investigation is large in comparison to prior investigations, it is still too small to allow an investigation of a hierarchy as fine-grained as the one in (20). Furthermore, the first two positions of this hierarchy cannot be instantiated in our corpora because all objects are *den*-objects and sentences with subject pronouns were removed from the analysis. We therefore collapsed the personal hierarchy into two categories, with inanimate NPs subsuming the two categories at the lower end of the hierarchy (inorganic matter and abstracts) and animate NPs subsuming all remaining categories. This gives us the Animacy Constraint shown in (21).⁸

(21) The Animacy Constraint

animate > inanimate

To test the Animacy Constraint, all NP arguments were annotated as follows. First, a three-way classification was used: besides animate NPs and inanimate NPs, NPs denoting institutions, corporations, and the like were coded separately. As is well known, NPs of the latter type often behave like animate NPs. For example, they may fill the subject slot of agentive verbs, as shown in (22).

(22) The manager/The company tried to increase Christmas sales.

In a second step, we recoded all NPs of this type as animate. All the results that we present in this section were also computed with institutional NPs completely removed from the corpus. Since this did not affect the observed patterns, we present results with institutional NPs subsumed under animate NPs.

The relation between animacy and word order can be considered in two directions. First, do particular combinations of animacy features associated with subject and object imply the choice of either SO or OS word order?

⁸From here on, we will refer to hierarchies with only two elements as constraints.

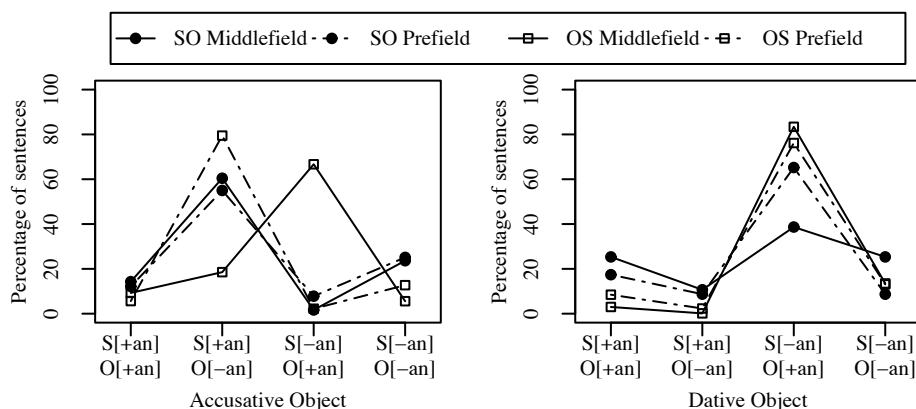


Figure 4: Distribution of all sentences with order and position of subject and object fixed across the four combinations of subject and object animacy ([+an] = animate; [-an] = inanimate). Results for sentences with accusative object are shown on the left and results for sentences with dative object on the right.

Second, are the two word orders SO and OS associated with particular combinations of animacy features on subject and object? We consider these two questions in turn, starting with the first one because this is the question which has to be answered in order to assess the validity of the Animacy Constraint. For each of the four combinations of animacy features and syntactic functions, Figure 3 shows the percentages of SO sentences for both middlefield and prefield sentences, separated into accusative and dative objects. Considering first sentences with an accusative object, we see high percentages of SO sentences in all conditions. When both the subject and the accusative object are in the middlefield, the percentage of SO sentences is almost at 100% with one exception, namely sentences with inanimate subject and animate object. But even in this condition, the rate of SO is at a still high 84%. For sentences with either subject or accusative object in the prefield, sentences where both are animate also reach 100% SO word order. For all other animacy combinations the rate is about 85%.⁹

Turning next to sentences with a dative object, Figure 3 presents a rather different pattern. When both subject and dative object are in the middlefield, the overwhelming majority of sentences with animate subjects show SO word order, regardless of whether the object is animate or inanimate. When the subject is inanimate, in contrast, there is no clear cut preference when the object is also inanimate but a reversal of the preference toward OS when the object is animate. Considering finally sentences in which either the subject or the dative object is located in the prefield, we see yet another pattern. While sentences in which both subject and object are animate have a strong preference for SO order, no other combinations of animacy features show strong preferences in favor of either SO or OS order.

We now turn to the reverse relation between animacy and word order and ask whether the two orders SO and OS imply particular animacy features on subject and object. Figure 4 shows the distribution of animacy features on subject and object separately for accusative and dative case. Each line in this figure represents one of the four possible combinations of order (SO vs. OS) and sentence type (middlefield vs. prefield), and the y-axis shows the percentages of sentences within the four combinations of animacy features (animate vs. inanimate) and syntactic function (subject vs. object) relative to each line. Thus, the values shown on the y-axis sum up to 100% for each line. The lines for SO sentences are based on the SO/OS corpora; the lines for OS sentences are based on the OS corpora.

For accusative sentences, which are shown on the left in Figure 4, a bimodal pattern is visible. With one exception, the dominant animacy pattern for sentences with an accusative object is the pattern S[+an]O[-an]. The one exception is provided by OS middlefield sentences: When an accusative object precedes the subject in the middlefield, the pattern S[-an]O[+an] dominates.

For all sentence types with a dative object, the combination S[-an]O[+an] is the most frequent. For both middlefield and prefield OS-sentences, the predominance of S[-an]O[+an] is very strong, and for SO prefield sentences, it is only slightly weaker. For SO sentences in the middlefield, the predominance of S[-an]O[+an] is still there, but it is

⁹We do not present chi-square tests because of several cells with too few observations.

much attenuated. Interestingly, the main reason for this is not that SO middlefield sentences have a particularly high rate of S[+an]O[+an]—although the rate of this animacy pattern is in fact highest for SO middlefield sentences—but instead because there are quite a few SO middlefield sentences with the pattern S[-an]O[-an].

A noteworthy aspect of these data is that a considerable percentage of dative objects is inanimate. Prototypical thematic roles for dative objects are recipient and experiencer and therefore dative objects normally refer to animate entities. Nevertheless, the rate of inanimate dative objects varied between 13% and 21% among our four corpora. Although some rare verbs with an inanimate dative object have often been cited in discussions of animacy and word order—in particular the two ditransitive verbs *jemanden-ACC einer Sache-DAT unterziehen* (‘to subject someone to something’) and *jemanden-ACC einer Sache-DAT aussetzen* (‘to expose someone to something’)—it still seems remarkable that between every fifth and every tenth dative object had an inanimate referent in our corpora. Two representative examples are shown in (23) and (24).

- (23) dass [den gewandelten Anforderungen]_{DAT} Rechnung getragen wird
 that the changed requirements account carried is
 ‘that the changed requirements are taken into account.’
- (24) dass [den Sanktionen gegen Österreich]_{DAT} jegliche Rechtsgrundlage fehlt
 that the sanctions against Austria any legal-basis lacks
 ‘that the sanctions against Austria lack any legal basis’

In summary, for both accusative and dative case, OS order in the middlefield occurs mainly when the subject is inanimate and the object animate. Furthermore, sentences with animate subjects occur mainly in the order SO. When comparing middlefield OS and prefield OS sentences, a difference depending on the case of the object emerges. For dative case, prefield OS and middlefield OS sentences show the same pattern; for accusative case, in contrast, prefield OS sentences pattern together with middlefield SO sentences. While our data thus clearly suggest that the Animacy Constraint has a strong effect on the order between subject and object in German, there are also several exceptions. This is in particular true for sentences with the pattern S[-an]O[+an]. Only dative middlefield sentences of this type exhibit a clear preference for OS (although a quarter of such sentences still occur with SO order). Accusative sentences with S[-an]O[+an] usually have SO order and for dative prefield sentences of this type no preference is visible.

With regard to possible correlations among word-order hierarchies, we consider in section 4.4 whether the effects of the animacy hierarchy might be side-effects of the factor weight. Since the animacy properties of subject and object are strongly constrained by the semantic roles that are assigned by the main verb of the sentence, the Semantic Role Hierarchy is another candidate hierarchy that might correlate with the animacy hierarchy. This issue is considered in depth in section 5.3.

4.2. The Semantic Role Hierarchy

We next consider the second of the two lexical-semantic hierarchies discussed by Siewierska (1993), the Semantic Role Hierarchy which is repeated in (25).¹⁰

- (25) The Semantic Role Hierarchy:
 agent > patient > recipient > benefactive > instrumental > spatial > temporal

For both practical and theoretical reasons, evaluating our corpus data with regard to the Semantic Role Hierarchy is anything but straightforward. As is well-known, various mutually incompatible variants of the Semantic Role Hierarchy exist. Levin and Rappaport Hovav (2005) provide a detailed overview of this issue and come to the conclusion that the existing disagreement is to a large extent due to the fact that role hierarchies have been postulated with different purposes in mind. The most important one is argument selection or linking, that is, predicting which of the arguments of a verb is realized as the subject, which as the direct object, etc. Among the other purposes to which semantic role hierarchies have been put to use is the issue of argument linearization. However, it is controversial whether an appeal

¹⁰A semantic role which is often found in the literature but which is missing in (25) is the role of ‘theme’. The theme role is subsumed by the patient role in (25). The labels ‘patient’ and ‘theme’ are notoriously confusing due to their inconsistent use (for a recent survey cf. Levin and Rappaport Hovav, 2005: 48f). Sometimes both terms are used interchangeably. However, when distinguished, ‘theme’ is used for entities undergoing movement or being located while ‘patient’ is used for entities being affected in a broader sense.

to semantic roles is really necessary for this purpose. According to Siewierska (1993), only the agent role is clearly implicated in linearization: NPs realizing agents typically precede all other NPs, but, as she notes, even this might be just a side-effect of the generalization that agents are typically subjects. Furthermore, argument realization and argument linearization seem to imply different orderings among some of the roles. The two roles which according to Levin and Rappaport Hovav (2005) have caused the largest amount of disagreement with regard to their relative ranking are patient and recipient. For current purposes, this is especially problematic because recipients are typically realized as dative objects, and dative objects play an important role within the current investigation.

A particularly clear illustration of this problem is provided by the behavior of ditransitive verbs. Normally, the arguments of a ditransitive verb in German are serialized in the order “subject > dative object > accusative object” (cf. references in footnote 1).

- (26) Gestern hat der Pfarrer dem Opa den neuen Hut gebracht.
 yesterday has the minister the grandpa the new hat brought
 ‘Yesterday, the minister brought grandpa the new hat.’

When a sentence like (26) is passivized, the former subject is lost and the former accusative object is promoted to subject, as expected. Importantly, however, passivization in German does not force the former accusative object to change its place. Thus, an unmarked sentence in the passive voice would have the word order in (27).¹¹

- (27) Gestern wurde dem Opa der neue Hut gebracht.
 yesterday was the grandpa the new hat brought
 ‘Yesterday, the new hat was brought to grandpa.’

(27) is a sentence with OS word order. Ditransitive verbs in the passive voice have accordingly been argued to provide a systematic source for exceptions to the otherwise canonical SO word order of German (cf. Lenerz, 1977). In terms of semantic roles, however, the active clause and its passive counterpart do not differ (exempting the original subject). As shown in (28), the recipient argument precedes the patient argument in the active as well as in the corresponding passive clause.

| | <i>Agent[+animate]</i> | <i>Recipient[+animate]</i> | <i>Patient[-animate]</i> |
|----------------|------------------------|----------------------------|--------------------------|
| <i>Active</i> | NOM | DAT | ACC |
| <i>Passive</i> | (von[‘by’]-PP) | DAT | NOM |

What (28) suggests is that semantic roles are more important for word order in German than syntactic functions. Note, however, that in the Semantic Role Hierarchy in (25) patient is higher ranked than recipient which would not give the right result for the order seen in (28). It might be tempting to simply rerank these two roles—and such reverse rankings have indeed been proposed—but then the hierarchy in (25) would no longer fulfill its primary purpose, namely to predict argument selection. Since Keenan and Comrie (1977) introduced the *syntactic function hierarchy*, syntactic functions have been assumed to be hierarchically arranged in the order “subject > direct object > indirect object”. The arguments of three-place verbs must therefore be arranged in the order “agent > patient > recipient” in order to match the syntactic function hierarchy. We are not in a position to resolve this issue here, but note that—as indicated in (28)—recipients are typically animate and patients typically inanimate. With the exception of the position of the subject, this correlation opens up the possibility that the ordering seen in (28) does not follow from the Semantic Role Hierarchy but from the Animacy Constraint (or, to put it more generally, the Personal Hierarchy) instead. We will come back to this issue at a later point.

Passivized ditransitive verbs are not the only verbs which have been argued to preferentially occur with OS order. Already Lenerz (1977) listed severally semantically defined verb classes with this property. Based on a recent survey

¹¹There are two alternative ways to translate a German passive clause like (27), either as in (ia) or as in (ib).

- (i) a. Yesterday, the new hat was brought to grandpa.
 b. Yesterday, grandpa was brought the new hat.

Whereas (ia) is faithful to the syntactic functions of the original German sentence, (ib) is faithful to the order between the two arguments. Because the subject obligatorily precedes the object in English, there can be no translation which is faithful to both syntactic function assignment and word order. In the following, we will give translations that preserve the assignment of syntactic functions.

of the relevant literature, Eisenberg (2004) identifies the three major subclasses of OS verbs shown in (29) (verbs from Lenerz, 1977 and Eisenberg, 2004; verbs selecting *sein* ('to be') as perfect auxiliary are printed in italics).

- (29) a. Object (Acc or Dat) = Experiencer:
 (i) Acc: *anekeln* ('to disgust'), *begeistern* ('to inspire'), *entsetzen* ('to appall'), *erfreuen* ('to please'), *überraschen* ('to surprise'), ...
 (ii) Dat: *auffallen* ('to strike'), *behagen* ('to please'), *einleuchten* ('to be clear'), *gefallen* ('to appeal to'), *naheliegen* ('to be important to'), *widerstreben* ('to jib'), ...
 b. Object (Dat) = Cause: *gelingen* ('to succeed'), *glücken* ('to succeed'), *mißraten* ('to turn out badly'), *passieren* ('to happen to'), ...
 c. Object (Dat) = Possessor: *gehören* ('to belong'), *fehlen* ('to be lacking'), *zukommen* ('to receive'), *zustehen* ('to be entitled'), ...

All three subclasses have in common that they have an animate object and a subject which is either obligatorily or at least prototypically filled by an inanimate NP. Probably the best-known class is the class of psych verbs that realize the experiencer as object (see (29a)). Three examples are given in (30).¹²

- (30) a. ... dass **den** Opa der Witz gelangweilt **hat**.
 that the-ACC grandpa the-NOM joke bored has
 '... that the joke bored grandpa.'
 b. ... dass **dem** Opa der Witz gefallen **hat**.
 that the-DAT grandpa the-NOM joke pleased has
 '... that the joke pleased grandpa.'
 c. ... dass **dem** Opa der Witz eingefallen **ist**.
 that the-DAT grandpa the-NOM joke occurred is
 '... that the joke occurred to grandpa.'

The examples in (30) illustrate two points. First, the experiencer object of a psych verb can occur with either accusative case (cf. (30a)) or with dative case (cf. (30b)/(30c)). Second, object-experiencer verbs with a dative object can select either *haben* or *sein* as perfective auxiliary; object-experiencer verbs with an accusative object always select *haben*, like verbs with accusative objects in general. Verbs selecting *sein* ('to be') as perfect auxiliary constitute the class of *unaccusative verbs*. As discussed in Grewendorf (1989) and subsequent work (summarized in Eisenberg, 2004), unaccusative verbs have a range of properties which set them apart from verbs selecting *haben* ('to have') as perfect auxiliary. The latter are sometimes also called *unergative*; to ease readability, we will use the terms *sein*-verbs and *haben*-verbs in the following. In the current context, the two most important properties of unaccusative or *sein*-verbs are that they do not have an agent argument and that—perhaps as a consequence of the first property—they typically occur with OS order. These two properties are characteristic for unaccusative verbs, but they are not restricted to them. As shown in (29), a small group of *haben* verbs has the same properties.

The other two classes of OS verbs are illustrated in (31).

- (31) a. ... dass dem Opa ein Malheur passiert ist.
 that the grandpa-DAT a mishap-NOM happened is
 '... that a mishap happened to grandpa.'
 b. ... dass dem Opa eine Villa gehört hat.
 that the grandpa-DAT a villa-NOM belonged has
 '... that a villa belonged to grandpa.'

In contrast to the experiencer class, the cause and possess class contain only verbs with a dative object. Verbs in the cause class share a basic meaning of 'something happened'. What happened is expressed by the subject; the dative object is causally involved but—in contrast to typical agents which are realized as subjects—it is not the controller of the event. Verbs in the possess class assert or deny a relation of possession in which the dative object plays the role

¹²Many accusative and some dative psych verbs with experiencer object can take on an agentive reading when the subject is a potential agent (in particular, a human being). A detailed discussion of this issue including relevant experimental results can be found in Scheepers (1996).

of possessor. Note finally that verbs in the ‘cause’ class typically are *sein*-verbs whereas the ‘possess’ class contains verbs with either perfect auxiliary.

In summary, several semantically defined classes of verbs occurring preferentially with OS order have been proposed in the syntactic literature. A common feature of these verbs is that they do not have an agent argument. Since subjects which realize the agent role typically precede all other arguments, having no agent is a prerequisite for licensing OS order in terms of the Semantic Role Hierarchy.

To a considerable extent, it is possible to identify OS verbs by morpho-syntactic criteria. Two classes of OS verbs are associated with unique auxiliaries. First, a passive auxiliary reveals a ditransitive verb in the passive voice. Second, the perfect auxiliary *sein* identifies an unaccusative verb. The third class of OS-verbs cannot be identified by purely formal means because verbs in this class behave like the majority of German verbs in selecting *haben* as perfect auxiliary. It might be possible to define this class in semantic terms, but in the following we will simply work with a list of common verbs in this class at relevant points.

We thus arrive at the following prediction. If the Semantic Role Hierarchy is involved in determining the order between subject and object in German clauses, sentences with SO word order should contain a different set of verbs than sentences with OS word order. In particular, sentences with OS order should show a particularly high proportion of sentences containing the three types of verbs discussed above. SO sentences, in contrast, should contain such verbs only rarely if at all. We next present a test of this prediction instead of a direct test of the Semantic Role Hierarchy because encoding all verbs in terms of their lexical-semantic properties is a task that is beyond the scope of the current paper. This test will proceed in two steps. In the first step, we determine the proportion of active sentences with *haben*-verbs, active sentences with *sein*-verbs, and passive sentences. The proportion of the latter two sentence types should be substantially higher for OS- than for SO-sentences. In the second step, we analyze the *haben*-verbs that are found in active sentences in more detail. OS-sentences should contain mainly non-agentive verbs of the kind discussed above; in SO-sentences, such verbs should occur rarely if at all.

In order to determine the proportion of the three sentence types identified above, we coded all sentences in the following way. First, all sentences with a past participle and a form of the passive auxiliary *werden* were classified as passive clauses. Secondly, the remaining active clauses were further subdivided according to the auxiliary that the main verb takes when it occurs in the perfect tense. To this end, a list of all main verbs occurring in our sentence sets was compiled. For each verb in this list, we coded whether it takes *haben* or *sein* in the perfect tense. Based on this list, the main verb in each corpus sentence was classified as either a *haben*- or a *sein*-verb.¹³

Note that passivized verbs—understood as a verb cluster consisting of the past participle of a main verb and a passive auxiliary—take *sein* as perfect auxiliary although the main verb in the active voice takes *haben*. Our three-way classification into passivized verbs, *haben*-verbs and *sein*-verbs thus reflects three binary distinctions: (i) The verb cluster is in the active or the passive voice, (ii) the verb cluster as a whole takes either *haben* or *sein* as perfect auxiliary, and (iii) the main verb in the verb-cluster takes either *haben* or *sein* as perfect auxiliary. Since these three binary distinctions are not independent of each other (e.g., with very rare exceptions, only *haben*-verbs can be passivized; the perfect auxiliary of the passive auxiliary itself is *sein*), we end up with only three of the eight possible combinations.

Figure 5 shows the distribution of voice/auxiliary in sentences with a dative object, broken down by sentence type and word order. Results for SO sentences are based on the SO/OS corpora, results for OS sentences on the OS corpora. Several noteworthy results are revealed by this figure. First, for middlefield sentences word order has a strong effect on the distribution of voice/auxiliary (SO: 61% active/*haben*-verb - 28% active/*sein*-verb- 11% passive vs. OS: 20% - 40% - 41%; $\chi^2 = 69$, $p < 0.01$). While a clear majority of SO-sentences has an active *haben*-verb, in sentences with OS order active *haben*-verbs are a minority. Just the opposite pattern holds for passivized verbs which barely occur in SO-sentences but account for about 40% of all OS-sentences. The proportion of active *sein*-verbs, finally, also shows an increase when going from SO- to OS-sentences, although not as strong as for passive clauses.

Sentences with either subject or object in the prefield show a similar picture to middlefield sentences, although the effects are much attenuated and the interaction is not significant (SO: 48% - 30% - 22% vs. OS: 35% - 28%

¹³The classification into active versus passive sentences is based on the actually occurring verb form. Active sentences were additionally coded for auxiliary selection in the perfect tense independently of whether a particular occurrence of a verb in the corpus was in the perfect tense or not.

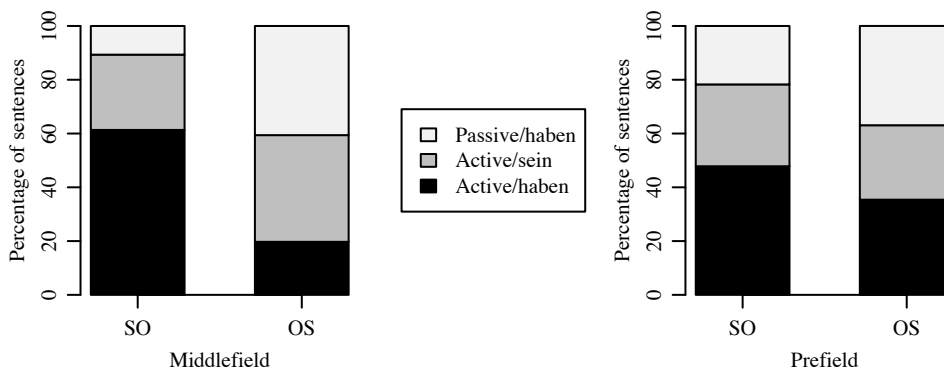


Figure 5: Distribution of verb properties in sentences with a dative object according to order (SO vs. OS) and position (middlefield vs. prefield).

Table 4: The *haben*-verbs which occur with highest frequency in sentences with either SO or OS order, for verbs with accusative object and verbs with dative object. Total numbers are given in parentheses.

| | Accusative | | | | Dative | | | |
|---|--------------|----------|---------------|----------|-------------|----------|-------------|-----------|
| | SO order | | OS order | | SO order | | OS order | |
| | verb (n=317) | f(n=546) | verb (n=40) | f (n=54) | verb (n=31) | f (n=46) | verb (n=41) | f (n=152) |
| 1 | haben | 19 | treffen | 8 | entsprechen | 9 | stehen | 41 |
| 2 | stellen | 11 | befallen | 2 | zustimmen | 4 | *fehlen | 31 |
| 3 | erhalten | 9 | beißen | 2 | gelten | 3 | liegen | 15 |
| 4 | finden | 9 | beschleichen | 2 | helfen | 3 | *zustehen | 8 |
| 5 | verlassen | 9 | drücken | 2 | | | *gefallen | 5 |
| 6 | bekommen | 8 | ereilen | 2 | | | stocken | 4 |
| 7 | bringen | 8 | interessieren | 2 | | | drohen | 3 |
| 8 | machen | 8 | plagen | 2 | | | *gehören | 3 |
| 9 | schaffen | 8 | | | | | winken | 3 |

- 37%; $\chi^2 = 2.2$, n.s.):¹⁴ The rate of active *haben*-verbs decreases when going from SO to OS whereas the rate of passivization increases. For active *sein*-verbs, there does not seem to be a difference between SO and OS. Overall, prefield sentences show a pattern that lies between the two extremes found in middlefield sentences; more specifically, prefield SO sentences are more similar to middlefield SO sentences and prefield OS sentences to middlefield OS sentences. We will come back to this finding in our attempt to model word order variation in German.

In summary, over 80% of all middlefield OS sentences with dative objects either contain a passivized ditransitive verb or an unaccusative verb—that is, one of those verbs which do not have an agent argument. This finding already suggests that OS word order in the middlefield is primarily driven by lexical-semantic factors, not by discourse factors. Since verbs which favor an OS structure for lexical-semantic reasons (either directly or indirectly via passivization) much more often have a dative object than an accusative object, the striking difference between the rate of accusative and dative case in OS middlefield sentences also follows.

We now come to the second step of our evaluation of the role that the Semantic Role Hierarchy might play for determining the order between subject and object. As revealed by Figure 5, for middlefield sentences the proportion of sentences with *haben*-verbs in the active voice is more than twice as high in SO-sentences than in OS-sentences. This finding is already in accord with the Semantic Role Hierarchy, but an even stronger test is possible. The proportion of *haben*-verbs should not only be larger in SO-sentences than in OS-sentences, but the set of *haben*-verbs occurring under the two orders should also be distinct. In particular, OS sentences should contain a high proportion of agentless verbs as listed in (29).

¹⁴The lack of a significant interaction for prefield sentences might also be due to the fact that the number of prefield sentences (n = 153) is much smaller than the number of middlefield sentences (n = 846).

A partial evaluation of this prediction is provided by Table 4. For each combination of order and case, this table lists the *haben*-verbs occurring with the highest frequency in sentences in the active voice. In each category, at most nine verbs are shown. In the two categories accusative/OS and dative/SO, fewer than nine verbs are shown because there were too many verbs in the next frequency range below the last one shown in Table 4 (28 verbs with a frequency of one for accusative/OS and 27 verbs with a frequency of one for dative/SO).

As shown by Table 4, there is no verb which appears among both the most frequent SO and the most frequent OS verbs. Overall, we found 12 accusative verbs and 8 dative verbs occurring with both SO and OS order. We can thus conclude that by and large SO and OS sentences in the active voice contain different sets of *haben* verbs, as predicted by the hypothesis that word order is intimately tied to verb-specific properties, in particular to verb semantics. Those verbs in Table 4 which are among the prototypical OS verbs listed in (29)—a list based on Lenerz (1977) and Eisenberg (2004)—have been marked with an asterisk. Four out of nine dative verbs in Table 4 and none of the accusative verbs are listed in (29). This low degree of overlap is due to the fact that Table 4 contains a substantial number of verbs which belong in the category of OS verbs only when used in an idiomatic or quasi-idiomatic way whereas the verbs in (29) are prototypical OS verbs in their basic usage.

For purposes of illustration, let us consider the verbs with the highest frequency. For accusative, this is the verb *treffen* which literally means ‘to hit’, but which almost exclusively occurred with the noun *Verschulden* (‘blame’) as subject, as illustrated in (32).

- (32) dass den Lehrer nur ein geringes Verschulden trifft.
 that the teacher only a small blame hits
 ‘that the teacher is to blame only to a small extent.’

In dative OS sentences, the most frequent verb is the verb *stehen* (‘to stand’) which takes a subject and a locative PP as arguments on its basic usage. In 29 out of 41 sentences found in our OS middlefield corpus, this verb occurred together with the PP *zur Verfügung* (lit. ‘to availability’), giving rise to an idiomatic use licensing an additional dative object. An illustrative example is shown in (33).

- (33) dass den Verbrauchern Qualitätsprodukte zur Verfügung stünden.
 that the consumers quality-products to availability stand
 ‘that quality products are available to the consumers.’

In summary, the detailed analysis of the *haben*-verbs occurring in the active voice provides further evidence for the assumption that verb semantics plays an important role in determining the order between subject and object. An obvious question raised by the findings of this and the preceding section is whether the Animacy Constraint and the Thematic Role Hierarchy—as reflected by the preferred use of distinct verbs in SO and OS sentences—are both necessary as determinants of word order. This question is addressed in section 5.

Let us finally note that the data presented in this section can provide an explanation of the striking interaction between case and order revealed by our corpus study as well as by the corpus studies of Hoberg (1981) and Kempen and Harbusch (2005): Why do sentences with OS word order in the middlefield occur much more often with a dative object than with an accusative object? Although Siewierska (1993) lists quite a few word order hierarchies, none of them makes reference to case. As we argue now, this is with good reason.

The three types of OS verbs distinguished above exhibit a clear asymmetry between accusative and dative case, as summarized in (34).

| (34) Verb type | Accusative Object | Dative Object |
|-------------------------------|-------------------|---------------|
| <i>haben</i> verbs | + | + |
| <i>sein</i> verbs | – | + |
| Passivized ditransitive verbs | – | + |

Dative objects can occur in all three constructions. Accusative objects, in contrast, are restricted to OS-licensing *haben*-verbs. If the verb list in (29) is representative, accusative OS verbs are further restricted to a single semantic class, namely the class of psych-verbs with experiencer object. In sum, (34) suggests that the preponderance of dative objects in OS middlefield clauses is a side effect of the structure of the lexicon which—for lexical-semantic reasons—provides many more dative OS verbs (simple or derived by passivization) than accusative OS verbs.

4.3. The Definiteness Constraint

The next constraint that we consider is the discourse-related Definiteness Constraint in (35).

(35) *Definiteness Constraint*

A definite NP precedes an indefinite NP.

The particular construction of our corpora does not allow a full fledged analysis of effects related to definiteness because we only sampled sentences containing a *den*-object and thus the object is always definite in the sentences under consideration. However, since the subject can vary freely with respect to definiteness, it is still possible to test the validity of the Definiteness Constraint. Given that we only consider definite objects, the Definiteness Constraint predicts that sentences with indefinite subjects should occur with OS order but not with SO order. When subject and object are both definite, in contrast, the Definiteness Constraint does not apply and order must be fixed by other factors.

In order to test this prediction, all subjects were coded as either definite or indefinite. The types of NPs falling into each class are shown in (36).

(36) a. Definite NPs:

NPs with definite article, proper names, NPs with demonstrative pronoun, NPs with definite quantifier (e.g., *jeder* ('every'), *die meisten* ('most'))

b. Indefinite NPs:

NPs with indefinite article, bare NPs, NPs with indefinite quantifier (e.g., *viele* ('many'), *keine* ('no'))

The overall rates of indefinite subjects in the four corpora under investigation are shown in (37). With one exception, about one quarter of subjects is indefinite. The exception is the middlefield OS corpus where the rate of indefinite subjects rises to 60%. This is already an indication that the Definiteness Constraint is at work in middlefield sentences, because the rate of indefinite subjects should be higher for a sentence set containing only OS sentences than for a sentence set containing a majority of SO sentences. Such a difference is absent from the prefield corpora, again indicating that there is a profound difference between word order in the middlefield and word order involving the prefield.

| (37) | Middlefield SO/OS-corpus | Middlefield OS-corpus | Prefield SO/OS-corpus | Prefield OS-corpus |
|------|-----------------------------|--------------------------|--------------------------|-----------------------|
| | 24% indef. subj. | 60% indef. subj. | 27% indef. subj. | 22% indef. subj. |

Table 5 shows the distribution of definite and indefinite subjects broken down by word order, case of the object, and sentence position. For OS sentences, values from both the SO/OS corpora and the OS corpora are shown. The former are needed for computing the rates of OS usage; given the small number of OS sentences within the SO/OS corpora, the values from the OS corpora allow an estimation of how reliable the values for OS are in the mixed SO/OS corpora. Considering middlefield sentences first, Table 5 reveals that definite subjects are much more frequent than indefinite subjects for SO sentences. In OS sentences, in contrast, the number of definite subjects is lower than the number of indefinite subjects. Furthermore, the case of the object does not interact with the distribution of definite and indefinite subjects.

For prefield sentences, the picture looks somewhat different. SO prefield sentences with an accusative object have about the same distribution of definite and indefinite subjects as their corresponding middlefield sentences. With OS order, in contrast, prefield sentences with an accusative object exhibit a clear difference: Whereas indefinite subjects are of somewhat greater frequency than definite subjects in the middlefield, prefield sentences show a very strong dominance of definite subjects. The number of dative prefield sentences is too small in the SO/OS prefield corpus to make any claims. In the OS prefield corpus, definite subjects and indefinite subjects occur about equally often, which is the same as what was found for dative OS sentences in the middlefield. In summary, for sentences with an accusative object, definiteness clearly has different effects for middlefield and prefield sentences. For sentences with dative object, middlefield and prefield sentences are more similar.

We now turn to an evaluation of the Definiteness Constraint according to which indefinite NPs should follow definite NPs. The table in (38) shows the percentages of OS order when the subject is indefinite and, for reasons of

Table 5: Distribution of definite and indefinite subjects in SO and OS sentences with either accusative or dative object, for middlefield and prefield sentences.

| | Accusative | | | Dative | | |
|--------------------|--------------|----------|-----------|--------------|----------|-----------|
| | SO/OS corpus | | OS corpus | SO/OS corpus | | OS corpus |
| | SO | OS | OS | SO | OS | OS |
| Middlefield | | | | | | |
| <i>Def. Subj</i> | 437 (80%) | 2 (33%) | 20 (37%) | 64 (85%) | 50 (48%) | 309 (40%) |
| <i>Indef. Subj</i> | 109 (20%) | 4 (67%) | 34 (63%) | 11 (15%) | 55 (52%) | 462 (60%) |
| Prefield | | | | | | |
| <i>Def. Subj</i> | 173 (75%) | 34 (89%) | 357 (87%) | 16 (70%) | 5 (26%) | 64 (49%) |
| <i>Indef. Subj</i> | 58 (25%) | 4 (11%) | 52 (13%) | 7 (30%) | 14 (74%) | 66 (51%) |

comparison, also when the subject is definite, although the Definiteness Constraint does not apply in this case. The percentages in this table are computed from the part of Table 5 showing the results for the two SO/OS corpora. Dative prefield sentences are left out because there were too few sentences to obtain reliable results.

(38)

| | Middlefield | | Prefield | |
|--------------------|-------------|--------|------------|--------|
| | Accusative | Dative | Accusative | Dative |
| <i>Indef. Subj</i> | 3.5% OS | 83% OS | 6.5% OS | - |
| <i>Def. Subj</i> | 0.5% OS | 44% OS | 16.4% OS | - |

Starting with middlefield sentences with a dative object, the table in (38) shows the effect predicted by the Definiteness Constraint. 83% of all sentences with an indefinite subject appear with OS, as expected for a definite object and an indefinite subject. As also shown in (38), the rate of OS falls to 44% when subject and object are both definite and the Definiteness Constraint cannot apply.

Sentences with an accusative object show a different picture. For middlefield sentences, the rate of OS is at a rather low 3.5% even when the subject is indefinite. Although this is still higher than the 0.5% OS rate when the subject is definite, it is clear from these results that the Definiteness Constraint is not obeyed by middlefield sentences containing an accusative object. The same point emerges even more strongly for sentences involving the prefield. Although the overall rate of OS is higher, the rate of OS with indefinite subjects is still at a rather low 6.5% which means that indefinite subjects most often precede a definite object, contrary to the Definiteness Constraint. Further, for these sentences the rate of OS is actually even higher when the subject is definite and thus the Definiteness Constraint cannot apply because subject and object are on equal footing with respect to definiteness.

As a final evaluation of the Definiteness Constraint, we look at the interaction of animacy and definiteness in dative sentences in the middlefield OS corpus. Only two combinations of subject and object animacy have enough instances for further investigation. For these two combinations, the numbers of definite and indefinite subjects are shown in (39).

(39)

| | O[+an]S[-an] | O[-an]S[-an] |
|----------------|--------------|--------------|
| O[+def]S[+def] | 284 (44%) | 17 (16%) |
| O[+def]S[-def] | 359 (56%) | 87 (84%) |

The table in (39) shows the expected interaction between animacy and definiteness. When the object is animate and the subject inanimate, there is a moderate preference for indefinite subjects (359 indefinite vs. 284 definite subjects; $\chi^2 = 8.7$, $p < 0.01$). When subject and object are both inanimate and thus do not differ in terms of animacy, indefinite subjects are strongly preferred (87 indefinite vs. 17 definite subjects; $\chi^2 = 47$, $p < 0.01$). This difference resulted in a significant interaction ($\chi^2 = 28$, $p < 0.01$). Thus, the Definiteness Constraint seems to be of only limited relevance when animacy strongly favors the use of the OS order, but it seems to have a substantial effect when there is a tie with respect to animacy.

In summary, despite the obvious limitations stemming from the fact that we only considered definite objects, the data presented in this section show clear effects of definiteness on the order between subject and object. The Definiteness Constraint, however, received only mixed evidence. A strong bias toward the ordering ‘definite NP before indefinite NP’ was found for middlefield sentences with a dative object. For accusative sentences, in contrast, the strong preference for SO is not overcome by definiteness properties of subject and object. This holds both for the middlefield and the prefield. In addition, we found evidence that the influence of definiteness is strongest when the choice between SO and OS is not already decided by the factor animacy.

4.4. The Weight Hierarchies

The idea that constituent weight influences word order dates back at least to Behagel’s ‘Gesetz der wachsenden Glieder’ (‘Law of Increasing Constituents’)(cf. Behagel, 1909/1910). In recent times, John Hawkins has built a far-reaching theory around the notion of constituent weight, declaring constituent weight to be the overarching word order determinant (Hawkins, 1994) or as one among several performance-based efficiency principles (Hawkins, 2004). We will consider Hawkins’s theory in the light of our results at the end of this section.

An important question with regard to constituent weight concerns the units in which to measure weight (cf. Wasow, 2002, for comprehensive discussion). The two general options that have been discussed in the literature are both reflected in the two weight-based hierarchies of Siewierska (1993) which are repeated in (40).

- (40) Weight Hierarchies
 - structurally simpler > structurally complex
 - short > long

Weight can be defined in terms of the syntactic complexity of phrases or in terms of constituent length. In both cases, there are several subtypes. For example, syntactic complexity might refer to number of phrasal nodes or to differences between categories, in particular between clauses and non-clauses. Length might, for example, be measured in number of syllables or number of words. In the following, we first consider weight in terms of structural complexity and then in terms of length.

4.4.1. Weight and Structural Complexity

We first consider the Complexity Constraint in (41).

- (41) Complexity Constraint
 structurally simpler > structurally complex

In order to test whether effects of this constraint can be observed in our corpora, we will present an analysis of the position of NP-related clauses, that is, clauses which are modifiers or complements of the subject or the object NP. An example of each type is shown in (42).

- (42) a. Peter hat *etwas* (gekauft) [_{CP} *das ich empfohlen habe*] (gekauft).
 Peter has something bought that I recommended have bought
 ‘Peter bought something that I had recommended.’
 b. Ich habe *die Hoffnung* (verloren) [_{CP} *das Rennen zu gewinnen*] (verloren)
 I have the hope lost the race to win lost
 ‘I have lost hope of winning the race.’

In sentence (42a), the NP-related clause is a relative clause; in sentence (42b), it is a complement clause. As also shown in (42), NP-related clauses can either follow their head noun immediately, or they can be extraposed across the clause-final verb, with the consequence that head noun and NP-related clause are separated by intervening material. Given the latter option, the following analysis not only addresses the Complexity Constraint in (41) as such, but also the more general question of where NP-related clauses are positioned within German clauses.

All subjects and objects were coded as either containing an NP-related clause (relative clause or complement clause) or not. Overall, the number of sentences containing NP-related clauses were as follows: 28 sentences in the middlefield SO/OS corpus, 62 in the middlefield OS corpus, 31 in the prefield SO/OS corpus, and 72 in the prefield OS corpus. Given these small numbers, we do not differentiate between the two types of NP-related clauses in

the following. There were only two sentences which contained more than one NP-related clause, both in the prefield SO/OS corpus. In these two sentences, one NP-related clause was located in the prefield and the other was extraposed, unambiguously in one and ambiguously in the other (see below).

The distribution of NP-related clauses in sentences with both subject and object in the middlefield is shown in (43). S-related in this table refers to NP-related clauses related to the subject, O-related to NP-related clauses related to the object. SO sentences are from the middlefield SO/OS corpus; OS sentences are from the middlefield OS corpus (numbers in brackets refer to the middlefield SO/OS corpus).

(43)

| | SO Sentences | | OS Sentences | |
|-------------|--------------|-----------|--------------|---------------|
| | S-related | O-related | S-related | O-related |
| NP internal | 0 | 3 | 7 (1) | 18 (2) |
| Extraposed | 1 | 18 | 37 (3) | 0 (1) |

When an NP-related clause belongs to a phrase directly in front of the clause-final verb (O with SO, S with OS), there is a clear preference for extraposition. The rate of extraposition is about 85% for both SO and OS sentences (3:18 for SO, 7:37 for OS). When an NP-related clause belongs to the clause-initial NP (S with SO, O with OS), the picture is different. The subject in SO sentences was accompanied by an NP-related clause only once, and this NP-related clause was extraposed. For the object in OS sentences, however, the OS corpus contains 18 NP-related clauses which all occur within the middlefield; in the SO/OS corpus, three OS sentences contained an NP-related clause belonging to the object, two within the middlefield and one in extraposed position.

With regard to the Complexity Constraint in (41), a clear majority of the data in (43) adheres to the ordering of ‘less complex’ in front of ‘more complex’. Given that none of the sentences under consideration contained more than a single NP-related clause, there are two ways to fulfill the Complexity Constraint. First, extraposed NP-related clauses are always in accordance with the Complexity Constraint because extraposed clauses necessarily follow all NP constituents, independent of where the originating NP is located within the middlefield. Second, NP-related clauses within the middlefield have to be part of the second NP (O in SO, S in OS) in order to adhere to the Complexity Constraint. Conversely, a middlefield internal NP-related clause which is part of the first NP (S in SO, O in OS) violates the Complexity Constraint; these conditions are printed in boldface in (43). Of the 91 sentences listed in (43), 71 comply with the Complexity Constraint. This leaves 20 counterexamples which are all of the same type: NP-related clauses belonging to the object in clauses with OS word order. While we do not know why all counterexamples come from OS sentences, we can at least conclude from these counterexamples that the Complexity Constraint does not hold categorically.

The existence of these counterexamples suggests that the pressure to extrapose—and thus to adhere to the Complexity Constraint—is counteracted by a ban on extraposition across a second argument. That is, extraposition is preferentially performed within a rather local window (for experimental evidence on this point, cf. Bader, 1996; Uszkoreit et al., 1998). As shown by the two examples where extraposition actually occurred across an intervening second NP (one SO example with extraposition from the subject and one OS example with extraposition from the object), this is not an absolute ban. On the other hand, these two examples are clearly outweighed by the 20 examples of non-extraposed NP-related clauses in middlefield-initial position. We thus conclude that the ban on extraposition across an argument has significantly more weight than the pressure to obey the Complexity Constraint.

An easy way to overcome the ban on extraposition across a second argument would be to switch word order. In particular, one might use a sentence with OS order for the very purpose of bringing a complex subject NP—in particular a subject NP containing an NP-related clause—into a position which makes extraposition easy. Evidence for this would be provided if OS sentences with a subject-related extraposed clause would exhibit a distribution of animacy that is untypical for OS sentences. An inspection of the data showed that this is clearly not the case. Of the 37 OS-sentences with a subject-related extraposed clause, 95% had an animate object followed by an inanimate subject. For the 18 OS-sentences with an object-related clause, this rate was 83%. These values are either equal or even somewhat higher than the base-rate of 82% sentences with inanimate subject and animate object among all OS sentences. Our data thus do not provide evidence for the hypothesis that OS order is used in order to bring an NP-related clause into a position optimal for extraposition.

We next consider prefield sentences. Since prefield sentences are main clauses, a complication arises with regard to classifying NP-related clauses as extraposed or not. As illustrated in (44), the verb cluster signaling the end of the

middlefield (the right bracket) is only filled when a sentence contains at least two verbal items.

- (44) Peter

| |
|-----|
| las |
| hat |

 ein Buch

| |
|-------------|
| \emptyset |
| gelesen |

, das der Onkel empfohlen hatte.
 Peter read/has a book read that the uncle recommended had
 ‘Peter read/has read a book that the uncle had recommended.’

As (44) shows, when the last phrase of a sentence is modified by a clause, the position of this clause is only decidable if the right bracket is not empty. Otherwise, the position of the clause is undecidable. In the following analyses, we treat all sentences with an NP-related clause that is ambiguous as to its position because of an empty verb cluster as a separate category.

The table in (45) shows the distribution of NP-related clauses in prefield sentences. In this table, the two categories ‘NP internal’ and ‘Extrapolated’ include only instances which could be unambiguously assigned to either of these categories. The last line represents NP-related clauses which are ambiguous between being NP internal (i.e., within the middlefield) and extrapolated.

(45)

| | SO Sentences | | OS Sentences | |
|--------------|--------------|-----------|--------------|---------------|
| | S-related | O-related | S-related | O-related |
| NP internal | 12 | <i>1</i> | 6 (0) | 34 (0) |
| Extrapolated | 0 | 8 | 14 (3) | 0 (0) |
| Ambiguous | – | 8 | 18 (1) | – |

A first striking feature seen in (45) is the low incidence of NP-related clauses unambiguously located within the middlefield, that is, within middlefields ending in one or more non-finite verbs (printed in italics in (45)). There is only a single example among 29 SO sentences (about 4%) and 6 examples out of 76 OS sentences (ca. 8%). It thus seems that when NP-related clauses can be removed from the middlefield both to the left (via fronting to the prefield) and to the right (via extrapolation), putting an NP-related clause into a closed middlefield is largely avoided. A second finding revealed by the table in (45) is that the two options of evacuating the middlefield—fronting to the prefield and extrapolation—occur with about equal frequency, at least if extrapolated NP-related clauses and ambiguous NP-related clauses are summed together.

As above, counterexamples to the Complexity Constraint are printed in boldface. In contrast to what we have seen for middlefield sentences, prefield sentences show a much higher percentage of counterexamples. However, these are of a different nature than the counterexamples found for middlefield sentences. Here, a more complex NP preceding a less complex NP is located in the prefield and not in the middlefield. 46 instances out of 105 total instances with NP-related clause are of this type, and thus almost 50%. It thus seems fair to conclude that the Complexity Constraint is not valid for sentences in which either subject or object is located within the prefield. We will come back to this issue when we discuss Hawkins’s weight-based theory in section 4.4.4. First, we consider the effect of weight defined in terms of constituent length.

4.4.2. Weight and Constituent Length

The Length Constraint is repeated in (46).

- (46) *Length Constraint*
 short > long

In the following, we will stick to the simplest measure of length, namely number of words. If constituent length is operative independent of structural complexity as defined in the preceding subsection, then it should show up even if we restrict our analysis to subjects and objects which do not contain NP-related clauses. We consequently removed all sentences containing NP-related clauses before computing the results that are presented in this subsection.

Table 6 shows the mean length of the first and the second NP in both SO and OS clauses. In addition, this table shows the length difference between the second and the first NP. Since an analysis of variance with the factors order and case would involve greatly varying sample sizes, the 95% confidence interval for the length difference is shown instead, separately for each combination of order and case. A length difference is significant on the 5% level if this

Table 6: Length of subject and object by order, case, and position

| | Length of first NP | Length of second NP | Difference (NP2 - NP1) | 95% Confidence Interval |
|---------------------|--------------------|---------------------|------------------------|-------------------------|
| Middlefield corpora | | | | |
| SO[Acc] (n = 524) | 2.72 | 3.67 | 0.95 | 0.69 – 1.21 |
| OS[Acc] (n = 52) | 3.40 | 3.96 | 0.56 | -0.40 – 1.50 |
| SO[Dat] (n = 64) | 2.98 | 3.55 | 0.56 | -0.15 – 1.27 |
| OS[Dat] (n = 725) | 3.25 | 3.19 | -0.06 | -0.30 – 0.18 |
| Prefield corpora | | | | |
| SO[Acc] (n = 213) | 3.93 | 3.66 | -0.28 | -0.92 – 0.36 |
| OS[Acc] (n = 372) | 4.78 | 3.87 | -0.91 | -1.45 – -0.38 |
| SO[Dat] (n = 19) | 3.05 | 4.21 | -1.16 | -2.84 – 0.52 |
| OS[Dat] (n = 122) | 4.57 | 4.25 | -0.32 | -1.19 – 0.55 |

Table 7: Number of sentences with subject shorter than object, subject and object of equal length, and subject longer than object, broken down by order, case and sentence type.

| | Middlefield | | | Prefield | | |
|---------|-------------|-----|------------|------------|-----|-----------|
| | S<O | S=O | S>O | S<O | S=O | S>O |
| SO[Acc] | 266 | 149 | 109 | 91 | 50 | 72 |
| SO[Dat] | 31 | 14 | 19 | 4 | 5 | 10 |
| OS[Acc] | 13 | 12 | 27 | 198 | 61 | 113 |
| OS[Dat] | 260 | 209 | 256 | 52 | 22 | 48 |

interval does not contain zero (these intervals are printed in boldface in Table 6). If the Length Constraint holds, the length difference should in fact always be significantly greater than zero because positive differences indicate that the second NP is longer than the first NP.

For sentences with both subject and object in the middlefield, only SO sentences with an accusative object show a significant length difference. OS sentences with an accusative object and SO sentences with a dative object still show a positive length difference, which, however, is rather modest and not significant. While the lack of significance for these sentences might be argued to be due to the small number of sentences, this is surely not the case for OS sentences with a dative object. For OS dative sentences, there is virtually no length difference at all.

Sentences with either subject or object in the prefield consistently exhibit a negative difference between second and first NP, meaning that on average the NP in the prefield is longer than the following middlefield internal NP. However, this difference is only significant for accusative OS sentences.

In summary, the Length Constraint is not supported by the data shown thus far. In order to strengthen this conclusion, we provide a further evaluation of the Length Constraint in Table 7. For the purpose of this table, all sentences were classified according to whether the subject was shorter than the object (S<O), whether the two were of equal length (S=O), or whether the subject was longer than the object (S>O). Numbers in cells which contradict the Length Constraint are printed in boldface.

Table 7 strengthens the conclusion that was already reached above on the basis of Table 6. Beginning with middlefield sentences, note first that in SO sentences with an accusative object, the subject is significantly more often shorter than the object than the other way around. When taken as a decision rule stating that SO should be used unless the subject is longer than the object, the Length Constraint would correctly classify 78% of all SO sentences. For OS sentences, the Length Constraint fares substantially worse. Applying the same decision rule to OS sentences leads to a correct classification of only 36%.¹⁵ We therefore conclude that constituent length is not a good predictor of the order between subject and object in the middlefield. A similar conclusion was already reached by Kurz (2000), who

¹⁵Following Hawkins (1994), one might suspect that SO order, as the canonical order, is retained even if the subject is longer than the object up to n words. Setting n to one word gives a revised decision rule according to which OS order is used if the subject is longer than the object by two words or more. This revised decision rule increased the success rate to 87% for SO sentences but reduced the success rate for OS sentences to 22%. It is thus of no more help than the original decision rule.

Table 8: NP length in number of words broken down by animacy, for subject NPs and object NPs. Numbers of instances are given in parentheses. Significant differences between animate and inanimate NPs are indicated by italics.

| | Subject NPs | | Object NPs | |
|-------------------|-------------|------------|------------|------------|
| | Animate | Inanimate | Animate | Inanimate |
| Middlefield SO/OS | 2.42 (417) | 3.35 (278) | 3.21 (217) | 3.75 (478) |
| Middlefield OS | 2.95 (38) | 3.25 (739) | 3.01 (661) | 4.66 (116) |
| Prefield SO/OS | 3.65 (167) | 4.10 (117) | 3.28 (75) | 3.75 (209) |
| Prefield OS | 3.68 (332) | 4.55 (162) | 3.99 (136) | 5.01 (358) |

Table 9: NP length in number of words broken down by definiteness, for subject NPs and object NPs (always definite). Numbers of instances are given in parentheses. Significant differences between animate and inanimate NPs are indicated by italics.

| | Subject NPs | | Object NPs |
|-------------------|-------------|------------|------------|
| | Definite | Indefinite | Definite |
| Middlefield SO/OS | 2.73 (553) | 2.96 (179) | 3.53 (732) |
| Middlefield OS | 3.46 (329) | 3.03 (496) | 3.23 (825) |
| Prefield SO/OS | 3.82 (228) | 3.75 (83) | 3.55 (311) |
| Prefield OS | 3.85 (421) | 3.95 (118) | 4.61 (539) |

also investigated the order between subject and object, and Fanselow (2000), who looked at the order between NP objects and PP objects.

For prefield sentences, the situation for the Length Constraint is even worse. There is only one condition for which the number of sentences in which the subject is longer differs significantly from the number of sentences in which the object is longer, namely the condition OS sentences with an accusative object. However, echoing the finding from Table 6, the difference is in the opposite direction from what would be expected under the Length Constraint. That is, when an accusative object is located in the prefield, it is most often longer than the subject, despite preceding it.

4.4.3. On the Relationship between Weight, Animacy and Definiteness

As already discussed in section 2, a recurrent theme with regard to word order variation is the question of possible correlations between the factors that have been proposed to determine word order. In particular, is there a single factor to which all other factors can be reduced? According to Hawkins (1994), this single factor is weight. We discuss Hawkins's theory in some detail in the next section. At this point, we only consider the more specific claim that animate NPs tend to precede inanimate NPs because the former are usually shorter than the latter (Hawkins, 1994, 337ff.), and definite NPs tend to precede indefinite NPs for the same reason (Hawkins, 1994, 111ff.). In the analyses presented in the preceding sections, weight did not turn out to be predictive for the order between subject and object whereas animacy and definiteness had an effect on order. This already casts serious doubts on the purported primacy of weight, and on the more specific claim that animacy and definiteness effects are epiphenomena of weight. To corroborate this point, this section presents the data that are necessary to assess the hypothesized relationship between animacy, definiteness and weight.

Table 8 shows the effect of animacy on length (measured in numbers of words) for both subject and object NPs in all of our four corpora. In all conditions shown in Table 8, animate NPs are shorter than inanimate NPs. This difference is significant in five of eight conditions. Overall, the data in Table 8 can be taken as support for Hawkins's hypothesis that animate NPs are on average shorter than inanimate NPs (for further evidence and discussion, see Rosenbach, 2005). Why then did animacy but not weight turn out to be a strong determinant of word order, at least for word order in the middlefield? The main reason for this lies in the fact that there are also certain length differences between subjects and objects. In the middlefield, subjects always show a somewhat smaller mean length than objects. In the middlefield OS corpus, in particular, the two categories that make up the main bulk of subjects and objects are of very similar length: animate objects with a mean length of 3.01 words and inanimate subjects with a mean length of 3.25 words. Although this still conforms to the general tendency of animate NPs being shorter than inanimate NPs, the difference is quite small and therefore not suitable to predict the order among subject and object in any useful way.

The effect of definiteness on length is shown in Table 9. Note that only subject NPs differ with regard to definiteness whereas object NPs are all definite due to the restriction of our corpora to objects introduced by the definite article *den*. In contrast to the factor animacy, the factor definiteness has no well-defined impact on length. First of all, only one corpus—the middlefield OS corpus—shows a significant difference between definite and indefinite subjects, but this difference is against the predicted direction with definite subjects in fact being longer than indefinite subjects. The differences exhibited by the other three corpora are quite small, non-significant, and not even in the same direction. When we also take objects, which are always definite, into account, the resulting picture gets even more diverging. In the middlefield SO/OS and the prefield OS corpora, definite objects are longer than definite subjects whereas the reverse is true in the middlefield OS and the prefield SO/OS corpora. We therefore conclude that weight and definiteness are not correlated in the corpora under consideration.

4.4.4. Weight and Order: Discussion

With regard to the two weight hierarchies listed in Siewierska (1993), the overall conclusion can only be a negative one. Although we have encountered several effects of constituent weight, the constraints covered by the weight hierarchies are not obeyed in any useful way. Despite this negative assessment of the weight hierarchies, the results presented in this section nevertheless indicate that weight plays some role for word order phenomena in German, although not in the way encoded in the weight hierarchies. In the remainder of this section we will show that a better understanding of some of the current results can be achieved by considering the parsing-based theory of word order developed by John Hawkins (for book-length presentations of this work, cf. Hawkins, 1994, 2004). For reasons of space, our discussion will be informal and short; this, however, does not affect the conclusions that we draw from our results.

According to Hawkins, parsing considerations are the main determinant of word order. In particular, orders are favored which allow the parser to structure the incoming input string as quickly as possible. This idea is captured by the principle of *Early Immediate Constituent (EIC)* which states that the parser prefers word orders which allow the recognition of a phrase and its immediate constituents as quickly as possible. Importantly, EIC does not imply a general short-before-long principle. Instead, according to EIC, the effect of constituent weight must be relativized to the position of the head licensing the phrases which are to be ordered. For purposes of illustration, let us consider the relative order between NP and PP. The situation obtaining in a strictly head-initial language like English is illustrated in (47) (the relevant heads are printed in boldface).

(47) Head-initial language

- a. [VP **send** [NP **books** which contain many pictures] [PP **to** Peter]]
 1 2 3 4 5 6 7 8
 └──┘
- b. [VP **send** [PP **to** Peter] [NP **books** which contain many pictures]]
 1 2 3 4 5 6 7 8
 └──┘

For the following considerations, it is assumed that a phrase is recognized on encountering its head. For (47a), where the NP precedes the PP, it takes seven words to recognize VP and its two immediate constituents NP and PP; note that the NP complement of P is not counted because the head P is sufficient for the recognition of PP. When the NP is put behind the PP, as in (47b), four words suffice to identify VP, NP and PP. Thus, in a head-initial language the preferred order is short before long.

For a strictly head-final language, the opposite is true. This is illustrated in (48) (for the sake of simplicity, we stick to English wording).

(48) Head-final language

- a. [VP [NP many pictures contain which **books**] [PP Peter **to**] **send**]
 1 2 3 4 5 6 7 8
 └──┘
- b. [VP [PP Peter **to**] [many pictures contain which **books**] **send**]
 1 2 3 4 5 6 7 8
 └──┘

Here, it takes only four words to recognize the VP in (48a) with the long NP preceding the short PP but seven words

supporting the predictions stemming from the left-to-right application of the EIC. However, neither our study nor the already cited studies by Fanselow (2000) and Kurz (2000) could confirm the claim that weight is a crucial determinant of middlefield-internal word order. The currently available evidence thus strongly suggests that word order within the middlefield cannot be explained by the kind of parsing considerations that form the core of Hawkins’s theory. We must leave it as a question for future research whether the recent advances of his theory (Hawkins, 2004) can handle these data.

4.5. Putting Everything Together: Logistic Regression Analysis

With a few exceptions, the analyses so far have investigated each constraint on word order in isolation from all other constraints. In order to capture the different constraints simultaneously, we performed a logistic regression analysis. Of the five constraints considered above, all were taken into account for the regression analysis with the exception of the Complexity Constraint. The remaining four constraints were recoded as factors in the following way before being entered into the analysis: (i) The Animacy Constraint was coded as the two factors *Subject Animacy* (animate vs. inanimate) and *Object Animacy* (animate vs. inanimate). (ii) The Semantic Role Hierarchy was again tested indirectly by the two factors *Voice* (active vs. passive) and *Perfective Auxiliary* (*haben* vs. *sein*). Note that for sentences in the passive voice the factor perfective auxiliary was coded as *sein* because the passive auxiliary forms its perfect with *sein*. (iii) The Definiteness Constraint was captured by the factor *Subject Definiteness* (definite vs. indefinite). (iv) For the Length Constraint finally we set up the factor *Length Difference* (length of subject minus length of object, measured in number of words).

The results of the logistic regression analysis with these six factors are shown in (50) for middlefield sentences and in (51) for prefield sentences.

(50) Results of logistic regression analysis for middlefield SO/OS corpus

| Coefficients: | Estimate | Std. Error | z value | Pr(> z) |
|----------------------|----------|------------|---------|-----------|
| (Intercept) | -3.7804 | 0.4803 | -7.87 | 3.5e-15 |
| Subject Animacy | 2.5644 | 0.4907 | 5.23 | 1.7e-07 |
| Object Animacy | -2.6546 | 0.4173 | -6.36 | 2.0e-10 |
| Voice | 1.0646 | 0.5298 | 2.01 | 0.045 |
| Perf. Auxiliary | 2.0855 | 0.4182 | 4.99 | 6.1e-07 |
| Subject Definiteness | 1.6509 | 0.3942 | 4.19 | 2.8e-05 |
| Length Difference | 0.0114 | 0.0591 | 0.19 | 0.847 |

(51) Results of logistic regression analysis for prefield SO/OS corpus

| Coefficients: | Estimate | Std. Error | z value | Pr(> z) |
|----------------------|----------|------------|---------|-----------|
| (Intercept) | -1.9196 | 0.3925 | -4.89 | 1e-06 |
| Subject Animacy | 0.6903 | 0.3370 | 2.05 | 0.041 |
| Object Animacy | -0.0184 | 0.3698 | -0.05 | 0.960 |
| Voice | 0.6484 | 0.9224 | 0.70 | 0.482 |
| Perf. Auxiliary | 0.7029 | 0.7249 | 0.97 | 0.332 |
| Subject Definiteness | 0.0583 | 0.3608 | 0.16 | 0.872 |
| Length Difference | -0.0810 | 0.0433 | -1.87 | 0.061 |

The results summarized in (50) and (51) largely confirm the discussion of the individual constraints. For word order in the middlefield, all factors were significant with the exception of the length difference between subject and object. This strengthens the picture that had emerged before: Word-order in the middlefield is mainly a matter of lexical-semantic factors—related both to the verb and to the verb’s arguments—with discourse-based notions like definiteness playing a role as well.

For word order involving the prefield, in contrast, only the factor Subject Animacy was fully significant. In addition, the factor Length Difference was marginally significant. This again shows that word order involving the prefield is not the same as word order involving the middlefield. The significance of the factor Subject Animacy shows that lexical-semantic factors also have an effect on prefield sentences. Given the strong SO preference for both middlefield and prefield sentences, we hypothesize that diverging from SO order in the middlefield is mainly (although

not completely) a matter of lexical conceptual factors whereas diverging from SO order by putting the object into the prefield is conditioned by different factors, with length playing an additional role here. We present a test of this hypothesis in the next section.

5. Modeling Word Order Variation

In this section, we discuss how the corpus results presented in the preceding two sections relate to grammatical conceptions of word order variation. Since a thorough discussion of the numerous approaches to word order is beyond the scope of the current paper, we will concentrate on what in our opinion is a particularly promising line of research.

5.1. Linguistic Accounts of Word Order Variation

A syntactic analysis of word order variation in German has to answer two major questions. The first one concerns the formal means for generating sentences with SO and OS order. In principle, there are two ways of generating a sentence with a particular order between subject and object. First, the sentence can be directly generated with subject and object in the correct order. Second, the sentence can be generated by applying one or more transformations to an underlying order which diverges from the observable surface order. Deciding between these two ways of handling word-order variation has a long tradition in the generative literature on German syntax. For example, already Thiersch (1978) distinguished between the *Fixed Order Hypothesis*, according to which a reordering transformation applies to a single underlying base structure, and the *Random Order Hypothesis*, according to which all logically possible orders among arguments can be base-generated, with filters selecting the ones that are actually grammatical.

After the transformational component had been reduced to a single movement operation ‘Move α ’ (for summary discussions, see Radford, 1981; Newmeyer, 1986), transformational accounts of word order variation have usually assumed that under certain circumstances the base-generated order between subject and object (typically, but not necessarily, SO) can be changed by moving the second argument to the left of the first one. When considering word-order variation within the middlefield, the term *scrambling* has established itself for this particular movement operation. This term was coined by Ross (1967) in his discussion of word-order variation in Latin. The scrambling transformation, a so-called ‘stylistic rule’, switches the linear order between two constituents. A detailed discussion of the various uses of this term with regard to German word order can be found in Haider and Rosengren (1998).

If one makes the assumption that certain word orders are transformationally derived, one has to specify the nature of the underlying representations. For example, are all possible orders among arguments derived from a single underlying base structure, or does word order variation already exist at the level of underlying representations? Considering only word order within the middlefield for the moment, we can discern at least three positions on this question.

First, on the basis of data concerning the binding of anaphoric elements, Müller (1995) claims that the arguments of a verb are base-generated in the fixed order ‘subject > accusative object > dative object’. The prototypical surface order with the dative object preceding the accusative object is derived from the underlying base structure by moving the dative object in front of the accusative object; other deviations from the fixed base structure are derived in a similar way. An implementation of this proposal in terms of Optimality Theory can be found in Müller (1999). For dative NPs which are not an argument of the verb (so-called ‘free datives’), Müller (1995) makes a different claim; such NPs are base-generated in a position higher than the accusative object.

According to an alternative position, word-order variation is found both on the level of underlying and on the level of derived structures. Such a position has been advocated by Haider (1993) (see also Haider and Rosengren, 2003). According to Haider (1993), each verb determines a particular base order between its arguments. For example, 2-argument verbs typically establish the order ‘subject before object’ as base order, but a small subset of verbs—including certain psych-verbs and unaccusative verbs (see section 4.2)—have OS order as base order. Deviations from a verb-specific base order are again derived by means of scrambling, that is, by moving an argument in front of some other argument(s). Heck (2000) also assumes that word-order variation is already found at the level of underlying structures, with additional variation brought about by scrambling. A crucial difference between Haider (1993) and Heck (2000) is that according to Heck (2000) the base-generated order between arguments is not determined by the verb, but instead by an optimality-theoretic competition between a set of word-order constraints like those that were discussed in the preceding section, including in particular the constraint ‘animate before inanimate’.

Third, mixtures of the first and the second position have also been proposed. For example, Vogel and Steinbach (1995) assume that the subject is always base-generated in front of the accusative object whereas a dative object can be freely inserted at any base-structure position.

If one takes the assumption that word-order variation is already found at the level of underlying, base-generated structures to its extreme, one ends up with the hypothesis that all possible orders among subject and object are base-generated (the Random Order Hypothesis of Thiersch, 1978). Even in syntactic frameworks which—like the Minimalist Program (e.g., Chomsky, 1995)—otherwise make heavy use of syntactic movement, analyses of this kind can be found (e.g., Fanselow, 2001a; Neeleman and Reinhart, 1998).

We now turn to the second major question that has to be addressed by an analysis of word-order variation, namely the question concerning the reasons for using one order instead of the other(s). In particular, how does one decide whether to use a sentence with SO order or a sentence with OS order? It seems uncontroversial that this question can only be answered by recourse to the notion of *information packaging*—the way sentences are divided up into topic and comment and into focus and background (cf. Erteschik-Shir, 2006a, for a recent overview). Among the formal means for signaling information packaging, syntactic means (word order) and prosodic means (sentence stress) are relevant in the current context. It is helpful at this point to take a closer look at recent cross-linguistic investigations of the relationship between syntax, prosody, and information packaging, as initiated in work by Vallduvi and Engdahl (1996) and Zubizarreta (1998) and developed further in the framework of Optimality Theory by, e.g., Büring and Gutiérrez-Bravo (2001) and Samek-Lodovici (2005). This line of research has identified two main strategies by which a language can resolve conflicts among syntactic and prosodic requirements: Either a rigid syntactic structure is combined with a flexible prosody, or a rigid prosody is combined with a flexible syntax.

English is a prototypical language of the first type; Italian instantiates the second one. For both languages, canonical word order is SVO and main sentence stress is canonically assigned in rightmost position. In many cases, canonical word order and canonical sentence stress coincide. This is true, for example, in out-of-the-blue utterances, as shown in (52).

(52) *Sentence-focus* (Context: What happened?):

- a. English: [_{Focus} John has LAUGHED]
- b. Italian: [_{Focus} Gianni ha RISO]
- G. has laughed

The two languages diverge, however, when it comes to narrow focus on the subject. As shown in (53a), English leaves the subject in its canonical preverbal position in this case. Since a focused constituent must contain the main stress of the sentence, this means that main stress can no longer be assigned in its canonical right-most position. In Italian, the opposite is found. In the Italian example (53b), main stress is still in rightmost position but word order is no longer canonical because the subject is now in clause-final position.

(53) *Subject-focus* (Context: Who has laughed?):

- a. English: [_{Focus} JOHN] has laughed
- b. Italian: Ha riso [_{Focus} GIANNI]
- has laughed G.

In terms of Optimality Theory, the two language types exemplified by English and Italian differ in the relative weight given to syntactic and prosodic constraints. Syntactic constraints specify the canonical word order of a language. For both English and Italian, these constraints specify SV(O) as canonical. Prosodic constraints specify the canonical position of sentence stress, which is clause-final again for both English and Italian. Where the two languages differ is with regard to the ranking of the constraints. In English, syntactic constraints are higher ranked than prosodic constraints. When a subject is narrowly focused, the sentence therefore still shows up with SV(O) order, and focus is assigned non-canonically to the sentence initial constituent. In Italian, in contrast, the prosodic constraint(s) which require clause-final sentence stress outrank the syntactic constraint(s) which require SV(O) order, leading to a constraint hierarchy in which prosodic and syntactic constraints are interleaved (cf. Samek-Lodovici, 2005). A narrowly-focused subject is therefore located sentence-finally, resulting in a sentence with non-canonical word order.

In contrast to English and Italian, both strategies coexist in the grammar of German (cf. Büring and Gutiérrez-Bravo, 2001). German sentences have the canonical order S(O)V and canonical sentence stress on the rightmost

argument (for qualifications of these characterizations, see the cited literature and below). When a sentence has wide focus, as the sentence in (54), German behaves like English and Italian. This sentence has canonical word order and canonical sentence stress on the rightmost argument.

- (54) *Sentence-focus* (Context: What happened?)
 [_{Focus} Der Pfarrer hat den Opa angerufen.]
 the minister has the grandpa called
 ‘The minister called grandpa.’

When the subject is in focus, however, German provides two alternative ways to resolve the conflict between syntax and prosody: Either the canonical subject-object (SO) order is maintained and sentence stress is assigned to the initial argument—as in English—or sentence stress is assigned to the final argument by switching to object-subject (OS) order—as in Italian. This is illustrated in (55).

- (55) Wer hat den Opa angerufen? (‘Who called grandpa?’)
 a. [_{Focus} Der PFARRER] hat den Opa angerufen. SO
 the minister has the grandpa called
 ‘The minister called grandpa.’
 b. Den Opa hat [_{Focus} der PFARRER] angerufen. OS
 The grandpa has the minister called
 (same meaning as (55a))

The kind of analysis sketched above assumes that under conditions of sentence-wide focus, a sentence shows up with the canonical order between subject and object as well as the canonical position of sentence stress. Under conditions of narrow focus, non-canonical orders and positions of sentence stress might be observed. An important question thus becomes what properties a sentence must have in order to allow sentence-wide focus. For English with its rigid syntactic structure, the answer to this question is straightforward; sentences with SVO order allow sentence-wide focus. For German, answering this question is somewhat more involved. In German, sentence-wide focus is prototypically associated with SOV word order—this is one of the reasons for considering SOV as the canonical word order of German—but this is not always so. This has been established in the large body of work—initiated by the seminal studies of Lenerz (1977) and Höhle (1982)—exploring the factors that determine whether a sentence with a given argument ordering allows sentence wide-focus (so-called *unmarked* word order) or only narrow focus on one of its constituents (so-called *marked* word order). Note also that SO or OS sentences with unmarked word order not only allow sentence wide focus but also narrow focus on the both the subject and the object. With regard to their focus potential, sentences with unmarked order thus subsume sentences with marked order.

As already pointed out at several points above, two main candidates for determining the unmarked word order of a sentence have been proposed. According to one position (e.g., Lenerz, 1977, Haider, 1993, Haider and Rosengren, 1998), the unmarked word order is determined by the verb of a sentence, as discussed in section 4.2 in connection with the Semantic Role Hierarchy. The alternative position, found for example in Müller (1999) and Heck (2000), is that the unmarked word order of a sentence is mainly a function of the position of subject and object on the Personal Hierarchy, as captured in a simplified way by the Animacy Constraint that was tested in section 4.1. Since there may be a tie between subject and object with regard to their position on the Personal Hierarchy, a limited amount of information provided by the verb must also be taken into account under this second position, for example about whether an argument is an agent (see the Agent Constraint of Heck, 2000) or about the case of an argument (see the Nominative Constraint of Müller, 1999).

Overall then, the theoretical developments summarized in this section lead to the following approach to word-order variation in German. Each sentence has an unmarked word order which is determined by lexical-semantic properties of the verb and/or lexical-semantic properties of the verb’s arguments. Deviating from the unmarked order is allowed under special information-structural conditions, for example in order to bring a narrowly focus-marked constituent into the default focus position which is assumed to be immediately in front of the clause-final verb. With regard to the reasons for using sentences with OS order, this has the following consequences. On the one hand, OS order can be unmarked in being licensed directly by lexical-semantic reasons. A sentence for which this is true allows for broad, sentence-wide focus, but also for narrow focus on either subject or object. It is thus usable in a wide variety of

Table 10: Observed word order × predicted word order, for the SO/OS middlefield corpus and the OS middlefield corpus.

| | Observed | Word order predicted by | | | |
|--------------|-----------|-------------------------|------------------|------------------|------------------|
| | | Animacy Rule | | Verb Rule | |
| | | SO | OS | SO | OS |
| SO/OS corpus | <i>SO</i> | 583 (79.7%) | 38 (5.2%) | 580 (79.2%) | 41 (5.6%) |
| | <i>OS</i> | 16 (2.2%) | 95 (13.0%) | 10 (1.4%) | 101 (13.8%) |
| OS corpus | <i>OS</i> | 146 (17.7%) | 679 (82.3%) | 57 (6.9%) | 768 (93.1%) |

contexts. On the other hand, OS order can be marked in being only licensed by certain information-structural reasons. The literature cited above has concentrated on the preferred position of a focused constituent in front of the verb as such an information-structural reason, but properties related to topichood are clearly relevant too in this respect.

In summary, when considering how this kind of approach to word order variation relates to the corpus results presented in the present paper, two questions must be addressed. First, are the instances of OS order which can actually be observed more a matter of discourse licensing or more a matter of lexical-semantic licensing? Second, what are the properties that define lexical-semantic licensing, verb properties or argument properties? We discuss these two questions in the remainder of this section.

5.2. Corpus Frequency and Models of Word Order Variation

We begin to address the questions raised at the end of the last subsection by determining how much of the observed word-order variation can be accounted for in terms of lexical-semantic properties. As discussed above, two hypotheses can be found in the pertinent syntactic literature with regard to the lexical-semantic properties responsible for whether or not a given order between subject and object is unmarked. We have translated these hypotheses into the *Verb Rule* in (56) and the *Animacy Rule* in (57).

(56) *The Verb Rule*

The subject precedes the object unless either of the following holds:

- The verb is in the passive voice.
- The verb takes *sein* as auxiliary.
- The verb belongs to a selected class of OS-licensing verbs (the complete lists are given in the columns headed ‘Active Dative *haben* verbs’ and ‘Accusative’ in the table shown on page 45 in the appendix).

(57) *The Animacy Rule*

The subject precedes the object unless the subject is lower on the personal (animacy) hierarchy than the object (that is, unless the subject is inanimate and the object is animate).

According to the *Verb Rule*, each verb has certain lexical-semantic properties which determine an unmarked order between its arguments. As in section 4.1, we identify the subset of verbs with unmarked OS order indirectly via voice and auxiliary selection in the perfect. In addition, some individual verbs which license OS order but are not captured by the aforementioned criteria are also included. According to the *Animacy Rule*, word order follows from lexical-semantic properties of subject and object itself. The Animacy Rule amounts to the claim that the object is put in front of the subject only if the object is higher on the Personal Hierarchy—which subsumes the Animacy Constraint as a special case—than the subject. Thus, the Animacy Rule is formulated on the assumption that SO is the default word order between subject and object, and that the default order can be overcome under special conditions. Note that the animacy rule has a predecessor in the word order template (7) of Hoberg (1981).

The results of applying these rules to sentences with both subject and object in the middlefield are shown in Table 10; counterexamples to the two rules under consideration are printed in boldface. This table shows how many SO sentences are correctly predicted to have SO order and how many are falsely predicted to have OS order, both by the Animacy Rule and by the Verb rule, and the same for sentences with OS order. For example, in the SO/OS corpus, there are 583 SO sentences for which the Animacy Rule predicts SO order and 38 SO sentences which are erroneously predicted to occur with OS order.

Table 11: Observed word order \times predicted word order, for the SO/OS prefield corpus and the OS prefield corpus.

| | Observed | Word order predicted by | | | |
|--------------|-----------|-------------------------|-------------------|--------------------|------------------|
| | | Animacy Rule | | Verb Rule | |
| | | SO | OS | SO | OS |
| SO/OS corpus | <i>SO</i> | 221 (71.1%) | 33 (10.6%) | 236 (75.8%) | 18 (5.8%) |
| | <i>OS</i> | 37 (11.9%) | 20 (6.4%) | 41 (11.6%) | 16 (6.8%) |
| OS corpus | <i>OS</i> | 431 (80%) | 108 (20%) | 427 (79.2%) | 112 (20.8%) |

For the SO/OS corpus, the Animacy Rule correctly predicts the word order of 92.6% of all sentences. At 93.0%, the percentage of correct predictions is practically the same for the Verb Rule. These numbers are higher than the baseline of 84.8% SO usage, although the differences are not statistically significant. Furthermore, a comparison of the Animacy and the Verb Rule reveals that in 92.2% of all cases, the two rules make the same prediction. For the OS corpus, the success rate of the Verb Rule is 93.1% and thus higher than the 82.3% success rate of the Animacy Rule. However, since the particular verbs that went into the Verb Rule were taken from the OS corpus, not much should be made of this difference. Furthermore, for the SO/OS corpus the Verb Rule also has a somewhat higher success rate for OS sentences, which however is counteracted by a slightly decreased success rate for SO. At 80.0%, the overlap between Animacy and Verb Rule is still quite high in the OS corpus. Given the high degree of overlap between the two rules, and the fact that both have virtually the same success rate when applied to the SO/OS corpus, it is impossible to decide between them based on the data in Table 10.

We next take a closer look at the counterexamples found in the OS corpus. For reasons of space, and given the strong correlation between the Animacy and the Verb Rule, we only consider counterexamples to the Animacy Rule. The most striking feature is that in 107 out of 146 counterexamples (73%) both the subject and the object are inanimate. These counterexamples will be considered in more detail in the next section. The remaining 39 counterexamples are too few to allow further statistical investigation, so we restrict ourselves to some cursory observations.

First, an example as in (58) argues that a personal hierarchy as fine-grained as the one given in (5) might indeed be needed.

- (58) ... dass den Touristen und Artenschutz-Aktivisten die Säugetiere auf dem Silbertablett vor Augen
that the tourists and animal-activists the mammals on the silverplate before eyes
geführt werden, ...
presented are
‘...that the mammals are presented to the tourists and animal activists for free.’

Sentence (58) contains two animate NPs, but the object NP is human and the subject NP non-human. Non-human NPs are lower on the personal hierarchy than human NPs, and in fact, the subject follows the object in (58).

Second, there are several middlefield sentences for which the observed OS order contradicts both the Personal Hierarchy and the Thematic Role Hierarchy. An example is shown in (59).

- (59) ... dass den letzten Suppentag eine so grosse Gästeschar besuchte.
that the last soup-day a so big guest-group visited
‘...that so many guests visited the last public soup day.’

In (59), an inanimate patient object precedes an animate agent subject. Sentences of this type are prototypical instances of ‘scrambling’ sentences in the sense discussed above, that is, sentences in which the object is put in front of the subject for purely information-structural reasons. While our data confirm earlier claims that such sentences are a rare event, examples as in (59) prove that ‘scrambling’ is a real phenomenon which can be observed when the analyzed corpus is sufficiently large.

Table 11 shows the results of applying the Animacy Rule and the Verb Rule to sentences in which either subject or object is located in the prefield. As before, order predictions by the Animacy Rule and order predictions by the Verb Rule coincide to a high degree. For the SO/OS corpus, the overlap is 90.0%; for the OS corpus, it is 92.6%. The

results in Table 11 differ markedly from the corresponding results for middlefield sentences. For the SO/OS corpus, the overall success rate is 77.5% for the Animacy Rule and 81.0% for the Verb Rule. These rates, which are even slightly less than the baseline rate of 81.7% SO sentences, are simply a side-effect of the high overall proportion of SO sentences. This becomes most evident when considering the corpus with OS sentences only. Here, the success rate for the Animacy Rule is at a low 20% which is almost the same as the 20.8% for the Verb Rule. Table 11 thus allows only one conclusion: The use of OS order with the object located in the prefield is not the result of lexical-conceptual factors.

5.3. Animacy versus Verb Semantics: A Final Test

We now undertake a final attempt at deciding between the two rules by taking a closer look at sentences which are especially problematic for them. For the Verb Rule, these are sentences which differ in order despite containing identical verbs. The Verb Rule is based on the assumption that each verb is associated with either SO or OS as the unmarked word order. This does not exclude the possibility that a verb may occur in a sentence with the opposite order. However, it should do so only under special discourse conditions. What should not be observed if the assumption underlying the Verb Rule is correct, is an animacy effect on word order for a verb which occurs in both orders.

In order to test this prediction, we consider verbs that occur both with SO and OS order in the middlefield and examine whether animacy determines the order between subject and object for those verbs. It was not possible to test this for individual verbs because no verb occurred with sufficient frequency with both orders. Since animacy of subject and object correlates with verb type, aggregating across all verbs in general might lead to spurious effects. We therefore restrict the analysis to those verbs which are predicted to occur with OS order by the Verb Rule in (56). When looking at the set of SO sentences from the middlefield SO/OS corpus and all sentences from the middlefield OS corpus, we found 18 verbs occurring both with SO and OS order holding voice constant. Six of these verbs occurred with a frequency of two or greater with both SO and OS order: *fallen* ('to fall'), *gehen* ('to go'), *kommen* ('to come'), *sein* ('to be'), *treffen* ('to hit'), *werden* ('to become'). The distribution of animacy properties in SO sentences (n=30) and OS sentences (n=242) containing one of those verbs is shown in (60).

| (60) | S[+an]/O[+an] | S[+an]/O[-an] | S[-an]/O[+an] | S[-an]/O[-an] |
|------|---------------|---------------|---------------|---------------|
| SO | 10.0% | 13.3% | 56.7% | 20.0% |
| OS | 4.1% | 0.4% | 81.0% | 14.5% |

For both SO and OS sentences, sentences of the type S[-an]/O[+an] occur with the highest frequency. For SO sentences, such an animacy pattern is quite unusual. Since we restricted our analysis to the subset of verbs which typically occur with OS order, and since such verbs typically have an inanimate subject and an animate object, this finding points to the importance of verb-specific word-order preferences. On the other hand, the peak at the combination S[-an]/O[+an] is much sharper for OS sentences than for SO sentences (81.0% vs. 56.7%; $\chi^2 = 7.92, p < .01$). This finding suggests that the choice between SO and OS order is influenced by animacy even when the analysis is restricted to SO and OS sentences containing the same set of verbs.

We next consider the Animacy Rule in (57) in more detail. As pointed out above, 107 out of 146 counterexamples to the Animacy Rule in the middlefield OS corpus were sentences with both subject and object inanimate. Since the formulation of the Animacy Rule in (57) implicitly states that SO is the default word order, it predicts that such sentences should occur with SO order. We formulated the Animacy Rule in this way because in the SO/OS corpus, of 149 sentences with both subject and object inanimate, 139 occur with SO order and only 10 with OS order.

Making SO the default order when subject and object tie with respect to animacy amounts to a constraint of the sort 'NOM > ACC/DAT'. Alternatively, properties related to the lexical semantics of the verb might decide about order in such cases. To test for this possibility, we look at the kind of verbs that occur in sentences with both subject and object inanimate. The relevant numbers for the 139 SO and the 107 OS sentences of this kind are shown in (61).

| (61) | Active: <i>haben</i> -verbs | Active: <i>sein</i> -Verb | Passive: <i>haben</i> -verbs |
|------|-----------------------------|---------------------------|------------------------------|
| SO | 94.6% | 4.0% | 1.3% |
| OS | 21.5% | 38.3% | 40.2% |

(61) reveals that SO and OS sentences differ strongly with regard to the distribution of verbs. While a vast majority of SO sentences contains a *haben*-verb in the active voice, for OS sentences active sentences with *sein*-verb and passive

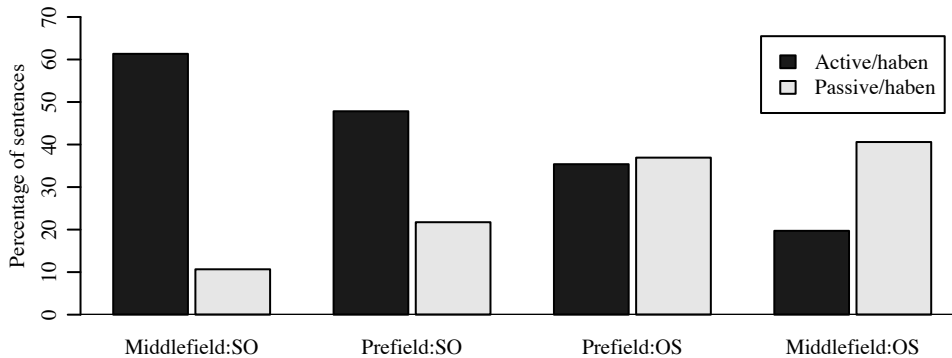


Figure 6: Percentages of sentences with active/*haben* and passive/*haben* verbs; data are taken from Figure 5.

sentences predominate. For sentences in which both subject and object are inanimate, the choice between SO and OS order thus clearly depends on verb-related factors (as far as these are reflected in auxiliary choice and voice). Note that invoking a constraint to the effect that the causer of an event (the agent in case of volitional causation) precedes all other arguments—as proposed, for example, by Siewierska (1993) and Heck (2000)—would not be sufficient to account for the pattern in (61). Such a constraint might explain the very high proportion of *haben* verbs in sentences with SO order, but it would not explain why agent-less verbs, that is, verbs in the other two verb classes, occur mainly in sentences with OS order, for the simple reason that the constraint under discussion would not apply to these verbs.

Summing up, the evidence presented in this section is somewhat equivocal. We tentatively conclude that the basic order between subject and object is jointly determined by lexical-semantic properties of verbs and animacy properties of their arguments. The evidence provided here does not indicate that either one can be dispensed with in favor of the other. More work is clearly needed to either confirm or refute this conclusion, and, in the former case, to specify in detail how verb and argument properties work together in determining word order.

5.4. Discourse-Licensed OS Order

A final point to consider is the role played by discourse factors for word order either within the middlefield alone or involving both prefield and middlefield. According to the results presented above, middlefield-internal order largely reflects the lexical-semantic properties of verbs and their arguments. For the minority of sentences diverging from the resulting order, we hypothesize that this is due to information-structural reasons. For prefield sentences, we hypothesize that the same lexical-semantic factors are at work but that arguments are reordered for discourse-based reasons much more frequently than in the middlefield, thereby masking the effect of the lexical-semantic factors to a substantial extent. We thus get the prediction that for sentences with either subject or object in the prefield the effect of lexical-semantic properties should be weaker than for sentences with subject and object within the middlefield.

This prediction is supported by the distribution of verbal properties (auxiliary selection and voice) that was investigated in section 4.2. Figure 6 (based on Figure 5; see there for further explanation) provides an alternative view of the percentages of active and passive *haben*-verbs for all four combinations of order (SO vs. OS) and sentence type (middlefield vs. prefield).

Seen from left to right, Figure 6 shows a steady decrease of the proportions of active *haben*-verbs and a steady increase of the proportion of passive verbs. Importantly, for both verb types the left and the right endpoints are marked by middlefield sentences: SO middlefield sentences have the highest proportion of active *haben*-verbs and the lowest proportion of passive *haben*-verbs, and the reverse is true for OS middlefield sentences. Prefield sentences are located in between, with prefield SO sentences nearer to middlefield SO sentences and prefield OS sentences nearer to middlefield OS sentences. Thus, prefield sentences still show the same distributional pattern of verb usage as middlefield sentences, but in a less extreme form.

We thus arrive at the hypothesis that both lexical-semantic and discourse-related factors have a substantial impact on word order when the prefield is involved. Note that this hypothesis is compatible with the conclusions reached by Weber and Müller (2004) in their study of the order between subject and object in main-clauses with either subject

or object in the prefield. Weber and Müller investigated the influence of three discourse-related factors—givenness, definiteness, and pronominalization—and found that neither of them is a strong determinant of word order. Weber and Müller (2004) therefore conclude that the discourse-related factors investigated in their study must be complemented by other factors, including information about animacy.

This leaves us with the task of specifying the particular discourse-related properties that determine the order between subject and object. Following previous work (Fanselow, 2001b; Frey, 2004b), we assume the existence of (at least) two ways to fill the prefield. First, under discourse-neutral conditions, the middlefield-initial constituent is moved to the prefield. In this case, prefield sentences preserve middlefield-internal word order and are therefore subject to the same linearization constraints. Second, a non-initial constituent can be moved to the prefield for discourse reasons. In this case, prefield sentences exhibit a word-order diverging from the order determined by lexical-semantic factors, whether this is the order SO or the order OS. Putting a phrase into the prefield is also known under the name of *topicalization*. While this term is inadequate when taken in an absolute sense—as already shown by (55), the prefield can clearly not only house topics—this term suggests that the prefield is typically occupied by a constituent which is a topic.¹⁷ While a special topic position has also been identified at the left edge of the middlefield (cf. Frey, 2000, 2004a), the main information-structural impact on middlefield-internal word order seems to be the desire to have the exponent of focus, that is, the constituent bearing main sentence stress, at the end of the middlefield and thus in a position immediately in front of the verb. Note that this can be achieved in different ways. First, a clause in which subject and object appear in their base-order fulfills this requirement without further adjustments if the clause is focused as a whole or if the argument which comes second according to the lexical-semantic factors is focused. Only if the first argument is focused would it be necessary to reorder the two arguments.

For practical reasons, it was not possible to test this hypothesis in a quantitative way. We therefore present only some representative examples from a cursory survey of our corpora with regard to the information-structural properties associated with OS sentences. In the prefield corpora, the majority of objects in the prefield are topics, mostly ordinary topics but sometimes also contrastive topics. An example illustrating the use of an ordinary topic is given in (62). The topic constituent in the prefield is printed in bold face.

- (62) Ob das am Montag vergangener Woche eröffnete Verfahren gegen den 55jährigen ehemaligen Mathematikprofessor wegen mehrfachen Mordes und vielfacher Körperverletzung durch Briefbomben fortgeführt wird, steht zur Zeit in den Sternen. **Den Prozeß** hat Kaczynski mit zwei Aktionen aus der Spur gebracht.

‘It is completely unclear at the moment if the lawsuit against the 55 years old former professor of mathematics for multiple murder and physical injury by means of letter bombs – it was opened on Monday last week – can continue. Kaczynski has disrupted **the trial** by two actions.’

(Frankfurter Rundschau, 14.01.1998)

In addition to ordinary topics, we also found some contrastive topics as in (63).

- (63) Beide Gremien hatten über zwei konkurrierende Anträge zu entscheiden. **Den ersten** haben die Grünen im Oktober 1997 gestellt.

‘Both committees had to decide between two competing proposals. **The first one** was proposed by the Green Party in October 1997.’

(Frankfurter Rundschau, 06.05.1998)

The situation is different in the middlefield OS corpus. Most often, the object is either directly given or inferable—which is expected given that all objects are definite—but it is not a topic. A representative example is given in (64).

- (64) Tausende von Urlaubern verließen nach einer Warnung des Nationalen Hurrikan-Zentrums in Miami (Bundesstaat Florida) die beliebten Urlaubsorte.

Die Meteorologen gingen davon aus, dass **den Küstenbewohnern** das Schlimmste erspart bleibt und der Hur-

¹⁷The notion of topic is not uncontroversial in the literature. At least two conceptions of topichood can be distinguished. One family of approaches considers aboutness as the crucial property of topics whereas the other family of approaches gives priority to givenness. While both properties often coincide, this is not necessarily so. Givenness is not a precondition for aboutness and neither does aboutness follow from givenness. In the following we adopt the first view and define topics as what the sentence is about.

rikan nicht auf das Festland stoßen, sondern nach Nordosten abdrehen wird.

‘Thousands of visitors left the popular destinations after a warning of the national hurricane center in Miami (Florida). The meteorologists assumed that **the coastal inhabitants** will be spared of the worst, and that the hurricane will not strike the main-land but will turn away to the northeast.’
(Frankfurter Rundschau, 31.08.1999)

The topic of the example in (64) is a hurricane threatening the coast. The clause-initial object *den Küstenbewohnern* (‘the coastal inhabitants’) is inferable in the given context but not the topic of its clause.

The identification of embedded topics is in fact rather difficult. Frey (2004a) offers a whole range of diagnostics, but most of them cannot be applied to our corpora since they involve specific lexical items (e.g., certain adverbials like *(un)glücklicherweise* ‘(un)fortunately’) which either occurred very rarely if at all or are excluded due to the restriction to definite objects and the exclusion of pronouns. One of the few examples in which the clause-initial object can clearly be identified as a topic is given in (65). In this passage, the topic is the German volleyball team. The NP *den jungen Spielerinnen* (‘the young players’) is a lexical variant referring to the topic.

(65) Unerfahrene deutsche Volleyballerinnen bei der Europameisterschaft Vierte/ Gegen Russland und Italien chancenlos

Was aber keinen Weltuntergang bedeutete, denn mit dem vierten Platz hatte die Mannschaft von Trainer Lee Hee Wan ihr Ziel sowieso schon übererfüllt. Und so fiel es nicht weiter ins Gewicht, dass **den jungen Spielerinnen** am Schluss ihre Grenzen klarer aufgezeigt wurden, als ihnen lieb sein konnte.

‘Unexperienced German volleyball players fourth at the European mastership/ Without chances against Russia and Italy.

However, this is no tragedy because with the fourth place the team of coach Lee Hee Wan had already achieved more than their original aim. And therefore it did not carry much weight that at the end **the young players** were shown their limits clearer than they would have liked.’

(Frankfurter Rundschau, 27.09.1999)

The examples in (64) and (65) have in common that subject and object appear in the order predicted by lexical-semantic considerations, independent of the additional information-structural properties of the sentence. This is different from the prefield examples in (62) and (63) in which the initial position of the object is only explainable with recourse to discourse properties. Given the results presented thus far, this difference between middlefield and prefield sentences is not unexpected. As a comparison between Table 10 and Table 11 shows, lexical-semantic factors account for most of the word-order variation when restricted to the middlefield but not when prefield and middlefield are both involved. One of the few middlefield examples in which animacy and verb semantics unambiguously predict SO order but the sentence nevertheless occurs with OS order is shown in (66) (see also example (59)).

(66) Guido Westerwelle, der umtriebige FDP-Generalsekretär, hat sein selbst gestecktes Ziel erreicht und während der Sommerpause den Segelschein gemacht. Nachdem sich der Polit-Profi, den auf der politischen Bühne kein Sturm erschüttern kann und der das immer wieder schlingende FDP-Schiff auf Kurs bringt, die Prüfung auf Anhieb bestanden hatte, legte er besonderen Wert darauf, dass **den Schein** nicht irgendjemand, sondern der für die Schifffahrt zuständige Bundesminister unterschreibt.

‘Guido Westerwelle, the secretary-general of the FDP who is always on the run, has achieved the aim he set himself: he obtained the sailing license during the sommer break. After the professional politician, who cannot be shaken by any storm on the political scene and who brings the always rolling FDP ship on course, had passed the exam at the first go, he insisted that not just anybody, but the federal minister responsible for shipping signs **the license**.’

(Mannheimer Morgen, 23.09.2000)

The crucial sentence in (66) has an animate subject which fills the agent role of the verb and an inanimate object. This is the prototypical configuration for SO order. That this sentence shows up with OS order instead can be attributed to the fact that the object is given in the preceding context and the subject is focused. This is thus a clear example of putting a discourse-given object in front of the subject in order to have the subject in the preferred preverbal focus

position.

In summary, the examples presented in this section give rise to the following hypotheses (cf. also Erteschik-Shir, 2006b). First, objects are put into the prefield when they are topics; second, objects are put in front of the subject within the middlefield in order to give way to a focused subject which thereby can end up in a preverbal position. In other words, within the middlefield objects are put in front of the subject for the subject's sake (see Jayaseelan, 2001, for a similar case from Malayalam), but they are put into the prefield for their own sake. Note that the latter hypothesis would explain one of the findings concerning the distribution of definite and indefinite subjects reported in section 4.3. As shown by (37) and Table 5, when either subject or object is in the prefield, the definiteness of the subject has no effect on the placement of the object which is always definite in our case. This finding follows from the hypothesis that objects are put into the prefield for their own sake: whether or not an object appears in the prefield only depends on discourse properties of the object itself, and the definiteness of the subject is therefore irrelevant.

5.5. Discussion

We summarize the findings discussed in this section in form of the two generalizations given in (67).

(67) *The OS Generalizations*

- a. Within the middlefield, an object occurs in front of the subject mainly in order to adhere to lexical-semantic constraints. In the rare case that an object precedes a subject for discourse-related reasons, this is due to the constraint requiring a focused constituent in preverbal position.
- b. For putting objects into the prefield, and thereby in front of the subject, lexical-semantic constraints and discourse-related constraints play both a significant role. The main discourse-related constraint is the constraint requiring topics to occur in clause-initial position.

According to the accounts of word order variation discussed above (e.g., Büring and Gutiérrez-Bravo, 2001; Samek-Lodovici, 2005), German is a language with a flexible syntax and a flexible prosody. In case of a mismatch between canonical syntactic structure and canonical prosodic structure, either structure can give way to the other. On a general level, our corpus analysis is in accordance with this description. However, it also reveals that the use that is made of the available flexibility is not evenly distributed across the different possibilities. In addition, our data also highlight certain pieces missing from the existing analysis. Most importantly, and in accordance with many linguistic studies of German word order, our data strongly support the assumption that the unmarked order between subject and object is—at least to a large degree—a function of lexical conceptual constraints instead of syntactic constraints. Furthermore, especially with regard to filling the prefield in main clauses, the constraints governing the position of focus must be complemented by constraints governing the position of topic. In summary, at least the three kinds of constraints on word order summarized in (68) must be distinguished.

- #### (68)
- a. Syntactic constraints
 - b. Lexical-semantic constraints in terms of verb semantics and animacy properties of the verb's arguments
 - c. Information-structural constraints in terms of focus and topic

For a language like English with its rigid syntactic structure, syntactic constraints must be ranked higher than both lexical-semantic constraints and information-structural constraints. In German, in contrast, the syntactic constraints must be ranked lowest, and lexical-semantic and information-structural constraints must tie in such a way that the available flexibility follows. As a final illustration of the different constraint rankings in English and German, consider the English sentence pair in (69) and its corresponding German translation in (70).

- #### (69)
- a. Hopefully, the book pleased Peter.
 - b. Hopefully, Peter was pleased by the book.
- #### (70)
- a. Hoffentlich gefiel Peter das Buch.
hopefully pleased P. the book
 - b. *Hoffentlich wurde Peter von dem Buch gefallen.
hopefully was P. by the book pleased

As shown in (69) and (70), the English verb *to please* and its German counterpart *gefallen* both realize the experiencer as object and the stimulus as subject. Due to the rigid word order of English, the stimulus has to precede the experiencer in a simple active clause like (69a). Since this is non-optimal in terms of the animacy features on subject and object, and perhaps also in terms of semantic roles, verbs like *to please* are among the verbs of English which occur frequently in the passive voice (cf. Bresnan et al., 2001; Dingare, 2001). As shown in (69b), reverting to passive voice results in an optimal order according to the lexical-semantic hierarchies. German, in contrast, allows the optimal order between experiencer and stimulus even in an active clause, as shown by (70a). In fact, verbs like *gefallen* do not passivize in German although verbs of this kind—dative verbs selecting *haben* as perfect auxiliary—normally do.

6. Summary

This paper has presented an in-depth corpus study of the order between subject and object in German. On the methodological side, a major innovation of the present corpus study in comparison to prior studies was that we sampled sentences in two ways: without any restrictions on order (SO/OS corpora) and with a restriction to object-initial order (OS corpora). We could therefore analyze sets of OS sentences which were much larger than the sets analyzed before. Furthermore, we went beyond prior studies in that we analyzed both sentences with subject and object in the middlefield and sentences where either subject or object was in the prefield.

The major findings of our study can be summarized as follows.

1. For both middlefield and prefield sentences, order and case interact, although not in the same way: When subject and object are both within the middlefield, SO sentences show a strong bias toward accusative objects whereas OS sentences have an even stronger bias toward dative objects. When either subject or object is located in the prefield, there is an accusative bias for both SO and OS sentences, but this bias is substantially stronger for SO sentences than for OS sentences.
2. For middlefield sentences, the use of OS order seems to be mainly conditioned by lexical-semantic factors, comprising both verb semantics and animacy properties of the verb's arguments. OS order with the object in the prefield, in contrast, does not seem to follow from lexical-semantic properties.
3. With respect to definiteness, we found mixed evidence. For sentences with an accusative object, SO was the favored order even if the subject was indefinite and the object definite, which should not occur according to the Definiteness Constraint. Sentences with a dative object, in contrast, showed the expected effect of definiteness: Sentences with definite object and indefinite subject occurred predominantly with OS order whereas sentences in which both subject and object were definite did not show any strong order preference.
4. With respect to weight, we also found mixed evidence. On the one hand, the positioning of NP-related clauses was clearly affected by weight insofar as NP-related clauses showed a strong tendency to occur outside the middlefield. On the other hand, weight in the form of constituent length (measured in number of words) was found not to be decisive for choosing between SO and OS order.

The present corpus study confirms and extends results that have been obtained in the earlier corpus studies by Hoberg (1981) and Kempen and Harbusch (2005). Because we not only analyzed sentence sets unrestricted with respect to order—for these, the rate of OS was rather low, as in the prior studies—but also sentence sets containing only OS sentences, the number of OS sentences that went into our analysis was much larger than in the earlier studies. This is in particular true for OS clauses in which the object is in the accusative case. In both Hoberg's study and in Kempen and Harbusch's study, the number of sentences of this kind was very small indeed (two and four, respectively). While the exceedingly low rate of OS sentences with an accusative object was also found in the present corpus study (< 1%), the absolute number was still much higher (n=54). The present study therefore allows the conclusion that accusative OS sentences follow a pattern which seems to hold for OS clauses in German in general. In particular—and in accordance with the grammatical literature on word order variation—middlefield sentences with an accusative object show OS order when the object is animate and the subject inanimate. This is exactly the same as what is found for sentences with a dative object. As already pointed out above, the fact that accusative OS sentences are much less frequent than dative OS sentences can therefore be seen as a side effect of the fact that there are simply many more opportunities for using dative OS sentences. For example, when passivized, almost all ditransitive verbs license sentences with OS order. In such sentences, the object can only bear dative case.

As a final point, we consider the finding presented in section 2 that sentences with three NP arguments—a subject, an accusative object, and a dative object—have a strong tendency to occur with SO order, in particular when all arguments are contained within the middlefield. For such sentences, the rate of SO was found to be at a high 95%. Given the importance of lexical-semantic factors on word-order within the middlefield, this finding comes at no surprise. The prototypical ditransitive verb has a subject which is an agent (an animate NP), a dative object which is a recipient (also an animate NP), and a patient (an inanimate NP). Furthermore, ditransitive verbs always select *haben* as perfective auxiliary. Thus, both the Verb Rule in (56) and the Animacy Rule in (57) predict that sentences with three NP arguments should occur with subject initial order, and this prediction is clearly confirmed by our corpus results.

References

- Allan, K., 1987. Hierarchies and the choice of left conjuncts. *Journal of Linguistics* 23, 51–77.
- Arnold, J. E., Ginstrom, R., Losongco, A., Wasow, T., 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76, 28–53.
- Bader, M., 1996. Prosodic effects and the distinction between primary and non-primary phrases. Poster presented at the Conference on Architectures and Mechanisms of Language Processing (AMLAP), Torino, University of Torino.
- Behagel, O., 1909/1910. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25, 110–142.
- Bierwisch, M., 1963. *Grammatik des deutsche Verbs*. Akademie Verlag, Berlin.
- Bornkessel, I., Schlesewsky, M., Friederici, A. D., 2002. Grammar overrides frequency: evidence from the online processing of flexible word order. *Cognition* 85, B21–B30.
- Bornkessel, I., Schlesewsky, M., Friederici, A. D., 2003. "And yet it moves" or why grammar overrides frequency: A reply to Kempen and Harbusch. *Cognition* 90, 211–213.
- Brants, S., Dipper, S., Hansen, S., Lezius, W., Smith, G., 2002. The TIGER Treebank. In: Hinrichs, E., Simov, K. (Eds.), *Workshop on Treebanks and Linguistic Theories (TLT)*. Sozopol, Bulgaria, pp. 24–41.
- Bresnan, J., Cueni, A., Nikitina, T., Baayen, R. H., 2004. Predicting the dative alternation. Manuscript, Rutgers Optimality Archive.
- Bresnan, J., Dingare, S., Manning, C. D., 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. In: Butt, M., King, T. H. (Eds.), *The Proceedings of the LFG '01 Conference*. URL <http://csli-publications.stanford.edu/LFG/6/lfg01.html>, University of Hong Kong.
- Büring, D., Gutiérrez-Bravo, R., 2001. Focus-related constituent order variation without the NSR: A prosody-based crosslinguistic analysis. In: Mac Bloscaidh, S. (Ed.), *Syntax at Santa Cruz 3*. Vol. 3. University of California, California, pp. 41–58.
- Chomsky, N., 1995. *The minimalist program*. MIT Press, Cambridge, MA.
- Dietrich, R., 1994. Wettbewerb - aber wie? Skizze einer Theorie der freien Wortstellung. In: Haftka, B. (Ed.), *Was determiniert die Wortstellungsvariation?* Westdeutscher Verlag, Opladen, pp. 33–47.
- Dingare, S., 2001. The effect of feature hierarchies on frequencies of passivization in English. M.A. thesis, Stanford University.
- Drach, E., 1937. *Grundgedanken der deutschen Satzlehre*, 4th Edition. Diesterweg, Frankfurt a. M.
- Eisenberg, P., 2004. *Grundriß der deutschen Grammatik*. Band 2: Der Satz, 2nd Edition. Metzler Verlag, Stuttgart.
- Engel, U., 1972. Regeln zur "Satzgliedfolge": Zur Stellung der Satzglieder im einfachen Verbalsatz. *Linguistische Studien* I:19, 17–75.
- Erteschik-Shir, N., 2006a. Information structure. The syntax-discourse interface. Oxford University Press, Oxford.
- Erteschik-Shir, N., 2006b. What's what? In: Fanselow, G., Féry, C., Vogel, R., Schlesewsky, M. (Eds.), *Gradience in grammar - generative perspectives*. Oxford University Press, New York, pp. 317–335.
- Fanselow, G., 2000. Does constituent length predict german word order in the middle field? In: Bayer, J., Römer, C. (Eds.), *Von der Philologie zur Grammatiktheorie*. Peter Suchsland zum 65. Geburtstag. Niemeyer, Tübingen, pp. 63–77.
- Fanselow, G., 2001a. Features, q-Roles, and free constituent order. *Linguistic Inquiry* 32, 405–437.
- Fanselow, G., 2001b. Münchhausen-style head movement and the analysis of verb second. In: Mahajan, A. (Ed.), *Proceedings of the Workshop on Head Movement*. Los Angeles: UCLA, Linguistics Department.
- Frey, W., 1993. Syntaktische Bedingungen für die semantische Interpretation. Über Bindung, implizite Argumente und Skopus. Akademie-Verlag, Berlin.
- Frey, W., 2000. Über die syntaktische Position der Satztopiks im Deutschen. *ZAS Papers in Linguistics* 20, 137–172.
- Frey, W., 2004a. A medial topic position for German. *Linguistische Berichte* 198, 153–190.
- Frey, W., 2004b. The grammar-pragmatics interface and the German prefield. *Sprache & Pragmatik* 52, 1–39.
- Grewendorf, G., 1989. *Ergativity in German*. Foris, Dordrecht.
- Haider, H., 1993. *Deutsche Syntax - generativ*. Vorstudien zu einer projektiven Theorie der Grammatik. Narr, Tübingen.
- Haider, H., Rosengren, I., 1998. *Scrambling*. Sprache und Pragmatik 49.
- Haider, H., Rosengren, I., 2003. Scrambling: Non-triggered chain formation in OV languages. *Journal of Germanic Linguistics* 15, 203–266.
- Hawkins, J. A., 1992. Syntactic weight versus information structure in word order variation. In: Jacobs, J. (Ed.), *Informationsstruktur und Grammatik*. Westdeutscher Verlag, Opladen, pp. 196–219.
- Hawkins, J. A., 1994. *A performance theory of order and constituency*. Cambridge University Press, Cambridge.
- Hawkins, J. A., 2004. *Efficiency and complexity in grammars*. Oxford University Press, Oxford.
- Heck, F., 2000. Tiefenoptimierung. *Linguistische Berichte* 184, 441–468.
- Heylen, K., 2005. A quantitative corpus study of German word order variation. In: Reis, M., Kepser, S. (Eds.), *Linguistic evidence*. de Gruyter, Berlin, pp. 241–263.

- Höhle, T. N., 1982. Explikation für "normale Betonung" und "normale Wortstellung". In: Abraham, W. (Ed.), *Satzglieder im Deutschen. Vorschläge zur syntaktischen, semantischen und pragmatischen Fundierung*. Narr, Tübingen, pp. 75–153.
- Höhle, T. N., 1986. Der Begriff 'Mittelfeld'. Anmerkungen über die Theorie der topologischen Felder. In: Schöne, A. (Ed.), *Kontroversen, alte und neue. Akten des VII. internationalen Germanisten-Kongresses, Göttingen 1985. Band 3*. Niemeyer, Tübingen, pp. 329–340.
- Hoberg, U., 1981. *Die Wortstellung in der geschriebenen deutschen Gegenwartssprache*. Hueber, München.
- Häussler, J., Bader, M., accepted. Grammar- versus frequency-driven syntactic ambiguity resolution: the case of double-object constructions, manuscript accepted for publication in: de Swart, Peter & Lamers, Monique (eds.) *Case, word order, and prominence. Psycholinguistic and theoretical approaches to argument structure*. Dordrecht: Springer.
- Jacobs, J., 1988. Probleme der freien Wortstellung im Deutschen. *Sprache und Pragmatik* 5, 8–37.
- Jayaseelan, K. A., 2001. IP-internal topic and focus phrases. *Studia Linguistica* 55, 39–75.
- Keenan, E. L., Comrie, B., 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8, 63–99.
- Kempen, G., Harbusch, K., 2002. An artificial opposition between grammaticality and frequency: comment on Bornkessel, Schlesewsky and Friederici (2002). *Cognition* 90, 205–210.
- Kempen, G., Harbusch, K., 2004. A corpus study into word order variation in German subordinate clauses: Animacy effects linearization independently of grammatical function assignment. In: Pechmann, T., Habel, C. (Eds.), *Multidisciplinary approaches to language production*. Mouton de Gruyter, Berlin, pp. 173–181.
- Kempen, G., Harbusch, K., 2005. The relationship between grammaticality ratings and corpus frequencies: A case study into word-order variability in the midfield of German clauses. In: Reis, M., Kepsen, S. (Eds.), *Linguistic evidence*. de Gruyter, Berlin, pp. 329–349.
- Konieczny, L., 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research* 29, 627–645.
- Kurz, D., 2000. A statistical account on word order variation in German. In: Abeillé, A., Brants, T., Uszkoreit, H. (Eds.), *Proceedings of the COLING Workshop on Linguistically Interpreted Corpora (LINC)*, Luxemburg.
- Lenerz, J., 1977. *Zur Abfolge nominaler Satzglieder im Deutschen*. Narr, Tübingen.
- Lenerz, J., 2001. Word order variation: Competition or co-operation? In: Müller, G., Sternefeld, W. (Eds.), *Competition in syntax. Vol. 49 of Studies in Generative Grammar*. Mouton de Gruyter, Berlin & New York, pp. 249–281.
- Lerot, J., 1985. Zur Wortstellungsnorm im Deutschen. *Deutsche Sprache* 13 (2), 137–142.
- Levin, B., Rappaport Hovav, M., 2005. *Argument realization*. Cambridge University Press, Cambridge.
- Lötscher, A., 1981. Abfolgeregeln für Ergänzungen im Mittelfeld. *Deutsche Sprache* 9 (1), 44–60.
- Mitchell, D. C., 1994. Sentence parsing. In: Gernsbacher, M. A. (Ed.), *Handbook of psycholinguistics*. Academic Press, San Diego etc., pp. 375–409.
- Müller, G., 1995. *A-bar syntax*. Mouton de Gruyter, Berlin.
- Müller, G., 1999. Optimality, markedness, and word order in German. *Linguistics* 37, 777–818.
- Neeleman, A., Reinhart, T., 1998. Scrambling and the PF interface. In: Butt, M., Geuder, W. (Eds.), *The projection of arguments: Lexical and compositional factors*. CSLI Publications, Stanford, CA, pp. 309–353.
- Newmeyer, F. J., 1986. *Linguistic theory in America*, 2nd Edition. Acedem Press, New York.
- Nunberg, G., Sag, I. A., Wasow, T., 1994. Idioms. *Language* 70, 491–538.
- Pafel, J., 2006. *Quantifier scope in German*. John Benjamins, Amsterdam.
- Primus, B., 1994. *Grammatik und Performanz: Faktoren der Wortstellungsvariation im Mittelfeld*. *Sprache und Pragmatik* 32, 39–86.
- Radford, A., 1981. *Transformational syntax*. Cambridge University Press, Cambridge.
- Reis, M., 1986. *Die Stellung der Verbalargumente im Deutschen: Stilübungen zum Grammatik:Pragmatik-Verhältnis*. *Lunder germanistische Forschungen* 55, 139–177.
- Rosenbach, A., 2005. Animacy versus weight as determinants of grammatical variation in english. *Language* 81, 613–644.
- Ross, J. R., 1967. *Constraints on variables in syntax*. Ph.D. thesis, MIT.
- Samek-Lodovici, V., 2005. Prosody-syntax interaction in the expression of focus. *Natural Language & Linguistic Theory* 23, 687 – 755.
- Scheepers, C., 1996. *Menschliche Satzverarbeitung: Syntaktische und thematische Aspekte der Wortstellung im Deutschen*. Doctoral dissertation, Universität Freiburg.
- Siewierska, A., 1993. On the interplay of factors in the determination of word order. In: Jacobs, J., von Stechow, A., Sternefeld, W., Vennemann, T. (Eds.), *Syntax: An international handbook of contemporary research*. Walter de Gruyter, Berlin, pp. 826–846.
- Skut, W., Krenn, B., Brants, T., Uszkoreit, H., 2002. An annotation scheme for free word order languagess. In: *Fifth Conference on Applied Natural Language Processing*. Washington, D.C., USA, pp. 27–28.
- Tanenhaus, M. K., Trueswell, J. C., 1995. Sentence comprehension. In: Miller, J., Eimas, P. D. (Eds.), *Speech, language, and communication. Handbook of perception and cognition*, 2nd. ed., vol. 11. Academic Press, San Diego, CA, pp. 217–262.
- Thiersch, C., 1978. *Topics in German syntax*. Doctoral dissertation, MIT.
- Uhmann, S., 1993. *Das mittelfeld im gespräch*. In: Reis, M. (Ed.), *Wortstellung und Informationsstruktur*. Niemeyer, Tübingen, pp. 314–354.
- Uszkoreit, H., 1986. Constraints on word order. *Linguistics* 24, 883–906.
- Uszkoreit, H., 1987. *Word order and constituent structure in German*. Chicago University Press, Chicago.
- Uszkoreit, H., Brants, T., Duchier, D., Krenn, B., Konieczny, L., Oepen, S., Skut, W., 1998. *Studien zur performanzorientierten Linguistik Aspekte der Relativsatzextraposition im Deutschen*. *Kognitionswissenschaft* 7, 129–133.
- Vallduvi, E., Engdahl, E., 1996. The linguistic realization of information packaging. *Linguistics* 34, 469–519.
- Vogel, R., Steinbach, M., 1995. On the (absence of a) base position for dative objects in German. *FAS Papers in Linguistics* 4, 99–131.
- von Stechow, A., Sternefeld, W., 1988. *Bausteine syntaktischen Wissens*. Westdeutscher Verlag, Opladen.
- Wasow, T., 2002. *Postverbal behavior*. CSLI Publications, Stanford.
- Weber, A., Müller, K., 2004. Word order variation in German main clauses: A corpus analysis. In: *Proceedings of the 20th International Conference on Computational Linguistics*, Geneva, Switzerland. pp. 71–77.
- Zubizarreta, M. L., 1998. *Prosody, focus, and word order*. MIT Press, Cambridge, MA.

Appendix

The table below shows the twenty most frequent verbs for each of the four categories, active *haben*-verbs with a dative object, active *sein*-verbs with a dative object, passive verbs with a dative object and accusative verbs. Some of these verbs occur in several more or less fixed combinations with certain objects or prepositional objects (so-called Funktionsverbgefüge and idiomatic expressions). Note that the translations given for each verb only cover the core meaning of the corresponding verb, whereas in the corpus the verbs occur with different meanings, often depending on the accompanying noun phrase or prepositional phrase. Take for instance the verb *stehen* (core meaning: ‘to stand’): The overwhelming majority of instances of *stehen* occurs in combination with *zur Verfügung* (‘at disposal’) meaning ‘to be at hand’. A few instances occur with *zur Seite* (‘at side’), meaning ‘to support’, *ins Haus* (‘into the house’), meaning ‘to be imminent’, *im Weg* (‘in the way’), meaning ‘to stand in sb.’s way’, and *zur Auswahl* (‘at choice’), meaning ‘to be up for choice’. Finally, we found a couple of instances of *stehen* in rather fixed idiomatic expressions: *Haare zu Berge stehen* (‘hair standing on end’, literally ‘hair stand at hill’) and *Wasser bis zum Hals stehen* (‘to be to one’s neck in it’, literally ‘water standing up to one’s neck’).

| | Active Dative <i>haben</i> Verbs | | Active Dative <i>sein</i> Verbs | | Passive Dative | | Accusative | |
|----|-------------------------------------|----|------------------------------------|----|-------------------------------|----|--------------------------------|---|
| | verb | f | verb | f | verb | f | verb | f |
| 1 | stehen ‘to stand’ | 41 | sein ‘to be’ | 53 | bieten ‘to offer’ | 21 | treffen ‘to hit’ | 8 |
| 2 | fehlen ‘to lack’ | 31 | bleiben ‘to remain’ | 41 | geben ‘to give’ | 12 | befallen ‘to affect’ | 2 |
| 3 | liegen ‘to lie’ | 15 | gelingen ‘to succeed’ | 27 | schenken ‘to donate’ | 12 | beißen ‘to bite’ | 2 |
| 4 | zustehen ‘to be entitled to’ | 8 | folgen ‘to follow’ | 21 | stellen ‘to put’ | 11 | beschleichen ‘to stalk’ | 2 |
| 5 | gefallen ‘to please’ | 5 | ausgehen ‘to run out’ | 15 | anbieten ‘to offer’ | 10 | drücken ‘to press’ | 2 |
| 6 | stocken ‘to halt’ | 4 | zukommen ‘to belong to’ | 14 | einräumen ‘to admit’ | 10 | ereilen ‘to overtake’ | 2 |
| 7 | drohen ‘to threaten’ | 3 | einfallen ‘to occur to’ | 10 | entziehen ‘to withdraw’ | 10 | interessieren ‘to interest’ | 2 |
| 8 | gehören ‘to belong’ | 3 | werden ‘to become’ | 9 | machen ‘to make’ | 10 | plagen ‘to plague’ | 2 |
| 9 | winken ‘to wave’ | 3 | entstehen ‘to arise’ | 8 | tragen ‘to carry’ | 9 | | |
| 10 | anhaften ‘to cling to’ | 2 | erwachsen ‘to accrue’ | 8 | zugestehen ‘to allow’ | 9 | | |
| 11 | anmerken ‘to let on’ | 2 | fallen ‘to fall’ | 8 | gewähren ‘to allow’ | 8 | | |
| 12 | gebühren ‘to be due to’ | 2 | gegenüberstehen ‘to face’ | 7 | vorlegen ‘to present’ | 8 | | |
| 13 | kleben ‘to stick’ | 2 | gehen ‘to go’ | 7 | nehmen ‘to take’ | 6 | | |
| 14 | nützen ‘to use’ | 2 | bevorstehen ‘to impend’ | 5 | ausbezahlen ‘to pay off’ | 5 | | |
| 15 | reichen ‘to hand’ | 2 | vergehen ‘to decay’ | 5 | verweigern ‘to refuse’ | 5 | | |
| 16 | schmecken ‘to taste’ | 2 | kommen ‘to come’ | 4 | entgegenbringen ‘to bring’ | 4 | | |
| 17 | | | passieren ‘to happen’ | 4 | erteilen ‘to give’ | 4 | | |
| 18 | | | platzen ‘to burst’ | 4 | legen ‘to lay’ | 4 | | |
| 19 | | | verbleiben ‘to remain’ | 4 | setzen ‘to place’ | 4 | | |
| 20 | | | | | übertragen ‘to convey’ | 4 | | |