

Thematic asymmetries do matter!
A corpus study of German word order

Elisabeth Verhoeven (Humboldt-University Berlin)

Abstract

This article addresses the question whether the influence of thematic roles (in particular, experiencers and patients) on word order is an epiphenomenal effect of other factors (such as information structure and animacy). For this purpose, we are investigating argument realization with different verb classes, including canonical verbs and either agentive or non-agentive experiencer-object verbs with varying case marking (dative or accusative), in a large corpus of written German. The obtained results indicate that at least a part of the experiencer-first effect is triggered by further factors, in particular animacy. However, after subtracting the effect resulting from these factors, the impact of the thematic properties remains and is therefore necessary to explain the whole range of data.

Keywords

argument realization, psych-verbs, experiencer, word order, referential type, animacy, definiteness

1. Introduction

The choice of word order in speech production is influenced by a large array of factors which are partly correlated with each other (Bresnan et al. 2007). A class of phenomena influencing the choice of word order relates to information structure. Several studies demonstrate that given information is more likely to precede new information (Clark & Haviland 1977 and subsequent work). Another relevant factor relates to the inherent properties of referential expressions: animate referents are more likely to precede inanimate referents (Bock & Warren 1985, van Nice & Dietrich 2003, Prat Sala 1997, Branigan & Feleki 1999, Prat-Sala & Branigan 2000, Prat Sala et al. 2000, Bornkessel et al. 2005, Scheepers 1997, Scheepers et al. 2000). Beyond that, some linearization preferences are attributed to specific verbs. At least some of these preferences are not just idiosyncratic properties of individual lexical items but appear with verb classes sharing particular thematic relations. Experiencer-object verbs (often called psych verbs, e.g., *frighten*, *please*, *concern*) are among the most intensively investigated of such verb classes, showing linearization preferences that crucially differ from the properties of canonical transitive verbs (c.f. Ferreira 1994, Scheepers et al. 2000, Verhoeven 2009, Lamers & de Schepper 2010).

The crucial claim is that particular preferences in linearization are attributed to the relational properties of the arguments, i.e., experiencer-object verbs differ from canonical transitive verbs due to their thematic properties. The experiencer role is taken by a participant who undergoes an event that affects consciousness, i.e., an event of emotion, cognition, volition, perception, or bodily sensation. Since

consciousness is a prerequisite for being an experiencer, the experiencer is by definition animate. In contrast to an agent, an experiencer does not have control over the event. Next to the experiencer role many experiencer verbs license a stimulus argument; the stimulus is a rather heterogeneous role comprising the cause, target or subject matter of an experiential situation and can be either animate or inanimate (see Pesetsky 1995).

In German, experiencer-object verbs appear in two different case-marking subclasses, namely those taking a dative-marked experiencer argument such as *gefallen* ‘please’ and those with an accusative experiencer argument such as *interessieren* ‘concern’. Dative psych verbs are often analyzed as showing a basic dative-before-nominative order (e.g., Lenerz 1977, Primus 1996, Fanselow 2001, 2003, Haider & Rosengren 2003, Haider 2010). The word order properties of accusative psych verbs are not uncontroversial. Some accounts treat them on a par with dative psych verbs for their basic order (Lenerz 1977, Haider & Rosengren 2003, Bayer 2004), while other accounts consider them as canonical transitive verbs with a higher nominative argument (Fanselow 2000).

Knowing that several linearization principles, such as animate-first, given-first, and experiencer-first, depend on each other to some extent, the empirical question is whether a possible influence of thematic roles applies independently. This is the challenge of the present article, which presents a large-scale corpus study that was designed to detect the influence of verb classes on linearization. The article proceeds as follows: section 2 summarizes the crucial assumptions about the coding of argument structure and the linearization of arguments in German and outlines the positions that have been taken regarding experiencer verbs in the previous literature. The research targets are presented in detail in section 3, while the design of the corpus

study is explicated in section 4. The following sections report the findings of the corpus study: section 5 investigates the frequencies of pronominal realization of subject and object arguments of the investigated verbs; section 6 examines word order as depending on the animacy, thematic role and definiteness of the verbal arguments and section 7 observes the choice of active vs. non-active verb forms as depending on the same factors. The results of these parts are discussed in detail in section 8 while section 9 provides general conclusions and highlights the findings of the present study.

2. Prerequisites in the grammar of experiencers

2.1 The syntax of experiencer verbs

Experiencer-object verbs, also called psych verbs, constitute a special class of verbs that have been analyzed as showing a particular syntactic structure which differs from that of canonical transitive verbs.¹ Starting with the seminal paper by Belletti & Rizzi (1988), several analyses have been proposed to account for the peculiar syntactic properties of psych-verbs (see Grimshaw 1990, Bouchard 1995, Pesetsky 1995, Reinhart 2001, Landau, 2010 among others).

Three classes of experiencer verbs have generally been distinguished (Belletti & Rizzi 1988): one class of experiencer-subject verbs, e.g., *love, hate, fear* etc., which code the experiencer as subject and the stimulus as object, and two classes of experiencer-object verbs, coding the experiencer either as an accusative object (e.g., *frighten, annoy, concern*) or a dative object (e.g., *appeal to*) while the stimulus takes the subject function. In the sense of Bickel (2004), object-experiencers are so-called

¹ Canonical transitive verbs are agentive verbs taking an agent subject (external argument) and a patient direct object (internal argument).

downgraded experiencers, including experiencer arguments that are coded by structural means normally used for objects, e.g., dative, accusative case, and adpositional marking.

In many languages, among them German, it has been observed that experiencer-object arguments display subject-like properties, often called psych properties. These properties include peculiarities in nominalization, reflexivization, passivization, extraction, binding, and argument linearization, among others (see Bayer 2004, Fanselow 2000, Grewendorf 1989, Haspelmath 2001, Klein & Kutscher 2002, Wunderlich 1997). Crucial for the purposes of this article, experiencer-object verbs display particular linearization preferences: for instance, in German it has been shown that both orderings ($S_{STIM} \prec O_{EXP}$ and $O_{EXP} \prec S_{STIM}$) are equally acceptable for accusative experiencer-object verbs while there is a preference for an early realization of the experiencer, as opposed to the stimulus, with dative experiencer-object verbs (Haupt et al. 2008:84, confirming earlier observations from Lenerz 1977, Hoberg 1981, Primus 2004).

The subject-like behaviour of some object-experiencers can be accounted for by assuming that dative and (at least some) accusative experiencers take a higher position than the nominative stimulus in the derivational structure of the clause. Works such as Belletti and Rizzi (1988) propose an unaccusative analysis for both dative and accusative experiencer-object verbs, where the experiencer occupies a higher position than the stimulus in the VP. Since Belletti and Rizzi's (1988) original proposal, several adaptations to this analysis have been proposed, among these most importantly Pesetsky (1995), who identifies the potential causal nature of the accusative experiencer-object verbs and explains their argument linking by the assumption that the causer originates in a higher position than the experiencer. The

more recent analysis in Landau (2010) combines the insights of both approaches: in this analysis all stative experiencer-object verbs (dative and accusative) receive an unaccusative account while the eventive accusative experiencer-object verbs with causative/agentive readings come with a higher causer argument and a lower experiencer argument. Analyses of German experiencer-object verbs in general follow the unaccusative approach, either for the dative verbs only (Fanselow 2000, 2003, Wegener 1998) or extending it to the accusative verbs (Grewendorf 1989).

Crucial for the non-canonical properties is their interaction with the agentivity of the experiencer-object verbs. While experiencer verbs selecting a dative experiencer are necessarily non-agentive, accusative experiencer-object verbs can be either agentive or non-agentive: the verb is agentive if the stimulus has control over the event, and this configuration is only possible with animate stimuli, see example (1a) vs. (1b).

- (1) a. *Maria ärgerte Peter (absichtlich).*
 ‘Maria annoyed Peter (intentionally).’
- b. *Marias Fragen ärgerten Peter (*absichtlich).*
 ‘Maria’s questions annoyed Peter (*intentionally).’
- c. *Maria interessierte Peter (*absichtlich).*
 ‘Maria concerned Peter (*intentionally).’
- d. *Marias Fragen interessierten Peter (*absichtlich).*
 ‘Maria’s questions concerned Peter (*intentionally).’

The distinction between \pm agentive and strictly non-agentive experiencer-object verbs depends on the respective verb. While verbs such as *ärgern* ‘annoy’ and

überraschen ‘surprise’ both have agentive and non-agentive readings, depending on the context and the animacy of the stimulus, a few verbs such as *interessieren* ‘concern’, *freuen* ‘please’ or *wundern* ‘astonish’ are strictly non-agentive. With such verbs, an animate stimulus cannot be interpreted as agentive, i.e., as bringing about the verbal event; see (1c)/(1d).

Non-canonical syntactic behaviour of experiencer objects only applies to non-agentive readings of experiencer-object verbs, but not to the agentive variants (see Arad 1998a, 1998b, Landau 2010). Since agentivity is restricted to animates (as shown in (1)), an interaction with animacy is observed consistently in these phenomena.

2.2 *Diathesis relations*

One of the peculiarities of transitive experiencer-object verbs is their behaviour in passivization. In German, regular passives are prototypically dynamic and agentive (see Zifonun 1992). This property interacts with the agentivity of experiencer-object verbs in such a way that the formation of a regular eventive passive (by means of the passive auxiliary *werden* ‘become’) requires a \pm agentive experiencer-object verb in its agentive reading. Thus, a regular passive is only grammatical with an animate stimulus as in (2a) and not with an inanimate stimulus as in (2b). Non-agentive experiencer-object verbs such as *interessieren* ‘interest’ do not form such a passive at all.

- (2) a. *Peter ist/wird von Maria genervt.*
‘Peter is bothered by Maria.’
- b. *Peter ist/*wird von den Möbeln genervt.*
‘Peter is bothered by the furniture.’

c. *Peter ist an Maria interessiert.*

‘Peter is interested in Maria.’

Next to the regular passive, German has a stative adjectival passive formed with the copula and the passive participle, also illustrated in example (2). Stative passives do not require an agentive interpretation, and hence non-agentive experiencer-object verbs such as *interessieren* ‘interest’ may form a stative passive, in which the stimulus constituent is a prepositional adjunct (not a passive agent), as in (2c). Similarly, \pm agentive experiencer-object verbs may form a stative passive both with an animate or an inanimate stimulus, as in (2a)-(2b).

In many languages transitive (accusative) experiencer-object verbs are more or less systematically related to anticausative (= deagentive) experiencer-subject verbs. This also applies to German, evidenced by verb pairs such as *interessieren* ‘concern’ ~ *sich interessieren* ‘be concerned’, *freuen* ‘please’ ~ *sich freuen* ‘be pleased’, *ängstigen* ‘frighten’ ~ *sich ängstigen* ‘be frightened’, *wundern* ‘astonish’ ~ *sich wundern* ‘be astonished’, *ärgern* ‘bother’ ~ *sich ärgern* ‘be bothered’ etc. Anticausativization is marked by the reflexive marker *sich* in these cases.

For the present investigation, diathetic alternations are relevant in so far as a choice among alternative diathetic forms is possible in order to express the same situation. It has been shown in previous studies that the choice of voice is influenced by several factors, such as animacy, thematic role, salience or givenness (see e.g., Ferreira 1994, Tomlin 1995, Prat-Sala 1997, Skopeteas & Fanselow 2009, van Nice & Dietrich 2003 on German). The special property of experiencer-object verbs is that transitive experiencer-object verbs alternate with intransitive anticausative experiencer-subject verbs, e.g., ‘x worries y’ vs. ‘y worries about x’ (Belletti & Rizzi

1988, Pesetsky 1995, Reinhart 2001 etc.).² It is not clear whether the choice between these latter alternations depends on the factors that have been reported to influence the choice of voice in general (such an investigation will be undertaken in our corpus in section 7).³

2.3 *Source of experiencer-first effects*

The challenge in the investigation of experiencer-object verbs is to determine the source of the exceptional behaviour group of verbs. As a starting point, two extreme positions are formulated in (3). The choice between the syntactic view in (3a) and the functional view in (3b) is an empirical question, and it is very likely that the answer to this question does not uniformly hold for all subsets of experiencer-object verbs.

(3) a. Syntax-based hypothesis

The peculiarities of experiencer-object verbs in linearization are the result of their thematic structure.

² A systematic relation has also been claimed to exist between lexically converse pairs of experiencer-subject and experiencer-object verbs such as *fear* vs. *frighten* etc. In this respect Levin and Grafmiller (2013) convincingly show by means of a corpus study that the difference between these verbs is not exhaustively captured by their converse thematic role structure. Rather they differ in more fine-grained semantics, especially with respect to the nature of the stimulus argument. With such putative converse verb pairs, the choice between them cannot obviously be (fully) accounted for by purely abstract features such as animacy and topicality relations.

³ See Engelberg 2014 for frequency distributions of this alternation in German corpora. This study does not specifically investigate the influence of the abovementioned factors on the occurrence of the alternating forms; however, frequencies of intransitive experiencer-subject sentences and passive sentences suggest a similar discourse function.

b. Functional hypothesis

The peculiarities of experiencer-object verbs in linearization result from the impact of factors affecting linearization (such as animacy and topicality) that independently hold true.

The hypothesis in (3a) corresponds to the core claim of syntactic accounts of experiencer verbs (see section 2.1) that explain the exceptional properties of these verbs as a reflex of the syntactic derivation of experiencer objects. In the alternative view in (3b) at least some properties of experiencer verbs are explained by the functional properties of their arguments, i.e., their animacy properties (Bouchard 1995) and/or their discourse prominence and its correlates in topicality (Bickel 2004, Haspelmath 2001).

The relevance of the factor animacy in the linearization of experiencer and stimulus has been convincingly shown in diverse experimental studies. The rate of an early realization of the object-experiencer, either through object preposing or through passivization, increases significantly in asymmetric animacy constellations, i.e., when the stimulus is inanimate (see Ferreira 1994; Lamers & de Schepper 2010; Scheepers et al. 2000; Verhoeven 2009). However, as outlined in section 2.1, the factors (in)animacy and (in)agentivity interact with accusative object-experiencer verbs and need to be systematically separated, which has not been done in most studies, so that the contribution of each of them to the resulting pattern cannot be identified in those cases.

Similarly, the relational properties of the participants of psych verbs, i.e., their involvement in the event in terms of thematic roles, have been tested for their impact on syntactic construction. Experimental studies show that experiencers tend to be

realized early in the clause or in a higher syntactic function. Evidence for an early realization of the experiencer through passivization has been reported for English in production and comprehension studies (see Cupples 2002; Ferreira 1994; Piñango 2000). For languages such as German, Dutch and Modern Greek, experimental as well as corpus studies report prominence effects of object experiencers related to word order (e.g., Bader & Häussler 2010, Bornkessel 2002, Bornkessel-Schlesewsky & Schlewsky 2009, Hoberg 1981, Kempen & Harbusch 2004, Lamers 2007, Lamers & de Hoop *forthc.*, Primus 1994, Scheepers 1997, Scheepers et al. 2000, Verhoeven 2009). These effects are reminiscent of the animacy effects and indeed may be in part due to the fact that experiencers frequently outrank stimuli in the animacy hierarchy.

Finally, prominence effects of experiencers in terms of early occurrence and high syntactic function have been related to the so-called ‘natural topicality’ of the experiencer vis-à-vis the stimulus (Bickel 2004, Haspelmath 2001). However, this observation has been essentially made on intuitive grounds, and systematic empirical evidence for this claim is still pending (one exception being the small corpus study on dative experiencer verbs in Nepali in Ichihashi-Nakayama 1994).

2.4 German clause structure and word order

The finite verb in declarative main clauses in German obligatorily surfaces in the second position in the clause, which determines two clausal domains: (a) the middlefield, i.e., the domain following the finite verb and preceding the (potential) non-finite verb, and (b) the prefield (Spec,CP), namely the position preceding the finite verb, which has to be filled by exactly one constituent.

Argument order in the middlefield has been shown to be influenced by factors such as animacy, thematic role, definiteness, focus, case etc. (see Lenerz 1977, Hoberg 1981, Bader & Häussler 2010 and many more). Reordering within the

middlefield has been explained as resulting from scrambling (Webelhuth 1995, Müller 1999) or as reflecting alternative base generated orders (Haider 1993, Fanselow 2001, Heck 2000). Furthermore, argument order in the middlefield is constrained (independently of the abovementioned factors) when nominative (unstressed) personal pronouns are involved: these obligatorily occupy the position immediately following the finite verb (C^0) whenever they are realized in the middlefield.

The prefield is obligatorily filled, which induces formal movement of the first eligible element of the middlefield (see Frey 2006). The first eligible constituent is the highest one, i.e., the subject constituent or a constituent scrambled past the subject; since the operation that leads this constituent to the prefield is purely formal (i.e., semantically vacuous), this type of syntactic operation does not involve any additional semantic or pragmatic features than the ones that led to the scrambling of the highest constituent in the middlefield. Next to formal movement, two further possibilities to fill the prefield have been identified: base generation of certain adverbials and A-bar movement, the latter indicating a contrastive interpretation of the moved material.

The profound difference between argument order in the prefield and the middlefield has been empirically proven in a number of studies, most recently in Bader and Häusler (2010). This study shows that object occurrence in the prefield is mainly driven by information structural aspects while object preposing (early position of the object; object < subject) in the middlefield is due to lexical-semantic reasons (including both verb semantics and animacy of the verbal arguments). Furthermore, it is shown that object < subject occurrences are rare with accusative objects and more frequent with dative objects, when both arguments are lexical NPs. With accusative objects, object < subject is significantly more frequent in the prefield than in the

middlefield, while the difference between the clausal domains is not significant with dative objects (Bader & Häussler 2010:727).

In sum, argument order in the prefield vs. in the middlefield is governed by different rules and thus these domains of clause structure need to be considered separately in a study of argument order in German.

3. Research targets

The central question of the present study is whether the fronting of the experiencer arguments relates to their thematic properties or is reducible to further properties that are characteristic of these arguments. Experimental studies on experiencer objects in languages such as German, Dutch, English, or Greek show that the fronting of experiencer-objects is a genuine property of these arguments since it is also manifested in configurations that do not involve asymmetries in terms of animacy or information structure (see Ferreira 1994; Scheepers et al. 2000, Lamers & de Schepper 2010, Verhoeven 2009). Experiencer objects frequently outrank the corresponding subjects on the animacy hierarchy; since animates are more likely to be discourse topics, a proportion of fronted experiencer-objects is certainly fronted for discourse-related reasons. Thus, the question at issue is to what extent the phenomenon of experiencer-object fronting is explained by the mediation of further factors that have an impact on linearization such as animacy and information structure. The abovementioned studies tested the impact of animacy on linearization and subject choice with experiencer verbs. Naturalistic data provide another rich source of evidence for the observation of discourse-related factors and hence should be considered to answer our research question.

In order to understand the discourse-related factors that may induce experiencer-object fronting, we need to empirically test the assumption that experiencers are likely

discourse topics. The answer to this question is a prerequisite for any further consideration since the claim that experiencer-fronting effects are just an instance of a larger phenomenon relating to the preferred discourse status of this argument requires showing evidence that the assumption about the discourse status holds true. The next aim is to disentangle the sources of the experiencer-fronting effects: it has to be examined to what extent experiencer fronting is a phenomenon independent from animate-first or given-first effects. Finally, and assuming that experiencer-first effects exist, we should address the question whether these effects relate to a preference for the early realization of the experiencer role that may be either satisfied by a marked word order or by a diathetic alternation of the verb or not. The research questions of this study are summarized in (4).

(4) Research questions

- a. Are experiencer-objects likely discourse topics?
- b. Is there an experiencer-first effect and if yes, is it independent from further semantic and pragmatic asymmetries?
- c. Is the frequency of non-canonical orders with experiencer-object verbs part of a general preference for the earlier realization of experiencers?

There are different classes of experiencer-object verbs and there is no reason to expect that these classes will have a uniform behaviour with respect to the phenomena at issue. In particular, we will compare canonical transitive verbs with dative and accusative experiencer-object verbs. Within the latter case of verbs, we will examine non-agentive experiencer-object verbs and experiencer-object verbs that are not specified for agentivity (see section 2.1). There are several empirical situations that

may appear in regard to the phenomena in (4): (a) if the crucial feature for a phenomenon lies in case marking, then we expect a contrast between dative objects and accusative objects (experiencers and patients), see CASE MODEL in (5); (b) if the crucial feature relates to the undergoer argument, then we expect experiencer objects (of any case) to contrast with patients, see UNDERGOER MODEL in (5); (c) if the crucial feature relates to the thematic properties of the actor, i.e., the agent/stimulus, then we expect a contrast between the verbs that can be agentive vs. the verbs that exclude an agentive involvement of the actor, see ACTOR MODEL in (5). The models in (5) are postulated in an abstract manner that will be applied on different phenomena: the symbol ‘<’ stands for an asymmetry in the occurrence of an experiencer-relevant property.

(5) Basic VERB-CLASS MODELS

CASE MODEL:	accusative			< dative
	canonical	EO.ACC ±ag	EO.ACC –ag	< EO.DAT
ACTOR MODEL:	potentially agentive			< non-agentive
	canonical	EO.ACC ±ag	< EO.ACC –ag	EO.DAT
UNDERGOER MODEL:	patient	< experiencer		
	canonical	< EO.ACC ±ag	EO.ACC –ag	EO.DAT

The basic models in (5) account for the three possible splits between the four categories at issue. Complex empirical situations are also possible, as exemplified in (6), implying that more than one factor in (5) is at issue.

(6) Complex VERB-CLASS MODELS

CASE & ACTOR:	canonical EO.ACC ±ag < EO.ACC –ag < EO.DAT
---------------	--

CASE & UNDERGOER: canonical < EO.ACC ±ag | EO.ACC –ag < EO.DAT

ACTOR & UNDERGOER: canonical < EO.ACC ±ag < EO.ACC –ag | EO.DAT

CASE & ACTOR & UNDERGOER: canonical < EO.ACC ±ag < EO.ACC –ag < EO.DAT

In order to answer the questions in (4), we examined a large corpus of written German. The methodological decisions for the sampling procedure, the annotation of the data, and the data analysis are presented in section 4. The following sections delineate the corpus data: section 5 presents the frequencies of personal pronouns and their implications for the question in (4a); section 6 deals with the question in (4b) and examines the frequencies of word orders. Finally, section 7 is devoted to the question in (4c) and examines the frequencies of diathetic alternations with the verb classes at issue. Section 8 discusses the implications of the observations in the corpus for the research questions of this article.

4. Method

4.1 Relation to prior corpus studies

There is a long history of quantitative corpus studies on German word order that constitute the empirical background of the present study. Hoberg (1981) is the first study in this spirit examining written texts (from the Mannheimer Corpus) which shows a clear asymmetry between SO and OS sentences in written German. This study identifies the role of animacy, case (different behaviour of accusative and dative arguments), and the particular word order preferences of pronouns as crucially determining word order distributions in the German middlefield. Similar results are contributed by the later and larger corpus study in Kempen and Harbush 2005, which investigates determinants of word order variation in the NEGRA II and the TIGER corpus. Next to the importance of pronominal realization and case for argument order

in the middlefield, these authors identify animacy, definiteness, and referential ease as triggers for argument reordering. A large recent corpus study on word order in German (based on the newspaper corpus of the IDS) is Bader and Häussler 2010, which pays particular attention to the factors determining OS order, which are investigated in a special OS corpus that was composed for this purpose. The authors show among other things that OS order in the middlefield is triggered by lexical-semantic factors such as animacy and verb semantics while OS order with O realization in the prefield is due to discourse-related constraints (such as topic-first) and lexical-semantic factors in equal measure. Next to these large scale studies, there are a number of further corpus studies that focus on more specific argument constellations or factors or limit themselves to specific syntactic domains, either prefield or middlefield (see e.g., Primus 1994, Heylen & Speelman 2003, Heylen 2005, Kempen & Harbush 2004, Weber & Müller 2004).

The present work contributes to this rich research paradigm by focusing on the role of different verb classes in determining word order variation. By observing argument order in relation to particular verb classes we want to gain more fine-grained statements about the influence of argument properties on word order preferences in speech production. Based on intuition data, syntactic studies have claimed that the thematic properties of the arguments have an impact on argument order; the aim of the present corpus study is to examine whether these claims correspond to preferences in speech production as reflected in corpora. Some differences have already been reported in previous studies (see, e.g., Bader and Häussler 2010 on the identification of particular verbs and verb classes such as experiencer-object verbs being associated with OS order, see also Lenerz 1977 for this observation); the present study will examine exactly these differences in a rich

database and will disentangle the effects of the thematic properties of the arguments from the effects of further relevant factors. In addition, the present study examines the impact of factors that have an influence on linearization not only with respect to word order phenomena but also with respect to the choice of subject along with the choice of active and passive voice, which is known to be an alternative means to deviate from the canonical configuration in which the actor precedes the undergoer (see Mathesius 1975, Prat-Sala 1997, van Nice & Dietrich 2003 on German).

4.2 *Choice of verbs*

In order to answer the research questions in (4), we compare the constructional behaviour of four different verb classes. The focus of our study is on experiencer-object verbs, for which the experiencer constituent may be either an accusative or a dative object. Accusative experiencer-object verbs are further distinguished in \pm agentive and non-agentive, since this distinction is crucial for syntactic assumptions (see section 2.1). In sum, there are three interesting classes of experiencer objects: experiencer objects of \pm agentive transitive verbs, experiencer objects of non-agentive transitive verbs, and dative experiencer-objects. The behaviour of the experiencer-object verb classes will be compared with that of canonical transitive verbs in order to detect deviations from the neutral pattern.

For each of these classes, we examined ten sample verbs. There are not many dative experiencer-verbs in German. For the purposes of our study we selected the verbs in (7) from a larger group of 33 verbs mentioned in the literature (Klein & Kutscher 2002). We excluded verbs occurring rarely in written texts (e.g., *dämmern*

‘begin to dawn on so.’, *stinken* ‘be cheesed off with’) and verbs with frequent homonymic forms (e.g., *auffallen* ‘catch so.’s eye’, *entfallen* ‘slip so.’s mind’).⁴

(7) dative experiencer-object verbs

behagen ‘be to so.’s liking’, *einfallen* ‘come to mind’, *einleuchten* ‘make sense to so.’, *gefallen* ‘please’, *imponieren* ‘impress’, *leidtun* ‘feel bad about’, *missfallen* ‘displease’, *nahegehen* ‘affect so.’, *schmecken* ‘taste’, *widerstreben* ‘be reluctant to’

The subclassification of accusative experiencer-object verbs in \pm agentive and non-agentive is not uncontroversial. Speakers (and linguists) vary in their intuitions about the semantic properties of individual verbs. For this reason, we created an inventory of 20 accusative experiencer-object verbs and conducted an acceptability test for the division of this inventory in two subgroups. Native speakers of German were instructed to estimate on a Likert-scale (1= very bad; 7 = very good) the well-formedness of control-test sentences; see illustrative example with the experiencer-object verb *erstaunen* ‘astonish’ in (8). The experiment was performed online (created with OnExp 1.2; <http://onexp.textstrukturen.uni-goettingen.de/>). Each speaker gave a single estimate for each verb and was presented all 32 verbs. The test sentences were

⁴ The verbs in (7) are heterogeneous with respect to auxiliary selection in the perfect form, a grammatical property that is often taken to distinguishing unaccusative from unergative verbs: two verbs form perfect tenses with *sein* (*einfallen*, *nahegehen*), the remaining verbs with *haben*. Note, however, that the present study is not designed to test hypotheses about the behaviour of *haben* vs. *sein* selecting dative verbs but rather about dative vs. accusative experiencer verbs.

pseudo-randomized. 32 native speakers participated in this study in July 2013 (20 female, age range: 17-52, age average: 25.7).

(8) *Die Kollegen haben beschlossen, Klaus zu erstaunen.*

‘The colleagues decided to astonish Klaus.’

The obtained results are summarized in Table 1. The ten verbs with the higher scores will be classified as EO ±agentive verbs in the following analyses while the ten verbs with the lower scores will be classified as EO –agentive verbs.

Table 1. Agentivity of accusative experiencer-object verbs

	EO ACC –ag			EO ACC ±ag			
	mean	SD	<i>n</i>	mean	SD	<i>n</i>	
<i>freuen</i> ‘give pleasure’	1.3	.6	32	<i>enttäuschen</i> ‘disappoint’	4.1	1.8	32
<i>wundern</i> ‘astonish’	1.3	.7	32	<i>aufregen</i> ‘excite’	4.6	1.4	32
<i>interessieren</i> ‘concern’	1.8	1.1	32	<i>irritieren</i> ‘confuse’	4.9	1.7	32
<i>befremden</i> ‘alienate’	2.5	1.7	32	<i>amüsieren</i> ‘amuse’	5.2	1.6	32
<i>bedrücken</i> ‘depress’	2.5	1.3	32	<i>langweilen</i> ‘bore’	5.4	1.5	32
<i>entsetzen</i> ‘appall’	3.1	1.5	32	<i>reizen</i> ‘stimulate, provoke’	5.5	1.7	32
<i>ekeln</i> ‘disgust’	3.3	1.7	32	<i>erschrecken</i> ‘frighten’	5.6	1.6	32
<i>faszinieren</i> ‘fascinate’	3.5	1.6	32	<i>ärgern</i> ‘annoy’	6.0	1.6	32
<i>empören</i> ‘outrage, anger’	3.5	1.7	32	<i>überraschen</i> ‘surprise’	6.0	1.4	32
<i>anwidern</i> ‘nauseate’	3.6	1.6	32	<i>nerven</i> ‘bother’	6.1	1.7	32

The canonical transitive verbs in our study serve as a control condition. It is crucial that the point at issue is not to make generalizations about the typical behaviour of canonical transitive verbs but to examine whether particular properties are exactly associated with experiencer-object verbs or equally hold true for other verbs without an experiencer object. For this purpose we selected the verbs in (9), which occur with similar animacy configurations as experiencer-object verbs, meaning they occur with an animate undergoer and they allow for an animate or

inanimate actor. These verbs involve the affectedness of an animate undergoer, but in contrast to experiencer-object verbs the action at issue does not refer to a cognitive or emotional process that takes place within the consciousness of the affected animate.

(9) canonical transitive verbs

beeinträchtigen ‘impair’, *behindern* ‘hinder’, *blenden* ‘blind’, *infizieren* ‘infect’,
heilen ‘cure, remedy’, *retten* ‘rescue’, *schützen* ‘protect’, *vergiften* ‘poison’,
wecken ‘wake’, *zerquetschen* ‘squash’

4.3 Sampling

We extracted a dataset of 40 (verbs) × 1,000 (tokens) = 40,000 tokens from the IDS corpus (COSMAS-Database, Corpus *W-öffentlich*, containing a total of 2,291,520,000 word forms).⁵ The tokens for each verb were extracted by using the randomization function of the web interface on lemma-based queries for each verb. Further properties of the data in order to identify the critical datasets for the hypotheses at issue were manually annotated. We restricted ourselves to main declarative clauses with two (lexical, pronominal or clausal) arguments, in which the respective verb occurs in a finite form (including infinitives and participles only if they are part of a periphrastic tense form, e.g., future and perfect). This excludes occurrences of the target verbs in subordinate clauses, questions, exclamatives, headlines etc. as well as sentences in which the verbs at issue occur without two overtly realized arguments. For experiencer-object verbs in particular, we excluded non-experiential occurrences

⁵ The corpus *W-öffentlich* contains written language, mainly from newspapers and written prose. The material used in this article was extracted between May and September 2010.

of the verbs in our inventory. The subset of valid tokens contains 9,761 sentences (24.4% of the entire dataset), see Table 2, second line (classified according to verb class). For the hypotheses in section 7, we are using the subset of tokens that display two arguments realized as lexical noun phrases, which is a dataset of 4,319 sentences (44.2% of the tokens with two arguments), see Table 2, third line. In section 6, we will examine the word order option of the subset of active clauses in this dataset, which comprises 2876 sentences (66.6% of the sentences with two lexical arguments); see Table 2, fourth line.

Table 2. Categories in the corpus

	canonical	EO ACC \pm ag	EO ACC -ag	EO DAT	total
total extract	10,000	10,000	10,000	10,000	40,000
main decl. clauses, two args.	1,248	2,047	2,873	3,593	9,761
two lexical arguments	990	974	1,191	1,164	4,319
active voice	767	527	418	1,164	2,876

4.4 Annotation

The valid tokens were annotated for word order, diathesis, and two properties of noun phrases, namely animacy and NP-type. Two word order properties are relevant for the purposes of our study, namely the ORDER between subjects and objects/non-subjects and the FIELD in which these arguments appear. With respect to the relative order we distinguish between sentences in which the subject precedes the object (SO) or vice versa (OS), see (10a-b) vs. (10c-d). The notion of subject (S) refers to the traditional understanding of the term in German grammar, i.e., it is always the nominative argument, which is not a claim about the syntactic status of non-nominative experiencers (see section 2.1). The notion of object (O) comprises objects of canonical verbs as well as accusative or dative (object) experiencers. Furthermore, we

use the notion of non-subject for agent and stimulus phrases in non-active voice (e.g., *die Malerin ist von den Naturgestalten fasziniert* ‘the paintress is fascinated by the natural figures’) or with anticausative experiencer verbs (e.g., *die Staatssicherheit interessiert sich für den Regisseur* ‘national security was interested in the young producer’). The feature FIELD refers to the position of the arguments in the clausal domains determined by the finite verb. We distinguish two cases that are relevant for German syntax (see section 2.4): either both arguments are in the middlefield of the clause, i.e. they follow the finite verb (see (10b) and (10d)), or one argument (S or O) occupies the prefield, i.e., the syntactic position preceding the finite verb (see (10a) and (10c)).

- (10) a. ORDER: SO; FIELD: prefield

*Das geschichtsträchtige Gebäude fasziniert auch A.*⁶

‘The historical building fascinates A, too.’ (A97/OKT.29430)

- b. ORDER: SO; FIELD: middlefield

In den A. faszinierte die Vielzahl der 52 Orchideenarten das Publikum.

‘In A. the multitude of 52 orchid types fascinated the public.’

(A01/SEP.30280)

- c. ORDER: OS; FIELD: prefield

Mich haben die Brunnen in der Altstadt fasziniert.

‘The fountains in the old town fascinated me.’ (A01/AUG.22096)

- d. ORDER: OS; FIELD: middlefield

Schon damals faszinierten sie märchenhafte Gestalten.

⁶ Proper nouns are anonymized in the cited examples.

‘Already at that time she was fascinated by the fabulous figures.’

(A99/AUG.58970)

The feature DIATHESIS has two possible values in our scheme, namely ‘active’ and ‘non-active’, see (11). The verbs in our sample occur in either active or non-active forms, except for the dative EO verbs that only occur in the active. The value ‘non-active’ comprises dynamic passives of transitive verbs (*er wurde enttäuscht* ‘he was disappointed’), stative passives (*er ist enttäuscht* ‘he is disappointed’) as well as anticausatives (*er interessiert sich für etwas* ‘he is interested in sth.’).

(11) a. DIATHESIS: active

A. enttäuschte sein Publikum auch in B. nicht.

‘A. did neither disappoint his audience in B.’ (I96/MAI.18738)

b. DIATHESIS: non-active

Über diese Brüskierung ist Nationalrat A. sehr enttäuscht.

‘The national councillor A. is very disappointed by this affront.’

(E98/OKT.27811)

With respect to the inherent properties of the arguments, we will concentrate on two features that are crucial for our research questions in (4) and can be annotated in a large corpus, namely ANIMACY (animate vs. non-animate) and NP-TYPE (pronominal vs. lexical; DEFINITENESS of lexical NPs). A factor that is known to have an influence on German word order and is not included in our annotation scheme is information structure, in particular focus and topic (see section 2.4). Although these concepts are certainly crucial for the choice of word order in German, the annotation of

observational data for information structural categories is by no means straightforward. They can be approximated by indirect measures, most importantly by the availability of a referent in the preceding text (Givón 1994 ed., Weber & Müller 2004, Bresnan & Hay 2008, Gundel et al. 2005), but they cannot be unambiguously determined by contextual cues, since speakers have the choice between different information structures under identical contextual conditions. In the present large-scale corpus study, we did not annotate the contextual categories. The distinction between DEFINITENESS levels is expected to be an indicator of contextual influences: definiteness correlates with givenness, such that definite descriptions are more likely to relate to given referents than indefinite descriptions.⁷

ANIMACY properties are known to play an important role in word order and are necessary for our annotation scheme in order to disentangle the effects of animacy and the pure effect of experiencer-objects (see research question in (4b)). We annotated ANIMACY as a binary feature for each argument with the following values: ‘animate’ and ‘inanimate’ referent. So-called metonymically used animates such as institutions, organizations etc., which participate in the same events as animate beings and may exercise control over events as conscious entities (e.g., *Verband* ‘organization’, *Behörde* ‘public authority’, *Firma* ‘company’, *Regime* ‘regime’, etc.) were annotated as ‘animates’ (see also Bader & Häussler 2010: 730, who show that this class of NPs behaves like ‘animates’).

⁷ The relevant issue for the corpus frequencies is exactly this correlation; givenness is certainly not a sufficient condition for definite descriptions (the referent must also be uniquely identifiable in the context) and – at least in some accounts – is the result of a conversational implicature (see Abbott 2008).

NP-TYPE refers to the distinction between lexical and pronominal NPs and the DEFINITENESS levels of lexical NPs. Pronominal NPs contain personal pronouns. The reason for separating this category is that personal pronouns follow particular word order rules in German that do not hold for other types of NPs. Furthermore, we distinguished between local (3rd person) and non-local (1st or 2nd person) pronouns and we annotated 3rd person expletives. Lexical NPs are classified into definite and indefinite. Definite NPs contain NPs with definite articles (e.g., *der* ‘the’), demonstratives (e.g., *dieser* ‘this’), definite quantifiers (e.g., *jeder* ‘every/each’), possessive pronouns (e.g., *mein* ‘my’) as well as proper nouns without a determiner. Indefinite NPs comprise NPs with indefinite articles (e.g., *ein* ‘a’), indefinite quantifiers (e.g., *einige* ‘some’, *viele* ‘many’) as well as bare common noun NPs (see Bader & Häussler 2010:738 for a similar classification).

In the annotation of ANIMACY and DEFINITENESS at the NP level, we restrict ourselves to the distinctions between lexical NPs (pronominal NPs will be dealt with separately, because they are only relevant for a subset of our research questions). The distinctions of ANIMACY and DEFINITENESS relate to prominence scales as indicated in (12) (see Comrie 1981 and Dahl & Fraurud 1996 for animacy; Gundel et al. 1993, Fraurud 1990 for definiteness).

(12) Prominence scales

- a. ANIMACY: animate > inanimate
- b. DEFINITENESS: definite > indefinite

The central issue is how these scales are mapped onto the thematic role hierarchy *actor* > *undergoer* (see Foley & Van Valin 1984, Dowty 1991). There are

three logical possibilities for the status of actor and undergoer with reference to these scales, see (13): (a) the actor outranks the undergoer on a prominence scale; (b) both actor and the undergoer have equal status on a prominence scale; and (c) the undergoer outranks the actor on a prominence scale.

- (13) a. actor $>$ _{prominence} undergoer (non-disharmonic mapping)
b. actor \equiv _{prominence} undergoer (non-disharmonic mapping)
c. actor $<$ _{prominence} undergoer (disharmonic mapping)

The basic assumption is that deviations from the canonical pattern are expected whenever the prominence scales in (12) are disharmonically mapped to the thematic role hierarchy (Aissen 1999). In particular, this study will observe the effect of disharmonic mapping on the choice of subject and on the choice of word order. For this reason, we distinguish between two classes of sentences (separately for ANIMACY and DEFINITENESS): sentences that represent disharmonic mapping between the prominence scale and the thematic role hierarchy and sentences that do not represent disharmonic mapping in this respect. The resulting permutations of ANIMACY and NP-TYPE at the clause level are exemplified in (14).

- (14) a. ANIMACY = disharmonic; DEFINITENESS = disharmonic

Den Regisseur interessiert von nun an eine Frage.

‘From now on, the director is interested in one question.’

(SOZ06/AUG.00423)

- b. ANIMACY = disharmonic; DEFINITENESS = non-disharmonic
Der Versuch reizt den Trainer jedenfalls.
 ‘The attempt appeals to the trainer anyway.’ (NUN06/APR.02092)
- c. ANIMACY = non-disharmonic; DEFINITENESS = disharmonic
Drei ... Wildschweine sowie ein ... Rehkitz faszinierten vor allem die Kinder.
 ‘Three boars and a fawn fascinated especially the children.’
 (RHZ03/JUN.19102)
- d. ANIMACY = non-disharmonic; DEFINITENESS = non-disharmonic
Der Offizier A. rettete 1945 die Stadt.
 ‘The officer A. saved the city in 1945.’ (RHZ04/OKT.20270)

4.5 Data analysis

Sections 5-7 present the findings of the corpus study with respect to the research questions in (4). For the statistical analysis we used the same procedure in all sections. Corpus frequencies are analyzed through *generalized logit mixed models*. In each phenomenon at issue, we start with the question of which VERB CLASS model in (5)-(6) reaches the maximal goodness of fit in the data. We answer this question in comparing the alternative VERB CLASS models with a maximal specification of the fixed and random factors at issue. Model selection is based on the AIC values and on the results of log-likelihood tests on the goodness-of-fit of the compared models. After selecting the optimal VERB CLASS model, we conducted a backwards selection procedure by testing whether removing particular fixed factors or interaction effects has a significant impact on the informativity of the model, based again on the AIC-values and the results of the log-likelihood tests. Calculations were made with the function *glmer* of the R-package *lme4* (Bates et al. 2011).

5. Experiencers and discourse-prominence

There is a well-known asymmetry in the preferred discourse status of the verbal arguments: subject constituents typically introduce the topic of the utterance (Lambrecht 2000), which follows from the preference for subjects to express the running topic in a maximally coherent discourse chain (see a clear formulation of this principle in *Centering Theory*, Walker et al. 1998). The reflex of this preference in discourse is the empirical observation that subject constituents are, more frequently than object constituents, either dropped or realized as pronominal NPs – depending on the available options in the grammar at issue (Du Bois 1987, Gordon et al. 1993, Poesio 2008). This background motivates a straightforward prediction for the frequencies of pronominal arguments in our data. Since the topic at issue in a discourse chain is preferably realized as subject, pronominal subjects are expected to occur more frequently than pronominal objects, see the examples in (15a-b).

- (15) a. Pronominal subject, canonical verb

Damit schütze er die Interessen der Patienten.

‘With this he would protect the interests of the patients.’

(RHZ98/JUL.01488)

- b. Pronominal object, canonical verb

Tatsächlich schützte ihn der Bürgermeister...

‘The mayor protected him indeed ...’ (Title unknown)

The crucial question for our considerations is how the arguments of experiencer-object verbs behave with respect to the frequency of pronominal

realization, see (4a). If experiencers are particularly prominent in discourse (see Haspelmath 2001, Bickel 2004:77), then experiencer-object verbs are expected to show the mirror image of canonical transitive verbs concerning the pronominal realization of their arguments: pronominal objects are predicted to be more likely than pronominal subjects; the cases at issue are exemplified in (16a-b). The notion of ‘discourse prominence’ should be understood as high activation of the referent through the context or the discourse situation, such that it can be identified as the content of a pronominal expression (Gundel et al. 1993).

(16) a. Pronominal subject, experiencer-object verb

Gerne imponiert er den Touristen.

‘He likes to impress tourists.’ (BRZ05/NOV.00289)

b. Pronominal object, experiencer-object verb

Ihr imponierte deren einfache, aber selbstbewusste Lebensweise.

‘She was impressed by their simple but self-confident way of living.’

(A08/MAR.03673)

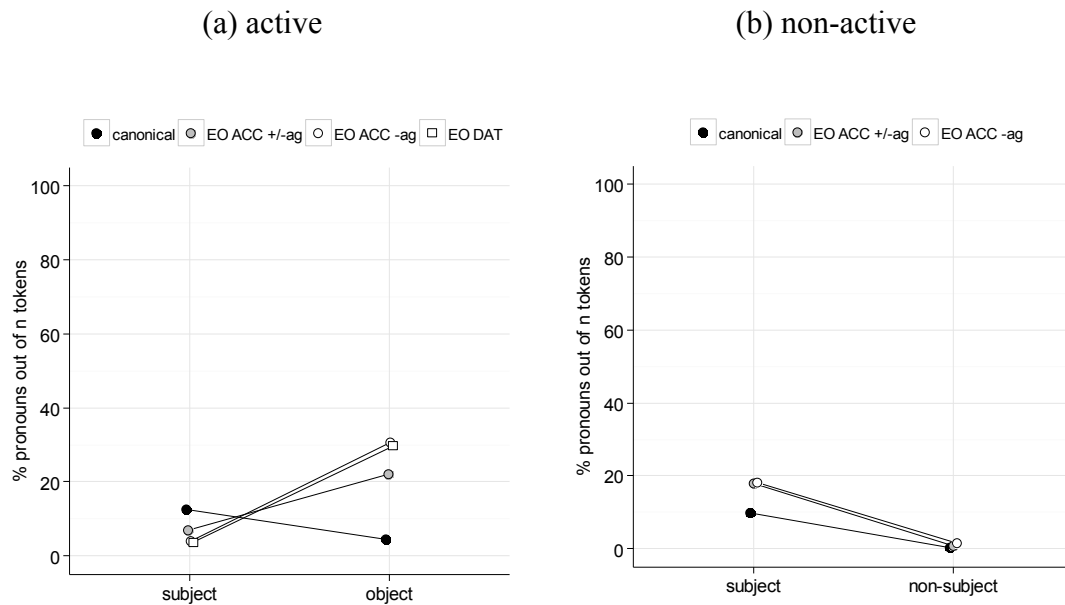
The distinction between lexical and pronominal NPs is coded in our data (see NP-TYPE in section 4.4). In order to examine the difference between canonical verbs and classes of experiencer-object verbs, we queried the proportions of pronominal arguments in the valid data (9,761 sentences with two overtly realized arguments, i.e., 19,522 arguments; see Table 2). The hypothesis relates to the preference for topics to be realized as subjects in a discourse chain. The local persons (first and second person) are not informative for this hypothesis, since they are not necessarily continuous topics. Furthermore, the hypothesis relates to referential uses of the third

person and not to third person expletive pronouns (see also Speyer 2007: 87). For this reason, local person pronouns (545 tokens with subjects; 2,349 with objects) and expletive uses of the third person pronoun (1,070 tokens found in subjects) were excluded from this analysis. The properties of the remaining tokens (15,558 arguments) are reported in the following.

The relevant data lies in the relative frequency of referential pronouns out of the third person NPs, see Figure 1 (counts are listed in Appendix A). The left panel presents the results for active clauses, in which the subject is the actor constituent (i.e., the agent of canonical verbs or the stimulus of EO verbs); the right panel presents the results of non-active clauses, in which the subject is the undergoer constituent (i.e., the patient of canonical verbs or the experiencer of EO verbs). The canonical verbs make up the baseline in order to estimate the properties of experiencer-object verbs. The observed pattern in this verb class corresponds to our expectations: subjects of canonical verbs are more frequently pronominal than objects, both in active voice (left panel) and non-active voice (right panel).

All experiencer-object classes show a mirror image of the canonical transitive verbs in the active voice. The object (experiencer) is more frequently pronominal than the subject (stimulus). The same asymmetry between experiencers and stimuli is maintained in the non-active voice, which suggests that – in contrast to canonical transitive verbs – the information state of the arguments of experiencer-object verbs (as reflected in pronominalization) does not account for the choice of voice. Independent of the diathetic realization of the verb, the most frequently pronominalized argument is the experiencer (either as object or as subject).

Figure 1. Pronouns out of third person NPs



In order to examine the validity of the observations in Figure 1, we fitted a *generalized logit mixed model* on the choice of pronominal vs. non-pronominal arguments with FUNCTION (subject; object), and VERB CLASS as fixed factors, and VERB as a random factor (including the intercept and slope with FUNCTION). Since verbs with dative experiencer-objects only appear in the active voice, it is not possible to calculate a single statistical model with all verb groups in both voice options. For this reason, we report separate models for the active and non-active data.

As a first step, we compared the alternative models of verb class in (5) and (6). The maximal goodness-of-fit for the active data was reached by the ACTOR & UNDERGOER MODEL, i.e., the threefold contrast ‘canonical vs. EO accusative (\pm agentive) vs. EO accusative/dative ($-$ agentive)’ verbs ($AIC = 7,110.4$, $df = 9$ compared to $AIC = 7,113.2$, $df = 11$ of the CASE & ACTOR & UNDERGOER MODEL that contains all contrasts; a log-likelihood test reveals that the ACTOR & UNDERGOER MODEL does not result in a significant loss of information). The estimates of the

ACTOR & UNDERGOER MODEL for the active data are given in Table 3. A log-likelihood test on the goodness of fit reveals that the interaction effect between VERB CLASS and FUNCTION cannot be removed from the model, since the loss of information in a model without this interaction effect is significant ($\chi^2(2) = 42.4$; $p < .001$).

Table 3. Pronominalization in active clauses

factor	estimated level	estimate	S.E.	z-value	p (> z)
intercept		2.6	0.1	20.5	< .001
V-CLASS	EO.ACC \pm ag	0.7	0.3	1.9	< .05
	EO -ag	0.9	0.3	2.9	< .01
FUNCTION	object	-0.8	0.2	-4.5	< .001
V-CLASS : FUNCTION	EO.ACC \pm ag & object	-2.2	0.5	-4.7	< .001
	EO -ag & object	-1.4	1.1	-3.7	< .001

In the non-active clauses, the maximal goodness of fit is reached by the UNDERGOER MODEL, which captures the contrast between patients (canonical verbs) and experiencers (both classes of experiencer-object verbs) ($AIC = 2,281.4$, $df = 7$ compared to $AIC = 2,284.7$, $df = 9$, of the ACTOR & UNDERGOER MODEL; a log-likelihood test reveals that the loss of information of the model with the fewer parameters is not significant).⁸ Log-likelihood tests reveal that the interaction between VERB CLASS and FUNCTION is not significant, whereas both factors have significant main effects (a comparison between a model with two main effects and a model without FUNCTION: $\chi^2(1) = 63.1$; $p < .001$; between a model with two main effects and

⁸ Since EO dative verbs are not part of this data, model comparisons relate to those models in (5) and (6) that do not have the CASE contrast.

a model without VERB CLASS: $\chi^2(1) = 4.6$; $p < .05$). The estimates of the model with the maximal goodness of fit are given in Table 4.

Table 4. Pronominalization in non-active clauses

Factor	estimated level	estimate	S.E.	z-value	p (> z)
Intercept		2.1	.3	8.3	< .001
V-CLASS	EO.ACC	-.6	.3	-2.2	< .05
FUNCTION	non-subject	3.4	.3	12.8	< .001

The findings confirm the prediction that third person pronouns are more frequent with subject constituents of canonical verbs, which reflects the preference for encoding the running topic through the subject in a maximally coherent discourse chain; the difference between active and non-active voice replicates previous findings on the asymmetry in the contextual properties of voice (see, in particular, corpus studies in Givón 1994). The novel contribution of the presented data is that this generalization is restricted to canonical verbs, since experiencer-object verbs show a preference for third person pronominal experiencers independent of their syntactic realization as nominative or non-nominative constituents (in non-active or active voice respectively). This conclusion is justified by the significant interaction effect between VERB CLASS and FUNCTION in the active clauses.

The frequencies of pronominal realization of third person arguments in our corpus confirm that experiencer arguments relate more frequently than stimulus arguments to referents that are highly activated in discourse. This discourse preference implies that experiencers are more likely discourse topics than stimulus arguments. Having thus established the discourse-prominence of experiencers, which holds independent of their grammatical realization as subject or object, we are now in

the position to disentangle the sources of the grammatical behaviour of experiencers in more detail.

6. Experiencer-first and word order

This section examines the question whether experiencer-object verbs possess different word order properties from canonical transitive verbs; see research question (4b). It has been shown that experiencer-objects display subject-like properties in several languages – among other things occurring in the subject position in languages like Icelandic (see discussion in section 2.1). For German in particular, a large number of studies on word order claim that dative experiencers precede nominative arguments (Fanselow 2000, 2003; Primus 1994, Lenerz 1977, Hoberg 1981). Regarding accusative experiencers, the empirical situation is less clear. There is, however, experimental evidence that accusative experiencers have different linearization properties than accusative patients (Scheepers 1997, Scheepers et al. 2000, Haupt et al. 2008). The exceptional behaviour of experiencer objects has been attributed to their non-agentive readings; thus, the distinction between agentive and \pm agentive accusative verbs may be relevant for word order phenomena (see Arad 1998a, 1998b; Landau 2010; and also section 2.1).

In our corpus, deviations from canonical word order occur in all verb classes, as illustrated in (17). We expect to learn from the corpus frequencies whether the likelihood for these deviations to occur is equal across verb classes or not. If different verb classes behave differently, the question will be which model in (5)-(6) may capture these differences.

(17) a. OS order, canonical verb

As Kinder hat der blutrünstige B zerquetscht.

‘The bloodthirsty B crushed A’s children.’ (N97/DEZ.50201)

- b. OS order, EO accusative ±agentive verb

Den Stürmer selbst überraschte die Nachricht.

‘The news surprised the striker himself.’ (HMP07/NOV.02095)

- c. OS order, EO accusative non-agentive verb

Amerika-Urlauber freut der starke Euro.

‘American tourists are happy about the strong Euro.’ (M07/APR.02917)

- d. OS order, EO dative verb

A. selbst gefällt das Plakat gut.

‘A. himself likes the poster very much.’ (BRZ07/SEP.03714)

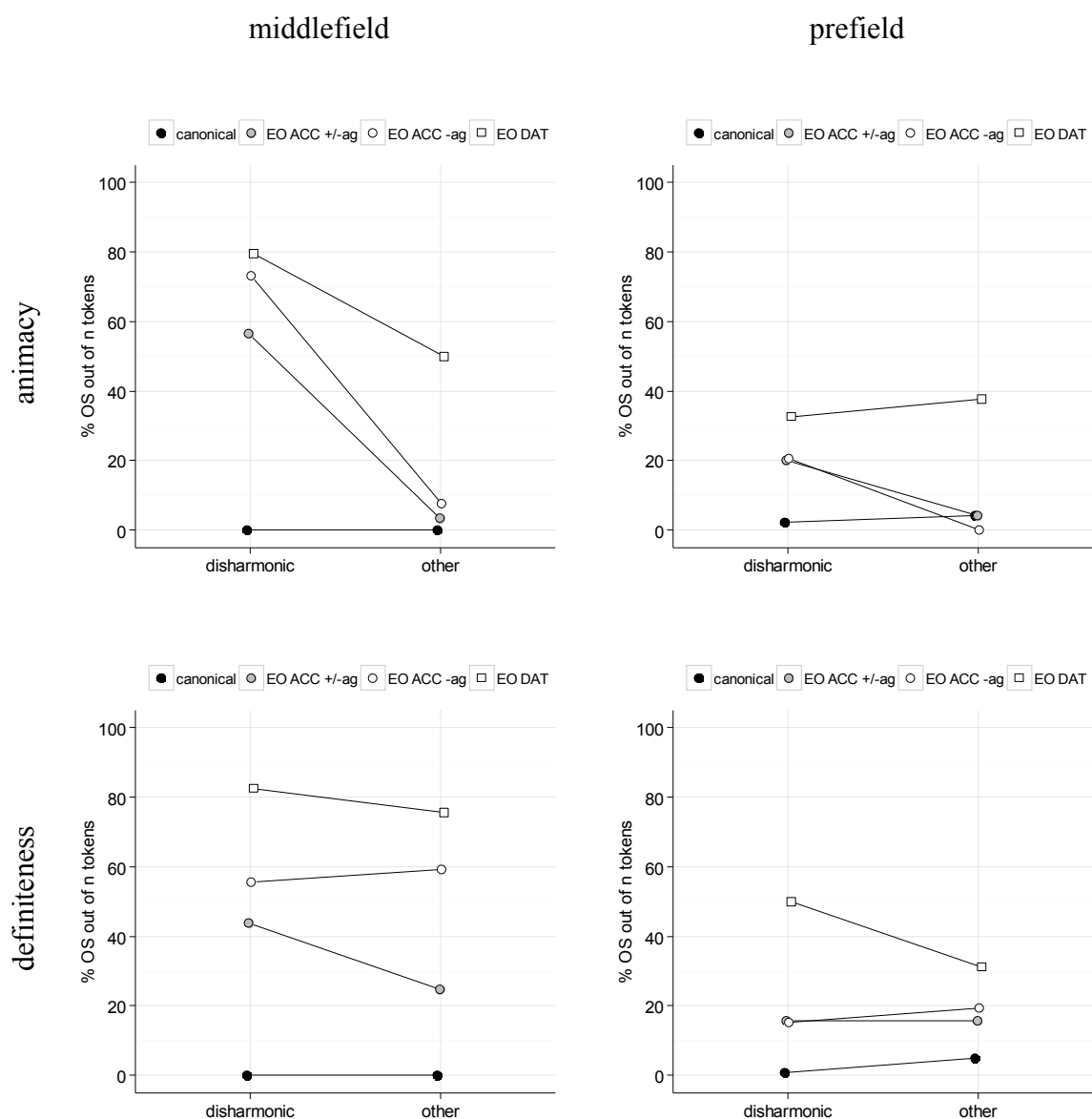
The relevant subcorpus for the questions of this section is the set of sentences with a verb in active voice and two lexical noun phrases (2,876 sentences, see Table 2). Personal pronouns follow particular rules in German syntax, and since the target of this study is to investigate the effect of verb classes on word order, the linearization preferences of personal pronouns are left out of consideration. A general inspection of the frequencies of the constructions in (17) confirms the hypothesized scale in the full VERB-CLASS MODEL containing all three contrasts between the four groups. The overall proportions of OS order are as follows: (a) canonical verbs (26 out of 767 tokens; 3.4%); (b) EO accusative ±agentive verbs (95 out of 527 tokens; 18.03%); (c) EO accusative non-agentive verbs (101 out of 418 tokens; 24.2%); (d) EO dative verbs (448 out of 1,164 tokens; 38.5%). These proportions descriptively confirm that VERB CLASS plays a role.

Three relevant factors must be considered before drawing any generalization from this data. As already introduced in research question (4b), we need to know

whether the effect of particular types of arguments on word order is a direct effect of the thematic properties of the arguments at issue or whether it can be traced back to further semantic and contextual properties of this type of arguments. Semantic properties are represented by ANIMACY in our dataset, which involves a clear differentiation between experiencers and patients, since the former are by definition animate while the latter are not necessarily so. Contextual properties can be approximated by DEFINITENESS under the assumption that it correlates with givenness (see section 4.4). Given that animate-first and given-first effects have been established in previous studies, the crucial issue is whether experiencer-first effects can be traced back to these more general principles. The third relevant factor is FIELD, namely the distinction between prefield and middlefield, since it is known that these two domains of the German clause possess different properties (see section 2.4).

The proportions of OS order in the corpus are summarized in Figure 2 (see counts in Appendix B). These data indicate that the order in which the object (O) precedes the nominative (S) is generally more frequent with experiencer-object verb classes. Furthermore, the effects are larger in the middlefield than in the prefield. Disharmonic animacy (undergoer outranks actor in animacy hierarchy) has a clear impact on the proportions of OS order of accusative experiencer verbs, which is manifested in a higher ratio in the middlefield than in the prefield. The effects of definiteness are less clear and not uniform across verb classes.

Figure 2. Choice of order (active clauses)



We fitted a *generalized logit mixed model* in this subset of the corpus data and started with a maximal model with ORDER as a dependent variable, all possible interaction effects of VERB CLASS, ANIMACY, DEFINITENESS, and FIELD as fixed factors and VERB as random factor (including the intercept and slopes of VERB with the fixed factors ANIMACY, DEFINITENESS, and FIELD). As a first step, we compared the alternative models of verb class in (5)-(6). The model with the maximal goodness

of fit is the CASE & UNDERGOER MODEL, i.e., the threefold contrast ‘canonical vs. EO accusative vs. EO dative’ verbs. The model ignoring the contrast between two subclasses of accusative experiencers (depending on agentivity) fits better to the data ($AIC = 2,470.0$ compared to $AIC = 2,475.7$ of the CASE & ACTOR & UNDERGOER MODEL that contains all possible contrasts) and does not display a significant loss of information (a log-likelihood test with $df = 7$ results in a chi-square value that does not correspond to a significant p -value), i.e., we do not have evidence that the contrast between two subclasses of EO accusative verbs is a necessary parameter for understanding the word order frequencies in the corpus. The goodness-of-fit of this model is also significantly better than the corresponding values of the models with less parameters, i.e., the CASE MODEL ($\chi^2(7) = 30.7, p < .001$), the ACTOR MODEL ($\chi^2(7) = 36.6; p < .001$) and the UNDERGOER MODEL ($\chi^2(7) = 25.3; p < .001$). Hence, a further reduction of the verb class distinctions is empirically not justified.

The findings of the model comparison indicate that the distinction between the two subclasses of EO accusative verbs in the lexicon does not have a significant impact on the word order frequencies. This result does not lead to the conclusion that agentivity does not play a role, since we cannot estimate the exact frequency of agentive and non-agentive occurrences of the \pm agentive verbs. The fact that \pm agentive verbs do not differ from $-$ agentive verbs in word order may be due to the fact that the non-agentive occurrences of the former verb class are particularly frequent. Hence, the precise conclusion is that the lexical distinction between two subclasses of EO accusative verbs (and not the distinction between agentive and non-agentive uses) does not have a significant role on word order choice.

In the following, we present an analysis of the data in terms of the CASE & UNDERGOER MODEL, i.e., ignoring the subclasses of accusative experiencer verbs.

Based on a backwards selection procedure, we removed all interaction effects that do not significantly affect the informativity of the model and came up with a final model containing three interaction effects: VERB CLASS and ANIMACY (log-likelihood test: $\chi^2(2) = 13.4$; $p < .01$), VERB CLASS and FIELD (log-likelihood test: $\chi^2(2) = 13.9$; $p < .001$), FIELD and ANIMACY (log-likelihood test: $\chi^2(1) = 9.1$; $p < .01$).⁹ All interactions with DEFINITENESS as well as the main effect of this factor could be removed from the model without significant loss of information. The final model containing the three significant interaction effects displays a standard error inflation, which is due to collinearity of the factors VERB CLASS and FIELD. Standard error inflation does not disappear through centering: hence, the interaction effect between VERB CLASS and FIELD is not informative in our data and has to be removed from the model. The estimates of the final model are listed in Table 5.

Table 5. Model estimates for the choice of order

factor	estimated level	estimate	S.E.	z-value	p (> z)
intercept		-2.5	.4	-6.4	< .001
VERB CLASS	EO.ACC	.7	.6	1.2	= .2
	EO.DAT	2.9	.6	5.1	< .001
FIELD	prefield	.1	.5	.3	= .8
ANIMACY	disharmonic	2.4	.6	4.2	< .001
FIELD : ANIMACY	prefield : disharmonic	-2.1	.5	-4.5	< .001
V-CLASS : ANIMACY	EO.ACC : disharmonic	2.6	1.3	2.1	< .05
	EO.DAT : disharmonic	-2.3	.6	-3.9	< .001

In sum, the presented findings indicate that three classes of verbs behave differently with respect to the frequency of OS order: canonical verbs, experiencer-

⁹ The log-likelihood tests relate to comparisons between a model containing all twofold interactions and a model in which the interaction effect of interest is removed.

object accusative verbs, and experiencer-object dative verbs. Verb class effects interact with animacy, which means that the effect of the individual verb groups differs across animacy options (disharmonic vs. other). This interaction effect as well as the main effect of verb class in our data indicate that the influence of verb classes on word order is not reducible to animacy. Visual inspection of the data shows that fronting dative experiencers is generally independent from animacy, while fronting accusative experiencers occurs much more frequently if the experiencer outranks the stimulus on the animacy hierarchy. Canonical verbs generally occur infrequently in the OS order (across ANIMACY levels).

The question is whether the role of case on the word order frequencies of experiencer-object verbs can be accounted for through the ambiguity potential of nominative and accusative NPs. The subset of the data relevant in this respect are the clauses without disharmonic animacy (see animacy: ‘other’ in Figure 2). These clauses are not disambiguated by animacy and display a contrast between dative and accusative verbs, since OS order only rarely occurs with the latter verb class. Structurally ambiguous clauses are frequent with accusative objects in German: a nominative-accusative contrast is only available in the masculine singular paradigm of determiners (*der* ‘DEF:M.SG.NOM’ vs. *den* ‘DEF:M.SG.ACC’; *ein* ‘INDEF:M.SG.NOM’ vs. *einen* ‘INDEF:M.SG.ACC’), adjectives (*große(r)* ‘big:(IN)DEF.M.SG.NOM’ vs. *großen* ‘big:M.SG.ACC’) and a subset of nouns (e.g., *Mensch* ‘human.M.SG.NOM’ vs. *Menschen* ‘human.M.SG.ACC’). In addition, verbal agreement can disambiguate the grammatical functions if subject and object differ in number. All other cases are structurally ambiguous and can be only disambiguated through the context, see (18a). In the 218 clauses with two lexical NPs that do not differ in animacy (both in the prefield and in the middle field), we found 55 (25.2%) structurally ambiguous clauses (see Table 6):

all these clauses are disambiguated by contextual cues and in all clauses the subject is the first argument. Ambiguity is rare with EO dative verbs: it arises if neither argument has a determiner, as is the case with proper nouns and if the two arguments do not differ in number, as in (18b). In our corpus, we found 2 structurally ambiguous clauses (2.9%) out of a total of 67 clauses with an EO dative verb having two lexical animate arguments (see Table 6). In both sentences, the first argument is interpreted as nominative based on contextual cues.

(18) a. Ambiguous EO.ACC

Und jedes Mal entsetzten die jugendlichen Täter die Richter mit ihrer Kaltschnäuzigkeit - von Schuldbewusstsein keine Spur.

‘And every time the adolescent delinquents appalled the judges with their coolness – no sense of guilt.’ (RHZ01/MAI.13687)

b. Ambiguous EO.DAT

Rink (der zuletzt im Sommer 1999 beim Konföderationen-Cup in den USA seine Länderspiele drei und vier bestritt,) hat Ribbeck bei seinen letzten Bundesligaauftritten imponiert.

‘Rink (who played his international matches three and four most recently in the summer of 1999 at the Confederations Cup in the USA) impressed Ribbeck with his last Bundesliga performances.’ (M00/MAR.04643)

The frequencies in Table 6 lead to the following conclusions. Structural ambiguity is more frequent with accusative (25.2%) than with dative (2.9%) verbs. Furthermore, structural ambiguity has an influence on word order, since all structurally ambiguous clauses in our corpus are disambiguated in the form of nominative-first clauses based on contextual cues. However, the role of ambiguity

does not explain the different frequencies of OS order with accusative and dative EO verbs. In the subset of non-ambiguous clauses, the OS order is significantly more frequent with dative verbs (38.5%) than with accusative verbs (4.9%); $\chi^2 = 42.3$, $p < .001$.

Table 6. Ambiguity and word order¹⁰

		SO		OS		total	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
EO.ACC	ambiguous	55	100.0	0	0.0	55	100
	non-ambiguous	155	95.1	8	4.9	163	100
EO.DAT	ambiguous	2	100.0	0	0.0	2	100
	non-ambiguous	40	61.5	25	38.5	65	100

Definiteness did not have a significant effect in this dataset: the frequencies in Figure 2 suggest that initial objects with EO accusative \pm agentive verbs in the middlefield and EO dative verbs in the prefield occur more frequently under disharmonic definiteness. However, since the effects related to DEFINITENESS are not significant, this conclusion is empirically not justified. The relevant result for our considerations is that the effect of verb classes on word order cannot be reduced to an effect of definiteness.

Finally, the data reveal a difference between prefield and middlefield in accordance with the established assumptions about German clausal syntax: the data in Figure 2 (in particular top panels) show that similar effects appear in the middlefield

¹⁰ This table contains active experiencer-object clauses with two lexical arguments that do not differ in animacy. The sums of EO.ACC and EO.DAT verbs in this table are identical with the sums of EO.ACC verbs (218) and EO.DAT (67) verbs in the appendices B.1 (middlefield) and B.2 (prefield) (column ‘other’). Clauses were classified as ‘ambiguous’ if the grammatical function of subjects and objects is not disambiguated by structural means (i.e., morphological case and verb agreement).

and in the prefield while the size of these effects is larger in the middlefield. This is in line with the view on German syntax that effects of semantic asymmetries show up in the middlefield. However, this observation cannot be validated statistically – due to the collinearity between these factors.

7. Experiencer-first and choice of subject

The analysis in section 6 has shown that fronted non-nominative arguments occur more frequently with experiencer-object verbs than with canonical verbs and that this difference cannot be reduced to further asymmetries in animacy or definiteness. The next question is whether the observed word order phenomena are part of a general preference for linearizations in which the experiencer role occurs earlier in the clause. An alternative strategy to deviate from the linearization of an active clause in the canonical word order is the choice of a diathetic alternation. The verbs in our sample occur in forms that license different argument structures in the corpus, e.g., stative or dynamic passives, deagentives etc. (see discussion in section 2.2). Diathetic alternations are not just linearization options but also possess additional semantic properties, meaning that active and non-active voice do not have the same extension. It is clear that diathetic alternations such as the ‘stative passives’ cannot be chosen for any type of propositional content in which the corresponding active occurs. However, the critical assumption for the comparisons in the following is that there is a subset of propositional contents that can be described with either diathetic form and to this extent the choice of voice in speech production can be influenced by linearization preferences (see discussion in section 1). What we expect to learn from the corpus frequencies is to what extent the factors determining the choice of order also have an influence on the choice of voice. Illustrative examples from the corpus are given in (19). The choice of voice can only be observed with accusative object verbs (either

canonical or experiencer-object verbs) that possess counterparts in which the undergoer (patient or experiencer) is realized as a subject; dative verbs do not have diathetic alternations of this type. We restrict our observations to the subset of data with two lexical arguments, see Table 2, third line.

(19) a. Non-active canonical verb

Die 34-jährige Frau wurde durch ihre Katze geweckt.

‘The 34-year old woman was woken up by her cat.’ (RHZ08/SEP.00132)

b. Non-active EO accusative \pm agentive verb

Fans wurden von den Profis enttäuscht.

‘The fans were disappointed by the professional players.’

(RHZ98/APR.47691)

c. Non-active EO accusative non-agentive verb

Aber die Leute sind so pervers interessiert an diesen Figuren.

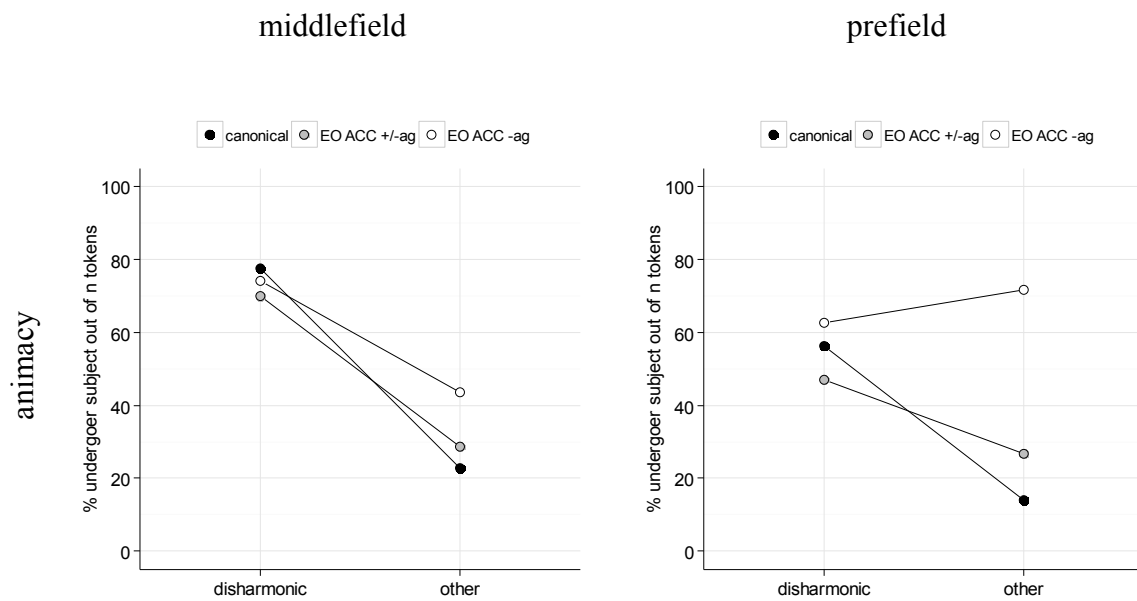
‘But the people are so perversely interested in these figures.’

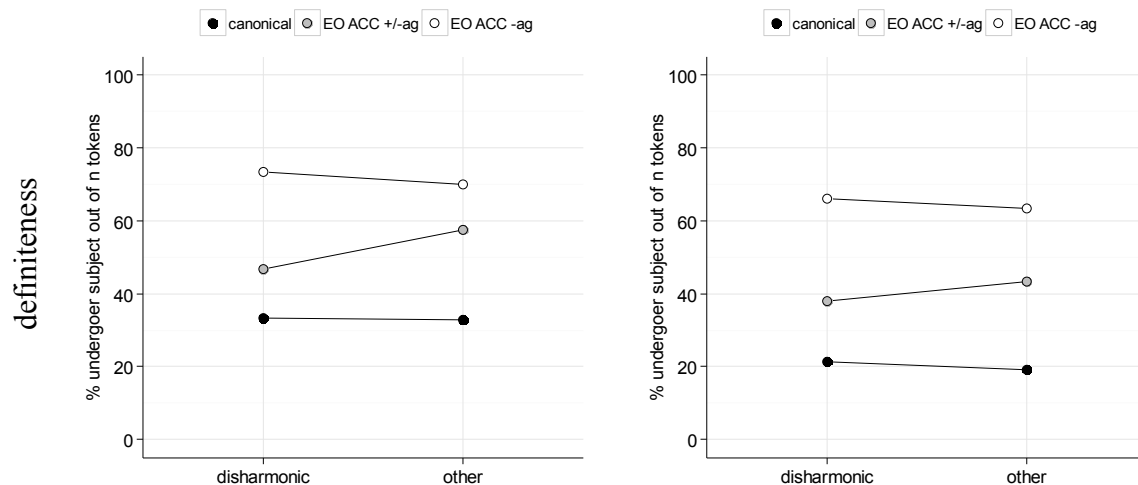
(NUN04/DEZ.00695)

The choice of voice was examined in the subset of tokens with two lexical arguments (4,319 clauses in Table 2), excluding the dative EO verbs (1,164 clauses) that only occur in active voice. The obtained proportions in the remaining (4,319 – 1,164=) 3,155 clauses are presented in Figure 3 (see counts in Appendix C). The proportions in the upper panels reveal that animacy plays an important role: the likelihood of a verb with an undergoer subject increases when the undergoer is lower in the animacy hierarchy than the actor (disharmonic animacy). This pattern is observed for all verb classes, except for the non-agentive EO verbs in the prefield.

The proportions in the lower panels indicate that definiteness does not have a consistent effect on the choice of voice. The proportions of non-active forms differ between verb classes: (a) canonical verbs: 22.5% (223 non-active out of total 990); (b) EO \pm agentive verbs: 45.9% (447 non-active out of total 974); (c) EO non-agentive verbs: 64.9% (773 non-active out of total 1191). The proportions of non-active forms also differ between the two clausal domains: (a) middlefield: 52.6% (340 non-active out of total 646); (b) prefield: 43.9% (1103 non-active out of total 2509) (see detailed counts in Appendix C).

Figure 3. Effects in the choice of subject





We fitted a *generalized logit mixed model* in this subset of the corpus data. We started with a maximal model with DIATHESIS as a dependent variable, the fourfold interaction between VERB CLASS, ANIMACY, DEFINITENESS, and FIELD as fixed factors and VERB as random factor (including intercepts and slopes with the fixed factors – expect for VERB CLASS). First, we compared the three reasonable VERB-CLASS MODELS; see (5)-(6), namely the UNDEGOER MODEL, the ACTOR MODEL, and the ACTOR & UNDERGOER MODEL, which is the full model with respect to this data.¹¹ The maximal goodness-of-fit is reached by the ACTOR MODEL ($AIC = 3,115.9$, $df = 26$ compared to $AIC = 3,127.4$, $df = 34$ of the full model; a log-likelihood test reveals that the ACTOR MODEL does not result in a significant loss of information). This result implies that the distinction between canonical verbs and EO \pm agentive verbs is an unnecessary stipulation for the understanding of the choice of diathesis.

After adopting the ACTOR MODEL, we examined whether all factors VERB CLASS, ANIMACY, DEFINITENESS, and FIELD and their interactions are indispensable

¹¹ Since EO dative verbs are not part of this data, model comparisons relate to those models in (5) and (6) that do not have the CASE contrast.

parameters of the model (based on model comparison, see section 4.5). All interactions containing DEFINITENESS as well as the main effect of this fixed factor can be removed from the model without significant loss of information. However, removing the threefold interaction effect between VERB CLASS, ANIMACY, and FIELD reveals a significant effect in the log-likelihood test ($\chi^2(1) = 4.1$; $p < .05$; comparing a model with only the interaction at issue with a model without any threefold interaction). The parameters of the final model, which contains this interaction and all embedded effects, are presented in Table 7.

Table 7. Parameters of the model for the choice of subject

factor	estimated level	estimate	S.E.	z-value	$p (> z)$
intercept		-1.3	.4	-3.2	< .01
V.CLASS	EO.ACC -ag	.8	.8	.9	= .3
ANIMACY	disharmonic	2.1	.3	5.9	< .001
FIELD	prefield	-.6	.2	-3.1	< .01
V.CLASS : ANIMACY	EO.ACC -ag : disharmonic	-1.3	.7	-.2	= .9
V.CLASS : FIELD	EO.ACC -ag : prefield	1.8	.7	2.9	< .01
ANIMACY : FIELD	disharmonic : prefield	-.5	.3	-1.9	= .05
V.CLASS : ANIMACY : FIELD	EO.ACC -ag : dis. : prefield	-1.6	.7	-2.2	< .05

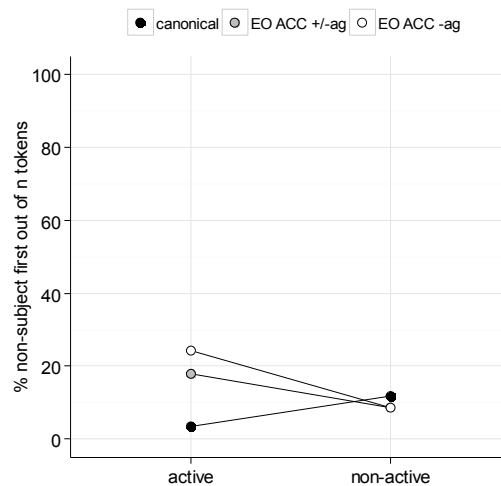
In sum, the data from the choice of subject reveal a contrast between non-agentive verbs and potentially agentive verbs (which comprise canonical and EO \pm agentive verbs). The statistical findings reveal a threefold interaction between VERB CLASS, ANIMACY, and FIELD, which accounts for the fact that non-agentive verbs occur very frequently in non-active forms across animacy levels in the prefield data, while all other verb classes involve an asymmetry in animacy. The observed difference cannot be straightforwardly accounted for by the current assumptions about German syntax. A source for differences between clausal domains (prefield vs.

middlefield) is the fact that two overtly realized arguments may be adjacent in the middlefield, but not if one of them is realized in the prefield. The processing difficulties involved in the adjacent realization of syntactic entities of the same type (Richards 2010) is a potential source for the higher frequencies of non-active voice in the middlefield. It is reported that OV languages display an intransitive verb bias accounted for through the processing difficulty of XYV linearizations (Ueno & Polinsky 2009). The fact that EO non-agentive verbs display higher proportions of non-active voice is in line with the view that earlier realization of the experiencer is more likely if the stimulus is not agentive. However, the fact that non-agentive verbs interact with animacy only in the middlefield is a result that cannot be easily accommodated through the known facts about German syntax. Since the prefield is the clausal domain that hosts topics and foci (see section 2.4), this result suggests that the choice of experiencers for these functions is independent from the animacy of the stimulus with non-agentive verbs. Taking into account that the overall frequency of non-active voice is highest with this verb group, it may be the case that this general preference overrides further distinctions in some parts of the dataset.

A final question is the interaction between choice of subject and choice of word order. The word order properties of the active clauses were discussed in section 6. To the extent that the choice of subject is an alternative strategy for fronting the lower argument, non-canonical word order is not expected to occur with non-active voice. However, the complementarity of voice and order is limited, since voice alternations may be accompanied with semantic differences, as discussed in section 2.2, i.e., the choice between active and non-active voice as alternative linearization options only holds for a subset of the possible situations.

The proportions of orders depending on voice are presented in Figure 4 (see counts in Appendix D). These frequencies reveal an interaction between verb classes and voice. With canonical verbs, the proportion of non-canonical order in passive voice (oblique agent preceding subject) is higher than the corresponding proportion of non-canonical order (object preceding subject) in active voice. With EO verbs, the proportion of non-canonical order with non-active verb forms (non-subject preceding subject) is lower than the proportion of non-canonical order (object preceding subject) in active voice. A *generalized logit mixed model* reveals that this interaction is highly significant (log-likelihood test, $\chi^2(2) = 63.7$; $p < .001$, comparing a model containing the interaction with a model without the interaction effect); furthermore, the contrast between the three verb classes is not reducible to a simpler model which only contrasts canonical and EO verbs (log-likelihood test, $\chi^2(2) = 7.9$; $p < .05$).

Figure 4. Voice and word order



The interaction in Figure 4 can only be interpreted if we consider the fact that active and non-active versions are not just linearization options. If this were the case, we would expect that fronting the non-subject constituent would not appear in non-

active voice, since in this case the thematic roles are linearized in the same way as in the unmarked expressions with active voice and canonical order. If we take into account, that the choice of voice is – at least in part – determined by semantic factors that are independent from linearization (see section 2.2), the difference between canonical and EO verbs in Figure 4 is just a further confirmation of the phenomenon observed in the word order facts of active clauses (section 6): canonical verbs display a preference to front agent constituents, which is reflected in the higher proportion of non-canonical order in the passive voice; EO verbs display a preference to front the experiencer, which is reflected in the higher proportion of non-canonical order in the active voice.

8. General discussion

The corpus facts presented in the previous sections generally gave affirmative answers to the questions outlined in section 3. Section 5 has shown that experiencer-object verbs differ from canonical verbs with respect to the distribution of referential third person pronouns. While these pronouns occur more frequently as subjects with canonical verbs, they occur more frequently as objects with EO verbs. This finding answers the question in (4a), i.e., it provides evidence that experiencer objects are likely discourse topics (in contrast to objects of canonical transitive verbs). Section 6 has demonstrated with evidence from word order frequencies that experiencer-first effects are present in the corpus and that they are independent from other related effects such as animate-first or given-first. A significant interaction between VERB CLASS and ANIMACY provides statistical evidence for the research question (4b) concerning the independency of the experiencer-first effect: if the thematic properties of experiencers were reducible to animacy, it should be possible to explain the obtained word order frequencies with an effect of ANIMACY alone. A direct inspection

of the descriptive data suggests that a part of this interaction is due to the contrast between accusative and dative EO verbs: while the effect of EO accusative verbs on non-canonical word order can be reduced to an animacy effect, this is not the case for EO dative verbs. A further part of this interaction relates to the difference between accusative EO verbs and canonical verbs: while disharmonic animacy induces marked word order in the former class of verbs, it does not have the same effect with the latter. Finally, section 7 has shown that disharmonic animacy has an impact on the choice of subject (and the corresponding diathetic form of the verb) and that this impact is different for EO non-agentive verbs in contrast to canonical verbs and EO \pm agentive verbs.

The major influence in the choice of order and the choice of subject comes from animacy asymmetries. The role of animacy in German syntax has been demonstrated in several experimental and corpus studies (e.g., van Nice & Dietrich 2003, Grewe et al. 2006, Bader & Häussler 2010). The facts of the present study give further support to these previous findings. Bader and Häussler (2010) show that disharmonic animacy correlates with a higher frequency of OS order. They further report a difference between accusative and dative objects, such that OS order is generally more frequent with the latter type of object. The results presented in section 6 are in line with these observations. In both studies, dative verbs behave similarly, showing a clear OS preference in middlefield sentences in the disharmonic animacy condition while OS order in the other animacy constellations is less frequent, but clearly noticeable. In the dative prefield data, the amount of OS sentences is smaller though still considerable – however, in contrast to our data, which do not show an animacy effect for dative verbs in the prefield, the prefield dative data in Bader and Häussler (2010) still shows such an effect. The accusative verbs in Bader and Häussler (2010) demonstrate an overall

clear preference for SO structures, which is similar to that of the canonical transitive verbs in our corpus. However, in contrast to our data, there is a (small) animacy effect with transitive verbs in both the middlefield and the prefield data in Bader and Häussler (2010). The fact that there is no object preposing with canonical transitive verbs in our middlefield data may be due to the fact that the overall frequency of sentences with disharmonic animacy was rather low (9 sentences). Furthermore, as Bader and Häussler (2010:757) underline, preposing of lexical accusative objects in the middlefield occurs mainly due to lexical-semantic reasons, which are visible with the EO verbs in our data. Very rarely does it occur for the discourse reason of focussing the subject in preverbal position (Haider 1993).

Definiteness does not have a significant effect in our data. This is in line with the fact that effects of definiteness are generally weaker – also in previous corpus studies on German word order. Weber and Müller (2004) report that clauses with indefinite subjects and definite objects occur in 8.5% of their SVO data (50 out of 591 sentences) and in 8.7% of their OVS data (48 out of 549 sentences). In Bader & Häussler's study (2010:739), definiteness plays a role with dative verbs in the middlefield, but is not considered to be a trigger of object-fronting with accusative verbs. Similarly, the descriptive facts of the present study suggest no definiteness effect: in the overall middlefield results, the OS order occurs in 41.9% of the disharmonic definiteness data and in 37.8% of the other cases (see totals in Appendix B.3). The difference in the prefield data is smaller and in the opposite direction.¹²

¹² The same difference of direction is also visible in Bader & Häussler's data (2010:728, Table 5), although the absolute results for dative and accusative verbs differ considerably. This is certainly due to the difference in the composition of the data sets: in contrast to the data set investigated in the present study, Bader and Häussler's data set is not restricted to certain verb groups.

19.9% for the disharmonic data, 20.5% for the other data. However, since DEFINITENESS does not have a significant effect in the statistical model, the validity of these observations is not empirically justified.

The difference between prefield and middlefield is established in German syntax. Following the findings from earlier corpus studies such as Bader and Häussler (2010), asymmetries in the semantic and thematic representation of the arguments are expected to have a straightforward impact in the middlefield while prefield constituents should be additionally influenced by further factors – in particular information structure. The word-order findings of the present study confirm these expectations: both clausal domains show the same general tendencies triggered by animacy and thematic asymmetries, but the effect size in the middlefield is larger. The comparison between clausal domains also involved an unexpected result in section 7: the frequency of non-active voice EO non-agentive verbs depends on animacy in the middlefield, but not so in the prefield (see discussion in section 7).

The main issue in the three substudies presented in this article was to identify and explain differences in the structural behavior between verb classes. Statistical analyses started with a comparison between the VERB-CLASS MODELS that were hypothesized to account for the obtained data. The models with the maximal goodness of fit are summarized in Table 8. Discourse prominence, as reflected in the frequency of pronominal realization (see Gundel et al. 1993), reveals a contrast between canonical and experiencer-object verbs and an additional influence of agentivity in the active voice. Order frequencies indicate a difference between canonical and experiencer-object verbs and an additional role of case-marking (accusative vs. dative). Frequencies of non-active forms are accounted for through the contrast between non-agentive verbs and the other verb classes.

Table 8. Verb-class models with maximal goodness-of-fit

	canonical		experiencer-object			
			accusative		dative	
			\pm agentive	non-agentive		
D-PROMINENCE (active)	α	<	β	<	γ	γ
D-PROMINENCE (non-active)	α	<	β		β	–
OS ORDER	α	<	β		β	< γ
NON-ACTIVE VOICE	α		α	<	β	–

The findings in Table 8 show that verb classes display different contrasts depending on the phenomenon at issue. The data from discourse prominence imply that subjects of active or non-active voice generally refer to highly activated referents, but experiencer-object verbs show the mirror image in the active voice. The discourse prominence data in the active voice involve an additional effect of agentivity, probably reflecting the fact that at least in a subset of their occurrences \pm agentive verbs behave like canonical verbs. Agentivity is also crucial for the choice of voice. Non-agentive verbs generally display a higher proportion of non-active forms, which overrides animacy contrasts in the prefield. The word order facts differ in that the crucial distinction does not relate to agentivity but to case-marking (accusative vs. dative). This finding cannot be accounted for through the greater likelihood of ambiguity between nominative and accusative NPs in German (see Table 6).

The word order effects involve two contrasts: canonical verbs and EO accusative verbs differ in that only the latter group comes with object-first order in disharmonic animacy configurations; EO accusative and EO dative verbs differ in that the object-first order occurs independently from animacy with the latter group. This is

in line with the view that the dative-before-nominative order is basic with EO dative verbs (e.g., Lenerz 1977, Primus 1996, Fanselow 2001, 2003, Haider & Rosengren 2003, Haider 2010) while the accusative-before-nominative order has to be triggered (by animacy in this data). The data from the choice of voice show that the early realization of experiencers is most frequent with non-agentive verbs, i.e., there is a general preference to realize the experiencer-argument early in the clause if it does not compete with an agent. This preference is alternatively fulfilled by the choice of voice or the choice of order with the EO accusative verbs. The complementarity between order and voice is empirically supported by the results in Figure 4 for EO accusative verbs. EO dative verbs do not provide a non-active option, which probably accounts for the difference in the frequencies of object-fronting between non-agentive accusative and dative verbs in our corpus (Lamers & de Hoop *forthc*; Lamers & de Schepper 2010).

Summing up, we assume that animate-first effects directly refer to the linearization of the arguments in the utterance (and not strictly to the constituent order). Such effects are expected to have an impact on the choice of subject as well as on the choice of word order. This intuition is confirmed by the fact that animacy-first effects are reflected in the choice of non-active voice with all classes (whereas the exceptional behaviour of the non-agentive verbs in the prefield is an additional phenomenon that is not captured by our assumptions). It is crucial that these effects also apply to canonical verbs, i.e., animacy-first effects do not depend on verb class.

Experiencer-first effects reflect the discourse prominence of experiencers, which applies to all experiencer-objects classes, as shown by the findings in the frequency of pronouns. The word order facts show a significant difference between initial patients and initial experiencers, which implies that an experiencer-first effect

on word order exists (which is part of the main effect of VERB CLASS and the interaction effect between VERB CLASS and ANIMACY). There is an additional empirical finding distinguishing accusative from dative experiencers: initial experiencers are very frequent across animacy configurations with the latter group, but almost restricted to disharmonic configurations with the former. The fact that this difference relates to case suggests that the source of the difference does not lie in the semantics of the different experiencer-object subclasses. The fact that the experiencer-first order is very rare with non-disharmonic configurations (which practically contain two animates) for accusative experiencer-object verbs possibly reflects a blocking effect: the accusative-nominative order is difficult to process in the absence of lexico-semantic cues (animacy asymmetry) for the disambiguation of thematic roles.

Finally, we will address the question of what consequences the empirical results of the present study have for assumptions about the basic order of EO verbs as a pure reflex of constituent structure. Assumptions about constituent structure are based on a large array of phenomena that are not dealt with in the present study (in particular, evidence for asymmetries relating to the hierarchical clause structure, such as binding or scopal phenomena). Corpus frequencies are relevant for testing the predictions of syntactic analyses of this type. Precisely, a word order that reflects the basic configuration in the syntax does not need to be licensed by a trigger whereas an order involving some reordering operation is expected to occur in a lexico-semantic or contextual configuration in which this operation is licensed.

A theory assuming that the basic order of dative-object verbs involves a higher experiencer that appears early in the utterance without a contextual trigger is confirmed in our data by the fact that dative-object verbs are most frequently experiencer-initial (across animacy levels). This theory must accommodate the fact

that experiencers also frequently appear early in sentences with accusative experiencer-object verbs, but not so with canonical verbs (which means that this latter effect cannot be accounted for by animacy alone). A theory assuming that the basic order of all experiencer-object verbs involves an initial experiencer must account for the fact that this generalization does not fit to the data of accusative experiencer-object verbs in non-disharmonic animacy configurations. The data pattern obtained in these configurations is predicted by the additional assumption of blocking due to case reasons.

9. Conclusion

This article presents a corpus study on experiencer-object verbs in German. Based on the frequencies of pronominal realization, it provides evidence that the objects of these verbs are discourse prominent. Word order frequencies reveal that the effect of experiencer objects interacts with animacy. Dative EO verbs occur in the OS order independent of animacy, and accusative EO verbs occur in the OS order mostly if this order is licensed by animacy. Both classes contrast with canonical verbs in which the OS order appears less frequently under identical conditions. Animacy effects also appear in the choice of voice. However, voice frequencies do not distinguish EO verbs from canonical verbs, but rather non-agentive from potentially agentive verbs.

These findings contribute to the previous research on the linearization properties of experiencer-object verbs in German. Previous accounts based on acceptability judgments already observed the difference between these classes of experiencer-object verbs. While the dative-first order is analyzed as basic for dative EO verbs, the accusative-first order has been a source of controversies. The results of the present corpus study support the intuition that these classes behave differently. Assuming that

corpus frequencies are informative for the necessity of licensing non-canonical word order, this study supports accounts restricting the exceptional behaviour to dative verbs (e.g., Fanselow 2000).

The reported verb class effects are relevant for the research on word order frequencies in corpora. The SO/OS frequencies in German text corpora have been the subject of numerous corpus studies. The present study shows that the role of verb classes on word order is not negligible. This result has important methodological consequences: word order generalizations must be tested against the impact of different verb classes and the role of individual verbs is at least a very reasonable random factor for word order studies.

Acknowledgements

Special thanks are due to Birgit Jänen and Julian A. Rott for their contribution to data annotation and to Nico Lehmann for proofreading. We are also very grateful to Amir Zeldes, Felix Golcher, Tom Ruetten, and an anonymous reviewer for helpful discussion and comments on the methodological and theoretical aspects of this article. Parts of this study were presented at the University of Vienna (June 2014) and the following conferences: *Quantitative Approaches in Corpus Linguistics and Psycholinguistics*, INALCO, CNRS, Paris, October 2013; *International Colloquium on Referentiality*, University of Erfurt, February 2014. We thank the audiences for very helpful comments and discussions. This article is part of the project VE 570/1 *Syntaktische und pragmatische Prominenz von Experiencern im Sprachvergleich*, sponsored by the German Research Foundation.

References

- Abbott, Barbara. 2008. Issues in the semantics and pragmatics of definite descriptions in English. In Jeanette K. Gundel & Nancy Hedberg (eds.), *Reference. Interdisciplinary Perspectives*, 61-72. Oxford: OUP.
- Aissen, Judith. 1999. Markedness and subject choice in optimality theory. *Natural Language and Linguistic Theory* 17. 673-711.
- Arad, Maya. 1998a. *VP-Structure and the syntax-lexicon interface*. London: University College of London doctoral dissertation.
- Arad, Maya. 1998b. Psych-notes. *UCL Working Papers in Linguistics* 10. 203-223.
- Bader, Markus & Jana Häussler. 2010. Word order in German: A corpus study. *Lingua* 120. 717-762.
- Bates, Doug, Maechler, Martin, and Bolker, Ben. 2011. lme4: Linear Mixed-Effects Models Using S4 Classes. R package version 0.999375-42.
- Bayer, Joseph. 2004. Non-nominative subjects in comparison. In Peri Bhaskararao & Karumuri Venkata Subbarao (eds.), *Non-nominative subjects*, vol. 1, 49-76. Amsterdam: Benjamins.
- Belletti, Adriana & Luigi Rizzi. 1988. Psych-verbs and θ -theory. *Natural Language and Linguistic Theory* 6. 291-352.
- Bickel, Balthasar. 2004. The syntax of experiencers in the Himalayas. In Peri Bhaskararao & Karumuri Venkata Subbarao (eds.), *Non-nominative subjects*, vol. 1, 77-111. Amsterdam: Benjamins.
- Bock, Kathryn J. & Richard K. Warren. 1985. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition* 21. 47-67.

- Bornkessel, Ina, Stefan Zysset, Angela D. Friederici, D. Yves von Cramon & Matthias Schlesewsky. 2005. Who does what to whom? The neural basis of argument hierarchies during language comprehension. *NeuroImage* 26. 221-233.
- Bornkessel, Ina. 2002. *The argument dependency model: A neurocognitive approach to incremental interpretation*. Leipzig: MPI-Series in Cognitive Neuroscience.
- Bornkessel-Schlesewsky, Ina & Matthias Schlesewsky. 2009. The role of prominence information in real time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass* 3(1). 19-58.
- Bouchard, Denis. 1995. *The semantics of syntax: A minimalist approach to grammar*. Chicago: The University of Chicago Press.
- Branigan, Holly P. & Elina Feleki. 1999. Conceptual accessibility and serial order in Greek language production. In Martin Hahn & Scott C. Stoness (eds.), *Proceedings of the 21st Conference of the Cognitive Science Society*, 96-101. Mahwah: Erlbaum.
- Bresnan, Joan and Jennifer Hay. 2008. Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua* 118(2). 245-259.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Boume, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69-94. Amsterdam: Royal Netherlands Academy of Science.
- Clark, Herbert & Susan Haviland. 1977. Comprehension and the given-new contrast. In Roy O. Freedle (ed.), *Discourse production and comprehension*, 1-40. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Comrie, Bernard. 1981. *Language Universals and Linguistic Typology*. Blackwell, Oxford.
- Cupples, Linda. 2002. The structural characteristics and on-line comprehension of experiencer-verb sentences. *Language and Cognitive Processes* 17(2). 125-162.
- Dahl, Östen & Kari Fraurud. 1996. Animacy in grammar and discourse. In Thorsten Fretheim, & Jeanette K. Gundel (eds.), *Reference and referent accessibility*, 47-64. Amsterdam & Philadelphia: Benjamins.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67. 547-691.
- Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 63(4). 805-855.
- Engelberg, Stefan. 2014. The argument structure of psych-verbs: A quantitative corpus study on cognitive entrenchment [to appear]. In Hans Boas & Alexander Ziem (eds.), *Constructional approaches to argument structure in German*. Berlin: De Gruyter.
- Fanselow, Gisbert. 2000. Optimal exceptions. In Barbara Stiebels & Dieter Wunderlich (eds.), *Lexicon in Focus*, 173-209. Berlin: Akademie Verlag.
- Fanselow, Gisbert. 2001. Features, θ -roles, and free constituent order. *Linguistic Inquiry* 32(3). 405-437.
- Fanselow, Gisbert. 2003. Zur Generierung der Abfolge der Satzglieder im Deutschen. *Neue Beiträge zur Germanistik* 112. 3-47.
- Ferreira, Fernanda. 1994. Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language* 33. 715-736.
- Foley, William A., and Robert D. Van Valin, Jr. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

- Fraurud, Kari. 1990. Definiteness and the processing of NPs in natural discourse. *Journal of Semantics* 7. 395-433.
- Frey, Werner. 2006. Contrast and movement to the German prefield. In Molnár, Valéria & Susanne Winkler (eds.), *The architecture of focus*. 235-264. Berlin: de Gruyter.
- Givón, Talmy (ed.). 1994. *Voice and inversion*. Amsterdam/Philadelphia: Benjamins.
- Givón, Talmy. 1994. The pragmatics of de-transitive voice: functional and typological aspects of inversion. In Talmy Givón (ed.), *Voice and Inversion*. 3-44. Amsterdam: Benjamins.
- Gordon, Peter C., Barbara J. Grosz & Laura A. Lilliom. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science* 17(3). 311-347.
- Grewe, Tanja, Bornkessel, Ina, Zysset, Stefan, Wiese, Richard, Cramon, Yves, D. von and Schlesewsky, Matthias. 2006. Linguistic prominence and Broca's area: The influence of animacy as a linearization principle. *NeuroImage* 32. 1395-1402.
- Grewendorf, Günther. 1989. *Ergativity in German*. Foris, Dordrecht.
- Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: MIT Press.
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274-307.
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 2005. Pronouns without explicit referents: How do we know when a pronoun is referential? In Antonio Branco, Tony McEnery & Ruslan Mitkov (eds.), *Anaphora Processing: Linguistic, Cognitive, and Computational Modelling*, 351-364. Amsterdam: Benjamins.
- Haider, Hubert & Inger Rosengren. 2003. Scrambling: Nontriggered chain formation in OV languages. *Journal of Germanic Linguistics* 15(3). 203-267.

- Haider, Hubert. 1993. *Deutsche Syntax - generativ: Vorstudien zur Theorie einer projektiven Grammatik*. Tübingen: Narr.
- Haider, Hubert. 2010. *The syntax of German*. Cambridge: Cambridge University Press.
- Haspelmath, Martin. 2001. Non-canonical marking of core arguments in European languages. In: Alexandra Y. Aikhenvald, R.M.W. Dixon & Masayuki Onishi (eds.), *Non-canonical marking of subjects and objects*, 53–83. Amsterdam and Philadelphia: Benjamins.
- Haupt, Friederike S., Matthias Schlesewsky, Dietmar Roehm, Angela D. Friederici & Ina Bornkessel-Schlesewsky. 2008. The status of subject-object reanalyses in language comprehension architecture. *Journal of Memory and Language* 59. 54-96.
- Heck, Fabian. 2000. Tiefenoptimierung: Deutsche Wortstellung als wettbewerbsgesteuerte Basisgenerierung. *Linguistische Berichte* 184. 441-468.
- Heylen, Kris & Dirk Speelman. 2003. A corpus-based analysis of word order variation: The order of verb arguments in the German middle field. In Dawn Archer, Paul Rayson, Andrew Wilson and Tony McEnery, *Proceedings of the Corpus Linguistics 2003 Conference*, 320-329. Lancaster: University Centre for Computer Corpus Research on Language.
- Heylen, Kris. 2005. A quantitative corpus study of German word order. In:
- Hoberg, Ursula. 1981. *Die Wortstellung in der geschriebenen deutschen Gegenwartssprache*. München: Hueber.
- Ichihashi-Nakayama, Kumiko. 1994. On dative ‘subject’ constructions in Nepali. In Carol Genetti (ed.), *Aspects of Nepali Grammar*. Santa Barbara Papers in Linguistics, Vol. 6, 41-76. California: University of California

- Kempen, Gerard & Karin Harbusch. 2004. A corpus study into word order variation in German subordinate clauses: Animacy affects linearization independently of grammatical function assignment. In Thomas Pechmann & Christopher Habel (eds.), *Multidisciplinary approaches to language production*, 173-181. Berlin: Mouton de Gruyter.
- Kempen, Gerard & Karin Harbusch. 2005. The relationship between grammaticality ratings and corpus frequencies: A case study into word order variability in the midfield of German clauses. In Stephan Kepser & Marga Reis. (eds.), *Linguistic Evidence*. 329-349. Berlin and New York: Mouton de Gruyter.
- Klein, Katharina & Silvia Kutscher. 2002. Psych-verbs and lexical economy. *Theorie des Lexikons* 122. Düsseldorf: University of Düsseldorf.
- Lambrecht, Knud. 2000. When subjects behave like objects: an analysis of the merging of S and O in sentence-focus constructions across languages. *Studies in language* 24(3). 611-682.
- Lamers, Monique & Helen de Hoop. forthc. Object fronting in Dutch. In Brian MacWhinney, Edith Moravcsik & Andrej Malchukov (eds.), *Competing motivations in grammar and usage*. Oxford: Oxford University Press.
- Lamers, Monique & Kees de Schepper. 2010. Argument linearization in Dutch and German: a multifactorial analysis. Paper read at the Conference on Competing Motivations, Leipzig, November 24, 2010.
- Lamers, Monique. 2007. Verb type, animacy and definiteness in grammatical function disambiguation. In Bettelou Los & Marjo van Koppen (eds.), *Linguistics in the Netherlands 2007*, 125-137. Amsterdam: Benjamins.
- Landau, Idan. 2010. *The Locative Syntax of Experiencers*. Cambridge, MA: MIT Press.

- Lenerz, Jürgen. 1977. Zum Einfluß von "Agens" auf die Wortstellung des Deutschen. In Hans Werner Viethen, Wolf-Dietrich Bald & Konrad Sprengel (eds.), *Grammatik und interdisziplinäre Bereiche der Linguistik. Akten des 11. Linguistischen Kolloquiums Aachen 1976*, 133-142. Tübingen: Niemeyer.
- Levin, Beth & Jason Grafmiller. 2013. Do you always fear what frightens you? In Tracy Holloway King & Valeria de Paiva (eds.), *From Quirky Case to Representing Space: Papers in Honour of Annie Zaenen*, 21-32. Stanford, CA: CSLI Publications.
- Mathesius, Vilém. 1975. *A functional analysis of present day English on a general linguistic basis* [ed. by J. Vachek]. The Hague, Paris: Mouton.
- Müller, Gereon. 1999. Optimality, markedness, and word order in German. *Linguistics* 37. 777-818.
- Nice, Kathy Y. van & Rainer Dietrich. 2003. Animacy effects in language production: From mental model to formulator. In: Holden Härtl & Heike Tappe (eds.), *Mediating between Concepts and Grammar*, 101-117. Berlin: de Gruyter.
- Pesetsky, David. 1995. *Zero syntax: Experiencer and cascades*. Cambridge, MA: MIT Press.
- Piñango, Maria Mercedes. 2000. Canonicity in Broca's sentence comprehension: The case of psychological verbs. In Yosef Grodzinsky, Lewis Shapiro & David Swinney (eds.), *Language and the brain. Representation and processing*, 327-350. San Diego, CA: Academic Press.
- Poesio, Massimo. 2008. Linguistic claims formulated in terms of centering: a re-examination using parametric CB-tracking techniques. In Jeanette K. Gundel & Nancy Hedberg (eds.), *Reference: interdisciplinary perspectives*, 216-245. Oxford: University Press.

- Prat-Sala, Mercè & Holly P. Branigan. 2000. Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language* 42. 168-182.
- Prat-Sala, Mercè, Richard Shillcock & Antonella Sorace. 2000. Animacy effects on the production of object-dislocated descriptions by Catalan-speaking children. *Journal of Child Language* 27. 97-117.
- Prat-Sala, Mercè. 1997. The production of different word orders: A psycholinguistic and developmental approach. University of Edinburgh: Ph.D. dissertation.
- Primus, Beatrice. 1994. Grammatik und Performanz: Faktoren der Wortstellungsvariation im Mittelfeld. *Sprache und Pragmatik* 32. 39-86.
- Primus, Beatrice. 1996. Dependenz und Serialisierung: das Deutsche im Sprachvergleich. In Ewald Lang & Gisela Zifonun (eds.), *Deutsch – typologisch. Jahrbuch des Instituts für Deutsche Sprache 1995*, 57-91. Berlin: de Gruyter.
- Primus, Beatrice. 2004. Protollen und Verbtyp: Kasusvariation bei psychischen Verben. In Rolf Kailuweit & Martin Hummel (eds.), *Semantische Rollen*, 377-401. Tübingen: Narr.
- Reinhart, Tanya. 2001. Experiencing derivations. New York: SALT lecture.
- Richards, Norvin. 2010. *Uttering trees*. Cambridge, Mass.: MIT Press.
- Scheepers, Christoph, Barbara Hemforth & Lars Konieczny. 2000. Linking syntactic functions with thematic roles: psych verbs and the resolution of subject-object ambiguity. In Barbara Hemforth & Lars Konieczny (eds.), *German sentence processing*, 95-135. Dordrecht: Kluwer Academic Publishers.
- Scheepers, Christoph. 1997. Menschliche Satzverarbeitung. Syntaktische und thematische Aspekte der Wortstellung im Deutschen. Universität Freiburg: Dissertation.

- Skopeteas, Stavros & Gisbert Fanselow. 2009. Effects of givenness and constraints on free word order. In Malte Zimmerman & Caroline Féry (eds.), *Information structure: Theoretical, typological, and experimental perspectives*, 307-331. Oxford University Press, Oxford.
- Speyer, Augustin. 2007. Die Bedeutung der Centering Theory für Fragen der Vorfeldbesetzung im Deutschen. *Zeitschrift für Sprachwissenschaft* 26. 83-115.
- Stephan Kepser & Marga Reis (eds.), *Linguistic Evidence*, 241-263. Berlin and New York: Mouton de Gruyter.
- Tomlin, Russell S. 1995. Focal attention, voice, and word order: An experimental, cross-linguistic study. In Pamela Downing & Michael P. Noonan (eds.), *Word order in discourse*, 517-554. Amsterdam and Philadelphia: Benjamins.
- Ueno, Mieko & Maria Polinsky. 2009. Does headedness affect processing? A new look at the OV-VO contrast. *Journal of Linguistics* 45. 675-710.
- Verhoeven, Elisabeth. 2009. Subjects, agents, experiencers, and animates in competition: Modern Greek argument order. *Linguistische Berichte* 219:355-376.
- Walker, Marilyn A., Aravind K. Joshi & Ellen F. Prince. 1998. Centering in naturally occurring discourse: an overview. In Marilyn A. Walker, Aravind K. Joshi & Ellen F. Prince (eds.), *Centering Theory in Discourse*, 1-28. Oxford: Oxford University Press.
- Webelhuth, Gerd. 1995. German is configurational. *The Linguistic Review* 4. 203-246.
- Weber, Andrea & Karin Müller. 2004. Word order variation in German main clauses: A corpus analysis. *Proceedings of the 20th International Conference on Computational Linguistics*.

Wegener, Heide. 1998. Der Kasus des EXP. In Marcel Vuillaume (ed.), *Die Kasus im Deutschen. Form und Inhalt*, 71-84. Tübingen: Stauffenburg Verlag.

Wunderlich, Dieter. 1997. Cause and the structure of verbs. *Linguistic Inquiry* 28. 27-68.

Zifonun, Gisela. 1992. Das Passiv im Deutschen: Agenten, Blockaden und (De-) Gradierungen. In: Ludger Hoffmann (ed.), *Deutsche Syntax*, 250-275. Berlin: de Gruyter.

Appendix A: Frequency of pronouns

A.1 Referential pronouns out of total third person NPs in active clauses

	subject				object			
	pronoun	other	total	% pron.	pronoun	other	total	% pron.
CANONICAL	120	837	957	12.5	42	925	967	4.3
EO.ACC ±ag	74	1,006	1,080	6.9	187	664	851	22.0
EO.ACC –ag	45	1,081	1,126	4.0	243	550	793	30.6
EO.DAT	101	2,688	2,789	3.6	637	1,503	2,140	29.8
total	340	5,612	5,952	5.7	1,109	3,642	4,751	23.3

A.2 Referential pronouns out of total third person NPs in non-active clauses

	subject				non-subject			
	pronoun	other	total	% pron.	pronoun	other	total	% pron.
CANONICAL	24	226	251	10.0	1	255	256	0.4
EO.ACC ±ag	119	548	667	17.8	6	845	851	0.7
EO.ACC –ag	231	1,045	1,276	18.1	22	1,532	1,554	1.4
total	375	1,819	2,194	17.1	29	2,632	2,661	1.1

Appendix B: Word order frequencies (active clauses)

B1. Animacy and verb class (middlefield)

	disharmonic animacy				other			
	SO	OS	total	% OS	SO	OS	total	% OS
CANONICAL	9	0	9	0.0	138	0	138	0.0
EO.ACC ±ag	20	26	46	56.5	53	2	55	3.6
EO.ACC –ag	12	33	45	73.3	12	1	13	7.7
EO.DAT	27	105	132	79.5	7	7	14	50.0
total	68	164	232	70.7	210	10	220	4.5

B2. Animacy and verb class (prefield)

	disharmonic animacy				other			
	SO	OS	total	% OS	SO	OS	total	% OS
CANONICAL	43	1	44	2.3	551	25	576	4.3
EO.ACC ±ag	248	62	310	20.0	111	5	116	4.3
EO.ACC –ag	259	67	326	20.6	34	0	34	0.0
EO.DAT	649	316	965	32.7	33	20	53	37.7
total	1,199	446	1,645	27.1	729	50	779	6.4

B3. Definiteness and verb class (middlefield)

	disharmonic definiteness				other			
	SO	OS	total	% OS	SO	OS	total	% OS
CANONICAL	26	0	26	0.0	121	0	121	0.0
EO.ACC ±ag	9	7	16	43.8	64	21	85	24.7
EO.ACC –ag	4	5	9	55.6	20	29	49	59.2
EO.DAT	4	19	23	82.6	30	93	123	75.6
total	43	31	74	41.9	235	143	378	37.8

B4. Definiteness and verb class (prefield)

	disharmonic definiteness				other			
	SO	OS	Total	% OS	SO	OS	total	% OS
CANONICAL	121	1	122	0.8	473	25	498	5.0
EO.ACC ±ag	48	9	57	15.8	311	58	369	15.7
EO.ACC –ag	56	10	66	15.2	237	57	294	19.4

EO.DAT	48	48	96	50.0	634	288	922	31.2
total	273	68	341	19.9	1,655	428	2,083	20.5

Appendix C: Frequency of the choice of undergoers as subjects

C1. Animacy and verb class (middlefield)

	disharmonic animacy				other			
	active	–active	total	% –act.	active	–active	total	% –act.
CANONICAL	9	31	40	77.5	138	41	179	22.9
EO.ACC ±ag	46	107	153	69.9	55	22	77	28.6
EO.ACC –ag	45	129	174	74.1	13	10	23	43.5
total	100	267	367	72.8	206	73	279	26.2

C2. Animacy and verb class (prefield)

	disharmonic animacy				other			
	active	–active	total	% –act.	active	–active	total	% –act.
CANONICAL	44	57	101	56.4	576	94	670	14.0
EO.ACC ±ag	310	276	586	47.1	116	42	158	26.6
EO.ACC –ag	326	548	874	62.7	34	86	120	71.7
total	680	881	1,561	56.4	726	222	948	23.4

C3. Definiteness and verb class (middlefield)

	disharmonic definiteness				other			
	active	–active	total	% –act.	active	–active	total	% –act.
CANONICAL	26	13	39	33.3	121	59	180	32.8
EO.ACC ±ag	16	14	30	46.7	85	115	200	57.5
EO.ACC –ag	9	25	34	73.5	49	114	163	69.9
total	51	52	103	50.5	255	288	543	53.0

C4. Definiteness and verb class (prefield)

	disharmonic definiteness				other			
	active	–active	total	% –act.	active	–active	total	% –act.
CANONICAL	122	33	155	21.3	498	118	616	19.2
EO.ACC ±ag	57	35	92	38.0	369	283	652	43.4
EO.ACC –ag	66	128	194	66.0	294	506	800	63.3
total	245	196	441	44.4	1,161	907	2,068	43.9

Appendix D: Frequency of word orders per voice

(NS= non-subject precedes subject; SN= subject precedes non-subject)

	active				non-active			
	NS	SN	total	% NS	NS	SN	total	% NS
CANONICAL	26	741	767	3.6	26	197	223	11.7
EO.ACC ±ag	95	432	527	18.0	39	408	447	8.7
EO.ACC –ag	101	317	418	24.2	67	706	773	8.7
total	222	1490	1712	13.0	132	1311	1443	9.1