Scales or features in verb meaning?

Verb classes as predictors of syntactic behavior

Elisabeth Verhoeven (Humboldt-Universität zu Berlin)

Abstract

Several syntactic properties of verbal heads are accounted for through their semantic properties. Verbal features such as agentivity, volitionality, stativity etc. have been proven a useful tool for predicting several aspects of their syntactic behavior such as passivization, auxiliary selection etc. In the context of the empirical turn in current linguistics, the assumption of discrete features is questioned by studies based on corpora or speakers' intuitions showing that the diagnostics of semantic features involve gradience. These findings are challenging for grammatical theory: are we justified to assume the existence of discrete verb classes or do the established properties indicate scalar dimensions of meaning? Based on two empirical studies – an acceptability study and a corpus study – the present article examines the role of *agentivity* in distinguishing verb classes and in predicting the syntactic behavior of verbs in German. Acceptability data show that the diagnostics of agentivity involve gradience, which cannot be reduced to random sources of variation. However, a comparison of scalar vs. categorical models of agentivity based on these diagnostics reveals that the syntactic variation in word order found in written corpus data is best accounted for through a model that assumes a binary division into a ±agentive and a non-agentive verb class.[1]

Keywords

acceptability judgments, corpus data, agentivity, OS order, non-active voice, psych verbs

1 Introduction

Properties of verbal meaning account for a large array of syntactic phenomena, such as auxiliary selection, impersonal passivization (Keller and Sorace 2003), locative inversion (Levin and Rappaport Hovav 1995), reflexivization and the causative alternation (e.g. Reinhart and Siloni 2005, Horvath and Siloni 2011) etc. Assuming that particular semantic features associated with subsets of verbs account for their syntactic behavior constitutes a valuable approach explaining the way lexical properties determine syntactic variation. A particular challenge arises when either the lexical classifications or the syntactic phenomena at issue involve gradience. Gradience at the lexical level is a source of *indeterminacy* in establishing boundaries between verb classes. For instance, certain readings may not be categorically excluded for a class of verbs but can be imposed on the verb meaning by an appropriate context. Certain lexico-semantic properties may be less stable than others (see Keller and Sorace 2003 on the lability of aspectual properties and argument role that cause gradience in auxiliary selection and impersonal passivization). Furthermore, gradience may arise from the fact that syntactic properties are determined by the interplay of various lexical factors, whose mutual interactions are indeterminate (see Sorace 2004 on auxiliary selection). The following two questions are crucial for our theoretical assumptions about the lexicon-syntax interface.

(1) a. Are the relevant semantic properties in the verbal lexicon features or scales?
b. Is the gradience of particular properties of meaning grammatically relevant or can we achieve equally adequate descriptions by abstracting away from it?

The present article examines the role of agentivity, a semantic property, that is known to be essential for the verbal syntax. Agentivity (or lack thereof) plays an important role

in understanding the syntactic behavior of psych verbs. Non-agentive experiencer-object verbs such as *concern, depress,* etc. show particular syntactic properties in many languages (e.g., Belletti and Rizzi 1988, Pesetsky 1995, Arad 1998, Landau 2010, Verhoeven 2014). These properties include peculiarities in nominalization, reflexivization, passivization, extraction, binding, and argument linearization, among others (for German see Grewendorf 1989, Wunderlich 1997, Fanselow 2000, Haspelmath 2001, Klein and Kutscher 2002, Bayer 2004, Verhoeven 2015, Temme and Verhoeven 2016). For the purposes of the present study, we focus on two properties of German experiencer-object verbs, namely properties of argument linearization and properties of argument alternation, which have been shown to interact with (non-)agentivity.

Several empirical studies presenting evidence from controlled and spontaneous language production and intuition (acceptability) have shown that accusative experiencer-object verbs display linearization preferences different from canonical transitive verbs. It hast, for instance, been shown that both argument orderings ($S_{STIM} \prec O_{EXP}$ and $O_{EXP} \prec S_{STIM}$) reach the same acceptability level with experiencer verbs in contrast to canonical verbs (Temme and Verhoeven 2016, Haupt et al. 2008, 84, confirming earlier observations from Lenerz 1977, Hoberg 1981, Primus 2004). A large-scale corpus study on the argument realization of experiencer-object verbs demonstrated that the OS order is more frequent with experiencer verbs than with canonical verbs (Verhoeven 2015). This difference in linearization properties between experiencer-object psych verbs and canonical transitive verbs is illustrated in example (2). Due to case syncretism between nominative and accusative plural, both sentences in (2) are globally ambiguous. In (2b), which has an agent and a patient argument, there is a strong bias towards the $S_{AG} \prec O_{PAT}$ interpretation. In contrast, in (2a), which contains the psych verb *interessieren* 'concern,

interest', a potential bias towards the $S_{STIM} \prec O_{EXP}$ interpretation is much less clear and an $O_{EXP} \prec S_{STIM}$ interpretation is easily available.

(2) a. *Damals interessierten die Lehrer die Schüler.*
back.then intrigue the teacher.NOM/ACC.PL the pupil.NOM/ACC.PL
(equally preferred readings) 'Back then teachers/pupils intrigued pupils/teachers.'

b. *Früher schlugen die Lehrer die Schüler.*
back.then beat the teacher.NOM/ACC.PL the pupil.NOM/ACC.PL
(preferred reading) 'Back then teachers beat pupils.'

(Non-)Agentivity has been argued to be a crucial factor in determining word order variation based on well-known properties of agents in contrast to properties of other thematic roles. The early occurrence of agents in linearization can be traced back to topic-hood along the following lines: agents tend to be topics, topics tend to occur first in an utterance resulting in an agent-first preference (e.g., Chafe 1976, Lambrecht 1994, Brunetti 2009). Furthermore, thematic role hierarchies are consistent in placing the agent on top so that agents outrank all other roles including experiencers. In as far as thematic role hierarchies influence argument linearization agent-like stimulus arguments are expected to precede experiencers while theme-like stimulus arguments are not (e.g., Grimshaw 1990, Van Valin and LaPolla 1997). For several languages, the relevance of agentivity in determining argument order has been experimentally shown (e.g., Bornkessel et al. 2005, Scheepers et al. 2000, Verhoeven 2009). Based on such evidence, the present study focuses on the role of agentivity and examines whether a binary or a scalar model of agentivity accounts best for the word order facts.

The second syntactic phenomenon at issue is the role of agentivity in the choice of subject. In many languages, transitive experiencer-object verbs regularly alternate with intransitive experiencer-subject variants (cf. Engl. *concern* ~ *be concerned with*). German transitive experiencer-object verbs alternate with reflexive (e.g. *interessieren* 'interest' ~ *sich interessieren* 'REFL interest') or stative passive structures (e.g. *faszinieren* 'fascinate' ~ *fasziniert sein von* 'be fascinated by'). Previous research demonstrated that agentivity influences the frequency of voice: experiencer-object verbs with (potentially) agentive stimuli occur more often in active voice than those with non-agentive stimuli (Grafmiller 2013, ch. 4.3. for English written and spoken corpus data; Verhoeven 2014 for experimental speech production in German; Verhoeven 2015 for German written corpus data; Pijpops and Speelman 2015 for Dutch written and spoken corpus data).[2]

The notion of agentivity leads to a binary classification between a subset of psych verbs that allow for agentive readings and another subset of verbs that do not do so: non-agentive verbs such as *concern* exclude an agentive interpretation in which the subject has conscious control over the event; verbs like *frighten* on the other side may occur with agentive or non-agentive readings. The conceptual distinction is binary and is expected to lead to a straightforward classification of every verb in one or the other class; see detailed discussion of the conceptual background in Section 2. The diagnostics that follow from this definition is the compatibility of the respective verb with a propositional content that entails an agentive contribution of the subject constituent. However, applying these diagnostics in an experimental setting reveals gradience: most experiencer-object verbs are not unambiguously compatible or incompatible with such contexts, but they are judged to be compatible to a certain extent. After establishing the existence of gradience in Section 3, we will address the question of whether this gradience is grammatically

relevant; see question (1b) above. In Section 4, we compare gradient and categorical models of agentivity in order to explain corpus frequencies on voice and word order properties of experiencer-object verbs in German. The results show that a binary notion of agentivity reaches the maximal fit in explaining the frequencies in the corpus.

2 Agentivity

Agentivity is generally conceived of as the capacity to control a situation. Several notions such as volition, intention, sentience, instigation, causation, and action have been identified as properties of agentive participants (e.g., Dowty 1991, Lehmann 1991, Van Valin and Wilkins 1996, Van Valin and LaPolla 1997, Primus 2012 etc.). These notions emphasize different aspects related to agentive situations. For the purpose of the present study we focus on volitional and intentional involvement as corresponding to control in a situation and hence as a prerequisite for agenthood. With respect to experiencer-object psych verbs, the agentivity of the stimulus is crucial; the question is whether the stimulus can be understood as controlling the accomplishment of the verbal event related to the experiencer. Recent studies on the semanto-syntactic properties of psych verbs distinguish between three subtypes of stimulus arguments, i.e. agentive, causative and theme/subject matter stimuli (Pesetsky 1995, Arad 1998, Reinhart 2001, 2002, Landau 2010, Verhoeven 2010). It has been shown that some psych verbs may be compatible with different stimulus flavors, implying agentive and non-agentive readings, while other psych verbs are clearly non-agentive (see examples below). The possibility of an agentive reading is lexically conditioned. The choice between agentive and non-agentive readings (with verbs that allow for both interpretations) is determined by the context.

Animacy plays a decisive role in this regard (Verhoeven 2010, Primus 2012): an agentive interpretation of a stimulus necessarily relies on its animacy, i.e. only animate participants can be agentive. A causer stimulus is susceptible to be interpreted as agentive when this role is filled by an animate participant (Holisky 1987, Van Valin and Wilkins 1996). Thus, in (3) the stimulus argument *Leonie* can be attributed some control over the event of impressing the audience (cf. Engelberg 2015 for similar corpus examples). For instance, Leonie can be thought of as having done her best in order to impress the audience. By means of controlling her action that provokes the audiences' impression, she indirectly controls the latter. However, an inanimate stimulus in (3) would necessarily invoke a non-agentive reading (e.g., 'the performance impressed the audience').

(3) *Leonie beeindruckte das Publikum mit ihrer Darbietung.*
Leonie impressed the audience with her performance
'Leonie impressed the audience with her performance.'

Some German experiencer-object psych verbs are more directly associated with (specific) actions bringing about the experiential change of state as e.g. *ärgern* 'annoy' or *erschrecken* 'frighten'; see examples in (4). In these examples, the animate stimulus participants are clearly interpreted as agentive.

(4) a. *Jugendliche erschreckten die Passanten (mit Masken).*
teenagers frightened the pedestrians (with masks)
'Teenagers frightened the pedestrians with masks.'
b. *Max wird oft von seinen Kameraden geärgert.*
May is often by his comrades annoyed
'Max is often annoyed by his comrades.'

At the other extreme, there are experiencer-object verbs that are definitely non-agentive. Their stimulus argument is necessarily interpreted as a subject matter/theme. Some of those verbs such as *wundern* 'astonish' and *freuen* 'please, give pleasure' do not even allow for animate participants in this function (see examples in (5)).

(5) a. *Die Geburt des kleinen Jungen/ *der kleine Junge freute die Eltern.*
the birth of the little boy/ *the little boy gave.pleasure the parents
'The birth of the little boy/ *the little boy gave pleasure to the parents.'

b. *Der Wahlsieg /?der Politiker wunderte niemanden.*
the election victory /?the politician astonished nobody
'The election victory /?the politician didn't astonish anybody.'

Agentivity is tested by evaluating the possibility of the stimulus' volitional or intentional involvement in the experiential event. Several structural frames are frequently used to test agentivity. We will focus on two frames, namely (a) the modification of an event with agent-related adverbs indicating volitionality (*intentionally*, *on purpose*), (see e.g., Roeper 1987, Talmy 1976; cf. Klein and Kutscher 2002, Verhoeven 2010), and (b) the embedding under subject control verbs of decision (Grafmiller 2013, ch. 5.2). These tests indicate whether an agentive reading of the verb at issue can be accommodated (in a possible context) or not. Studies on the acceptability of semantic compatibility generally produce gradient data. We assume that agentivity per se is not a scalar notion; a verb either allows for an agentive reading or not. The scalar judgments reflect the possibility to imagine a context in which the verb is used as agentive. If this is true, the gradience should correlate with the likelihood of such contexts in speech production. The same reasoning holds for a polysemy-based account of these verbs: the alternative meanings (agentive, non-agentive) may have differences in weight, i.e., it may be easier to retrieve

the one or the other meaning. Crucially, this does not imply that agentivity has more than two levels but that the strength of association of these levels with a particular lexical concept may vary. Consider the following examples in which experiencer-object psych verbs are embedded under a matrix verb involving subject control.

(6) a. **Die Polizistin entschloss sich, Sarah zu freuen*.
'The police woman decided to please Sarah.'
b. ??*Der Taxifahrer beschloss, Silke zu entsetzen*.
'The taxi driver decided to appall Silke.'
c. *Hannes hat beschlossen, die Lehrerin zu ärgern*.
'Hannes decided to annoy the teacher.'

While the sentence in (6a) is ungrammatical, (6b) is more acceptable and (6c) seems completely fine. The differences are brought about by different mechanisms and interpretations: *freuen* does not accommodate an animate stimulus subject, so control is out of question with this verb (see above); *entsetzen* 'appall' belongs to a subgroup of EO psych verbs which are potentially associated with activities that bring about the state of the experiencer of 'being appalled'. In richer contexts, such verbs allow for the interpretation/inference of an intentional involvement of the stimulus participant, e.g. by adding an instrumental prepositional phrase as documented in example (7) from the DeReKo (Deutsches Referenzkorpus, IDS 2010; see Grafmiller 2013, ch. 4.3.2 for similar observations with respect to English experiencer-object verbs).

(7) *Da entsetzt A. die linken Heerscharen seiner einstigen Bewunderer mit der vernünftigen Auskunft, Deutschland sei kein Zuwanderungsland.*

‘Then A. appalled the left-wing legions of his former admirers by providing the reasonable information that Germany is not a country of immigration.’ (L98/DEZ.24032)

In contrast, verbs such as *ärgern* are clearly associated with controlled activities. These verbs are similar to agentive verbs such as *teach*, which “denote an act performed with the intention of triggering a certain change of state” (Martin and Schäfer to appear). However, if these verbs occur with an agentive subject (in contrast to a causer subject) culmination of the change of state is not necessarily part of the truth conditions of the sentence. This can be shown by the felicity of cancelling the culmination of the result state in sentences with agentive subjects in contrast to causer subjects (see Martin 2015 for a comprehensive argumentation of the relation between agentivity and non-culminating causation). Thus, while denying that the object participant of the first clause in (8a) is in the state of being annoyed is pragmatically felicitous (in German), the same does not hold for (8b).

(8) a. *Anne ärgerte den Lehrer, aber er ärgerte sich gar nicht / war gar nicht ärgerlich.*
‘Anne annoyed the teacher, but he did not feel annoyed.’

b. *Die schlechten Schülerleistungen ärgerten den Lehrer, #aber er ärgerte sich gar nicht / war gar nicht ärgerlich*.
‘The low performance of the pupils annoyed the teacher, but he did not feel annoyed.’

The difference between (8a) and (8b) lies in the animacy of the stimulus participant: while animate participants can be interpreted as agents, this does not hold for inanimates, which are necessarily causers or subject matters (see above).

The previous discussion has demonstrated that individual experiencer-object verbs display crucial differences in their property of being (potentially) agentive. In the following sections, we will investigate whether these differences are systematically reflected in their grammatical behavior in natural language production as found in written corpora of German. We will explore whether agentivity should be conceived of as scalar or not (see (1a)) and, in particular, whether the grammatical behavior of experiencer-object verbs related to argument order and argument alternations is best explained by assuming coarse-grained or finer-grained lexical subclasses or even individual lexical distinctions (see (1b)).

3 Diagnostics of agentivity

3.1 Method

In order to test the impact of lexical differences on agentivity, we selected 20 experiencer-object verbs, which are listed in Table 1. The sample contains verbs that are frequently examined in the research on experiencer-object verbs. We avoided verbs that frequently occur in non-experiential readings (e.g., *bewegen* 'move physically or emotionally', (*be*)*rühren* 'touch physically or emotionally'). The list contains some verbs that are clearly non-agentive in German, e.g., *interessieren* 'interest', *freuen* 'give pleasure', *wundern* 'astonish', some verbs that clearly allow for agentive readings, e.g., *nerven* 'bother', *ärgern* 'annoy' and *erschrecken* 'frighten' as well as further verbs for which the intuitive classification is less straightforward.

Table 1. Sample verbs

amüsieren ‘amuse’	*bedrücken* ‘depress’	*enttäuschen* ‘disappoint’	*freuen* ‘give pleasure’	*nerven* ‘bother’
anwidern ‘nauseate’	*befremden* ‘alienate’	*entsetzen* ‘appall’	*interessieren* ‘interest’	*reizen* ‘irritate’
ärgern ‘annoy’	*ekeln* ‘disgust’	*erschrecken* ‘frighten’	*irritieren* ‘confuse’	*überraschen* ‘surprise’
aufregen ‘excite’	*empören* ‘outrage, anger’	*faszinieren* ‘fascinate’	*langweilen* ‘bore’	*wundern* ‘astonish’

The verbs were inserted in two sentential frames that test agentivity (see Section 2). The first test examines the compatibility with a subject-oriented adverb denoting the intentional involvement of the subject, namely *absichtlich* ‘on purpose’; see (9). If the lexical semantics of the verb is incompatible with an agentive reading, which is the case for the non-agentive verb *wundern* ‘astonish’ in (9a), this sentence is expected to be rejected by native speakers. If the lexical semantics of the verb allow for agentive readings, e.g., the verb *nerven* ‘bother’ in (9b), then it is expected to be compatible with the intentionality adverb.

(9) Compatibility with subject-oriented intentional adverb

a. *Gerda hat den Jugendlichen absichtlich gewundert.*
Gerda has the teenager on.purpose astonished
‘Gerda astonished the teenager on purpose.’

b. *Der Junge hat Björn absichtlich genervt.*
the boy has Björn on.purpose bothered
‘The boy bothered Björn on purpose.’

The second sentential frame tests the possibility to embed experiencer-object verbs under predicates implying that the subject has control over the event given in the subordinate clause; see (10). If the embedded psych verb excludes an agentive reading, e.g., the verb *ekeln* 'disgust' in (10a), then it is expected to be incompatible with a matrix control verb. Verbs that allow for agentive readings are expected to be compatible with this construction, e.g., the verb *nerven* 'bother' in (10b).

(10) Embedding under control verbs

a. *Der Lehrer hat beschlossen, Nathalie zu ekeln.*
the teacher has decided Nathalie to disgust
'The teacher decided to disgust Nathalie.'

b. *Hannes hat beschlossen, die Lehrerin zu nerven.*
Hannes has decided the female.teacher to bother
'Hannes decided to bother the female teacher.'

Each verb was inserted in the sentential frames in (9)-(10). Several proper names or definite common nouns were used as subjects and objects with each verb. All examples contained DPs denoting individuals (but not event-related nouns or inanimate concepts that are by definition non-agentive as subjects of experiencer-object verbs). Assuming that the choice of a particular individual (or the proper/common noun distinction) does not play any role for the compatibility of the verb with the agentive environment, we varied these lexicalizations in order to present the participants with a more diversified set of sentences.

Participants were presented the target sentences and were instructed to estimate their well-formedness on a 1-7 Likert scale (1=very bad, 7=very good). The experiment was implemented in *OnExp 1.2* (http://onexp.textstrukturen.uni-goettingen.de/) and was

performed online. Thirty-two native speakers (20 female, 12 male; age range 17-52, average 25.7) participated in this study (Berlin, July 2013).

3.2 Results

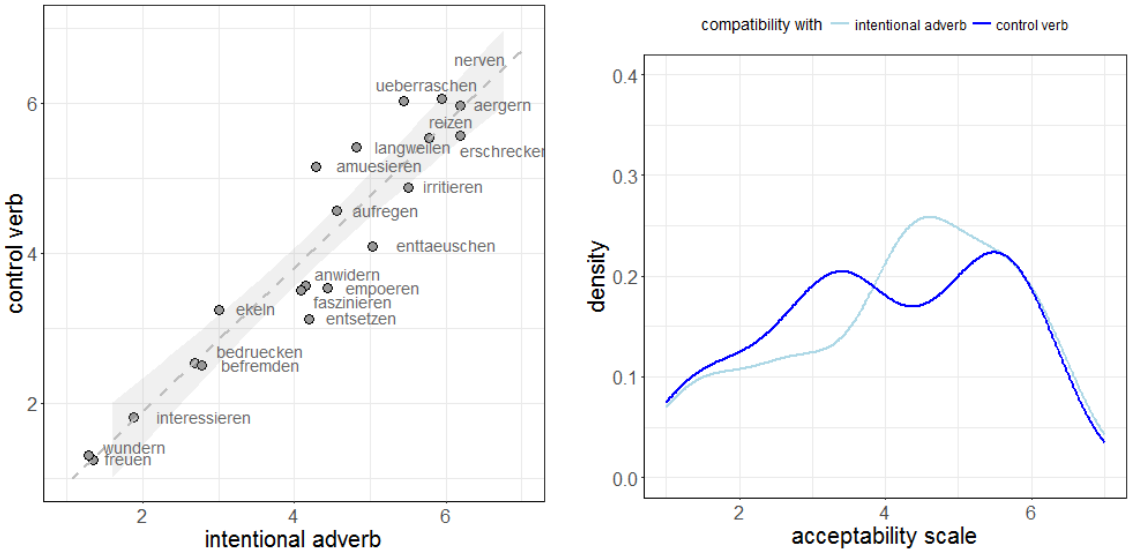
This study resulted in (32 speakers × 20 verbs × 2 tests =) 1280 judgments. There were no missing values, i.e., the final result was calculated on the basis of 1280 estimations on the 1-7 scale. The speakers exploited the whole range of the scale (1 = 216; 2 = 183; 3 = 148; 4 = 118; 5 = 185; 6 = 211; 7 = 219; total = 1280) with a slight preference for the extreme values. The mean values as well as the standard errors of the means are listed in Appendix I and presented in Figure 1. The estimations of the compatibility with an intentional adverb and the embedding under a control verb are strongly correlated (correlation r = .94); see Figure 1a. A linear regression on the obtained means by verb (with 'control verb' as *explanatory variable* and 'intentional adverb' as *response variable*) reveals an intercept of –.03 and slope of .96, which are the coefficients of the least-squares regression line plotted in Figure 1a. The acceptability with control verbs is a significant predictor of the acceptability with intentional adverbs: a linear regression reveals that the slope of .96 is highly significant (t = 12.03; p < .001).

The means by verb form a scale: in the control-verb test, the means range from 1.25 (*freuen* 'give pleasure') to 6.06 (*nerven* 'bother'); in the intentional-adverb test, the means range from 1.28 (*wundern* 'astonish') to 6.19 (*erschrecken* 'frighten', *ärgern* 'annoy'). However, the distribution of the means across this range is not even; see the density graph in Figure 1b. In the intentional-adverb test, seven verbs (35% of the verb sample) are judged within the score range [4, 5], which is reflected in the density graph. In the control-verb test, the means suggest a bimodal distribution around two central values: a group of five verbs (*entsetzen* 'appall', *ekeln* 'disgust', *faszinieren* 'fascinate', *empören* 'outrage',

anwidern 'nauseate') is found within the range [3.1, 3.6], another group of five verbs (*reizen* 'irritate', *erschrecken* 'frighten', *ärgern* 'annoy', *überraschen* 'surprise', *nerven* 'bother') is found within the range [5.5, 6.1]. This results in a bimodal density curve in Figure 1b. In order to test the statistical significance of the observed multimodality, we run Hartigan's dip test for unimodality.[3] This test examines the alternative hypothesis that the data is not drawn from a unimodal distribution. It revealed a D =.08 (associated with a p-value of .3), i.e., the data does not provide enough evidence that it is drawn out of a population that is bimodal.

Figure 1. *Agentivity tests*

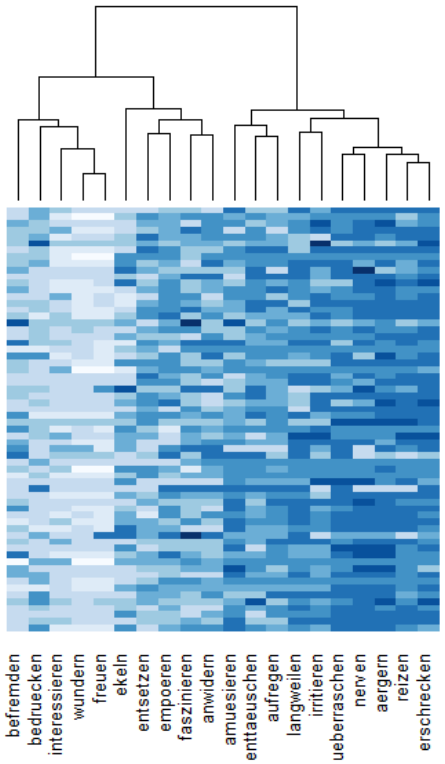
(a) mean values by verb (b) density graph



The straightforward implementation of the concept of verb classes is that the inventory of verbs contains clusters with distinctive behavior. In order to identify the classes, we computed a cluster analysis based on the Euclidian distance between verbs based on the obtained judgments in both tests. The result is plotted in Figure 2. The heatmap visualizes

all (32 speakers × 2 tests × 20 =) 1280 judgments through different darkness levels (light blue = low agentivity value; dark blue = high agentivity value). Each column in the heatmap contains the judgments for a particular verb, as shown in the *x*-axis. Each row in the heatmap contains the judgments by a single speaker in an agentivity test. A distance matrix reflecting the dissimilarity between the judgments was calculated for each pair of verbs. The binary-branching dendrogram on the top of Figure 2 displays the obtained clusters of verbs based on the dissimilarity values. The interpretable information of this figure is the clustering (and not the linear order of the verbs in the *x*-axis).

Figure 2. *Cluster analysis based on agentivity judgements*



The highest clustering reflects the distinction between the 10 non-agentive (left branch of the dendrogram) and 10 agentive verbs (right branch of the dendrogram). Among the

non-agentive verbs, *wundern* and *freuen* are maximally similar: as already discussed with respect to example (5), these verbs are incompatible with animate stimuli, which is confirmed by the low agentivity scores in both tests. Then *interessieren*, *bedrücken* and *befremden* are successively added: these verbs receive generally low scores in both agentivity tests. The next cluster contains verbs with acceptability scores between 3.25 and 4.4, i.e. *ekeln*, *entsetzen*, *empören*, *faszinieren*, *anwidern*. The right branch of the dendrogram contains a left branch with three verbs (*amüsieren*, *enttäuschen*, *aufregen*) with acceptability scores between 4.09 and 5.14 and a further cluster of 7 verbs with higher acceptability. This latter cluster is further subdivided (*langweilen, irritieren*) and (*überraschen, nerven, ärgern, reizen, erschrecken*). The latter verbs are most clearly associated with controlled actions, as explained in Section 2.

3.3 Discussion

The question is whether this data allows for a clear-cut distinction of two subsets of verbs in the examined verb sample. The fact that the experimental findings reveal a range of agentivity values does not permit a conclusion with respect to the nature of this category. Data collected through repeated observations are expected to involve variance just through the repetition of the task by different speakers. Hence, it is expected that the obtained scores vary. However, the significant correlation between the two tests in Figure 1a suggests that a part of this variance is indeed explained by some property of lexical semantics of the verb. If the variance was only explained by random factors of the experimental setting such as different speakers, lexicalizations etc., non-agentive verbs would be judged as less felicitous in both tests, but beyond this, the exact score within the range of non-felicitous configurations would not be predictable. But this is not what the data in Figure 1a show: the verbs *wundern* ‘astonish’ and *freuen* ‘give pleasure’ receive

scores around 1.3 in both tests, the verbs *bedrücken* 'depress' and *befremden* 'alienate' are judged with values around 2.6, the verbs *anwidern* 'nauseate' and *faszinieren* 'fascinate' around 4 etc. I.e., the acceptability level of the control-verb construction is a significant predictor of the acceptability level in the intentional-adverb test, as confirmed by the significant effect of the linear regression (see Section 3.2). This finding justifies the assumption of an agentivity scale. As discussed in Section 2, agentivity can only be a binary concept. However, it is possible for many verbs to occur in certain contexts with readings that are not predicted by their intension. The gradience in the speakers' judgments can only reflect this phenomenon, i.e., estimations of the likelihood that the respective verbs may be used in agentive contexts. The fact that the two diagnostics are strongly correlated indicates that speakers have very precise intuitions about the likelihood of individual verbs to occur in such contexts.

The density of the control-verb test displays a bimodal distribution. This type of distribution is potential evidence that the sample verbs are drawn from two populations of (German experiencer-object) verbs whose distributions are reflected in the result. However, the null hypothesis of uni-modality is not statistically rejected (see results of the dip test in Section 3.2), which means that we are not justified in inferring that the descriptively obtained bimodal distribution reflects a bimodal distribution in the population.

The cluster analysis has helped us to establish possible verb classifications based on the empirical data. The highest level of clustering corresponds to the difference between agentive and non-agentive experiencer-object verbs that is used in syntactic literature based on expert knowledge. From an empirical point of view, we cannot *a priori* know whether the grammatically relevant level of clustering is the highest branching or a more

detailed classification. The cluster analysis gives us the available options. Their grammatical relevance will be examined in the next section.

4 Predicting voice and order

In Section 3, we established the gradience of intuitions of agentivity in a sample of 20 experiencer-object verbs as a result of repeated observations on the acceptability of their occurrence in two agentivity tests. In this section, we will address the question of whether this gradience is grammatically relevant; see question (1b) above. In particular, we will investigate whether the frequencies of voice alternations (i.e. the occurrence of passives and reflexives) and argument order frequencies (i.e. stimulus- vs. experiencer-first orders) can be predicted based on the agentivity judgments.

In a nutshell, the question is how many levels of agentivity we need in order to understand the grammatical phenomena that are sensitive to agentivity distinctions. For this purpose, we use the gradience established through the acceptability tests, independently of the question where this gradience comes from. The reasoning of the empirical studies is the following: if the scales reflecting the strength of agentivity readings are grammatically relevant, then the individual levels of these scales should serve as explanatory variables for the frequency of grammatical alternations that are sensitive to agentivity distinctions (i.e. voice and order). A possible outcome of this investigation is that only a subset of the empirically established levels is relevant: if this subset only contains two levels (i.e. the minimum), then the grammatically relevant scale will not be different from a binary distinction between two groups.

After introducing our corpus and the methodology used (Section 4.1), we compare gradient and categorical models of agentivity in order to explain the frequencies of voice (Section 4.2) and word order (Section 4.3).

4.1 Method

The dataset of the present study consists of 20 (verbs) × 1 000 (tokens) = 20 000 tokens extracted from the IDS 2010 corpus (COSMAS II DeReKo database, Corpus *W-öffentlich*, containing a total of 2 291 520 000 word forms).[4] The tokens for each verb were extracted by using the randomization function of the web interface on lemma-based queries for each verb. The extracted data was manually annotated for several categories: clause type (declarative | interrogative | imperative), clause layer (main clauses | embedded clauses), voice (active | reflexive | regular_passive | stative_passive), word order (SVO | SOV | OVS | OSV | XVSO | XVOS). The aim of the present study is to examine the impact of verb classes on voice and order. For this purpose, we restricted the dataset to main declarative clauses featuring two non-clausal arguments, in which the respective verb occurs in a finite form (including periphrastic tense forms). This subset contains 3 942 sentences (19.71% of the entire dataset). Before examining the influence of agentivity on voice (section 4.2) and word order (section 4.3), we introduce the relevant annotation categories.

Experiencer-object verbs frequently appear in non-active voice; see counts in Appendix II. In the examined sample, we found 2049 tokens (51.98%) in active voice and 1893 tokens in non-active form (48.02%). Regular passive voice only rarely occurs with these verbs (93 tokens, mostly with the verb *überraschen* 'surprise'; i.e. 4.9% out of 1893 non-active forms); see illustrative example in (11).

(11) Regular passive verb[5]

Zu oft war er als Präsident von Trainern
to often was he.NOM as president by coaches
und Spielern schon enttäuscht worden.
and players already disappointed got
'It has been too often that he has already been disappointed by coaches and players.' (SOZ05/MAI.00807)

Experiencer-object verbs appear in two main voice alternations, i.e. stative/adjectival passive (744 tokens, i.e. 39.3% out of 1893 non-active) and reflexive (1056 tokens, i.e. 55.8% out of 1893 non-active); see examples in (12). Some verbs show both alternations (e.g. *interessieren* 'interest', *interessiert sein*, *sich interessieren* 'be interested'); see counts in Appendix II. Other verbs are restricted to one of both alternation types (e.g. *enttäuschen* 'disappoint', *enttäuscht sein* 'be disappointed' vs. *wundern* 'astonish' *sich wundern* 'be astonished').

(12) a. Stative/adjectival passive verb

Die sechsjährige A. ist fasziniert von den Akrobatinnen.
the six.year.old A.NOM is fascinated by the female.acrobats
'Six-year old A. is fascinated by the female acrobats.' (L99/JUN.28202)

b. Reflexive verb

Meine Tochter ekelt sich auch vor den Flöhen.
my daughter.NOM feel.disgusted REFL also by the fleas
'My daughter finds the fleas disgusting, too.' (A08/JUN.07923)

German is a language with flexible word order displaying object-before-subject linearizations triggered by diverse factors, among them animacy, definiteness, pronominality, heaviness, but also thematic role of the arguments (Lenerz 1977, Hoberg

1981, Bader and Häussler 2010, Verhoeven 2015). For the present study, we investigate the relative order of subjects and objects in the transitive occurrences of our target verbs. We distinguish between sentences in which the subject precedes the object (SO) or vice versa (OS).[6] Two options appear in German main declarative clauses: either one of the arguments (S or O) appears preverbally and the other postverbally (see (13a/b)); or, if the prefield is occupied by other material as e.g. by an adverbial, both arguments occur postverbally in either order (13c/d). We obtained 1567 tokens in SO order (i.e. 76.5% out of 2049 active transitive clauses) and 482 tokens in OS order (i.e. 23.5% out of 2049 active transitive clauses), see counts in Appendix II.

(13) a. Order: SO, preverbal S

Das vorgelegte Budget 1992/93 enttäuschte
the presented budget.NOM 1992/93 disappointed
die Anleger in der Vorwoche.
the investers.ACC in the previous.week
'In the previous week, the presented budget 1992/93 disappointed the investors.' (P92/MAR.07867)

b. Order: OS, preverbal O

Die russischen Abgeordneten bedrückt
the Russian deputies.ACC depress
vor allem der neue Verfassungsentwurf ...
above all the new draft.constitution.NOM
'The Russian deputies are especially depressed by the new draft constitution ....' (N93/JUL.27150)

c. Order: SO, both arguments postverbally

Seit vergangenem Jahr fasziniert die Cartoon-Reihe
since last year fascinates the cortoon-series.NOM
"*South Park*" *das amerikanische Publikum.*
South Park the American public.ACC

'Since last year, the cartoon series "South Park" fascinates the American public.' (N99/FEB.05347)

d. Order: OS, both arguments postverbally

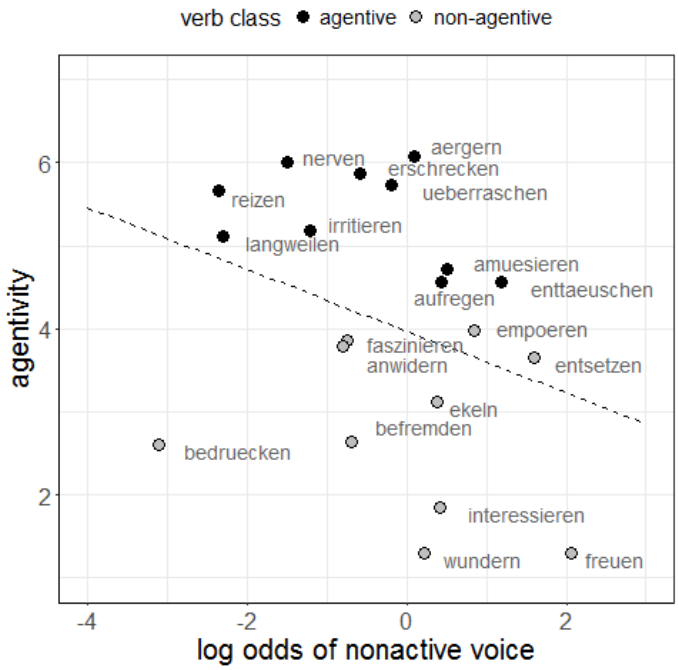
Am meisten ärgerte den Walliser das Lospech, ...
at most annoyed the Welshman.ACC the lot.bad.luck.NOM
'The Welshman was most annoyed by the bad luck.' (E96/FEB.02916)

4.2 Influence of agentivity on voice

The likelihood of non-active voice varies between verbs. Some verbs occur very rarely or not at all in non-active voice, e.g., *bedrücken* 'depress' (4.3%) or *reizen* 'irritate' (8.5%); other verbs most frequently appear in non-active voice, e.g., *entsetzen* 'appall' (83.2%) and *freuen* 'give pleasure' (88.7%) (see Appendix II for details). The question is whether a part of this variation is explained through the differences in agentivity. Figure 3 shows the relation between the likelihood of non-active voice and the average agentivity values by verb (i.e. the average of the results of both agentivity tests in Figure 1).[7] The two measures are inversely correlated: decreasing agentivity increases the likelihood of non-active voice. A logistic regression with agentivity as an explanatory variable and the occurrence of non-active voice as dependent variable reveals that the impact of agentivity on the occurrence of non-active voice is explained by the linear model in (14). The agentivity slope is associated with a significant p-value ($z = -10.5$; $p < .001$), i.e., the observed influence is beyond the chance level.

(14) $\log(p_{\text{non-active}}/1 - p_{\text{non-active}}) = .75 - .22*\textit{agentivity}$
($p_{\text{non-active}}$: likelihood of non-active voice)

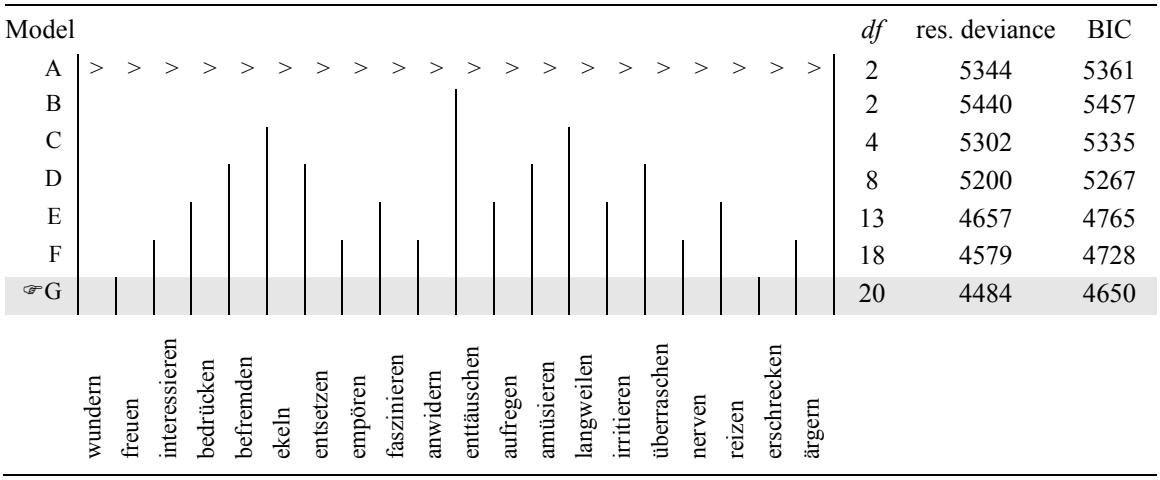
Figure 3. *Agentivity as explanatory variable for non-active voice*



However, Figure 3 indicates that a great amount of variation is not explained by the impact of agentivity as reflected in the large dispersion of the data points from the regression line. Agentivity is weakly correlated with the logarithmized odds ratio of non-active voice (r = –.32). Even if the agentivity value is a significant predictor for the frequencies of voice, it is not clear whether a model based on the gradient judgments of agentivity has a better fit on the data than a model assuming discrete classes of experiencer-object verbs. Hence, we compare the fit of the scalar agentivity model with the fit of models based on the assumption of verb classes with respect to agentivity. Linguistic theory assumes a binary contrast between –agentive and ±agentive verbs. The cluster analysis based on the diagnostics of agentivity reveals the possibility of a more fine-grained classification (see Figure 2). The model comparison in Figure 4 reports the model-fit measures for a model assuming the average agentivity value as explanatory

variable (model A) as well as for verb-class models that correspond to different depths of the cluster analysis (models B-F). The last model (model G) tests the possibility that the explanatory variable is the individual verb without reference to the agentivity value or to the different classes, i.e., the frequency of non-active voice is just an idiosyncratic property of the individual verbs. The degrees of freedom (*df*) display the levels of the explanatory variable, i.e., they reflect the conceptual complexity of the explanatory model. The residual deviance is a measure of the lack of model fit in logistic regression: a lower value indicates a better model fit. We observe in Figure 4 that the scalar model (model A) has a better model fit than a model based on a binary distinction between two verb classes (model B). In the models B-G, the lack of model fit as reflected in the residual deviance decreases, such that a fourfold distinction (model C) already reaches a better model fit than the scalar model A. However, adding new parameters to the model necessarily results in a better fit. The role of the BIC value (Bayesian Information Criterion) serves to weigh the model fit with an estimate of the model complexity in order to test whether the stipulation of new parameters is justified by the increase of the model fit (BIC = $Deviance + n_{\text{parameters}} \times \log(n_{\text{observations}})$). Lower BIC values indicate better model fit. The best (=lowest) BIC value is reached by model G, i.e., the assumption that each verb has an idiosyncratic impact on the likelihood of voice. Despite the large amount of parameters (20 different verbs), this model reaches the best BIC value due to large deviances of the agentivity-based models (models A-F).

Figure 4. *Model fit of logistic regressions on Voice*



The linguistic consequence of the statistics in Figure 4 is that agentivity generally correlates with the likelihood of non-active voice, but only a small amount of the variation can be explained by agentivity, independently of whether we model agentivity as a scale or as a discrete classification of two or more classes. Ultimately, the differences in frequency of voice relate to idiosyncratic properties of the individual verbs and can be best accounted for by a model that assumes that each lexical item has an individual influence on the frequency with which voice is chosen in discourse. This conclusion only relates to the putative role of agentivity and it does not exclude that further linguistic factors may explain this variation better (e.g., factors relating to lexical aspect, see Grafmiller 2013).

4.3 Influence of agentivity on order

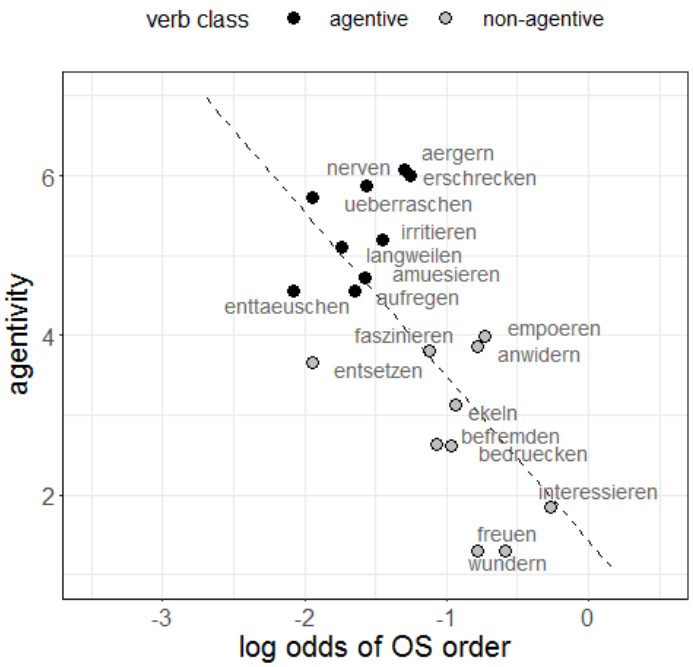
The word order facts were analyzed with the same procedure introduced for voice in Section 4.2. The likelihood of OS varies between verbs (see Appendix III). Some verbs rarely occur with OS order, e.g., *reizen* 'irritate' (3.7%), while with other verbs, the same order is quite frequent, e.g., *interessieren* 'interest' (43.4%).

Figure 5 shows that the average agentivity values per verb explain part of the variation in word order. Again, agentivity and order are inversely correlated: decreasing agentivity increases the likelihood of OS. A logistic regression with agentivity as an explanatory variable and the occurrence of OS as dependent variable reveals that the impact of agentivity on the occurrence of OS is explained by the linear model in (15); the effect of agentivity is associated with a significant *p*-value ($z = -6.8$; $p < .001$).

(15) $\log(p_{OS}/1 - p_{OS}) = -.2 - .24*\alpha$

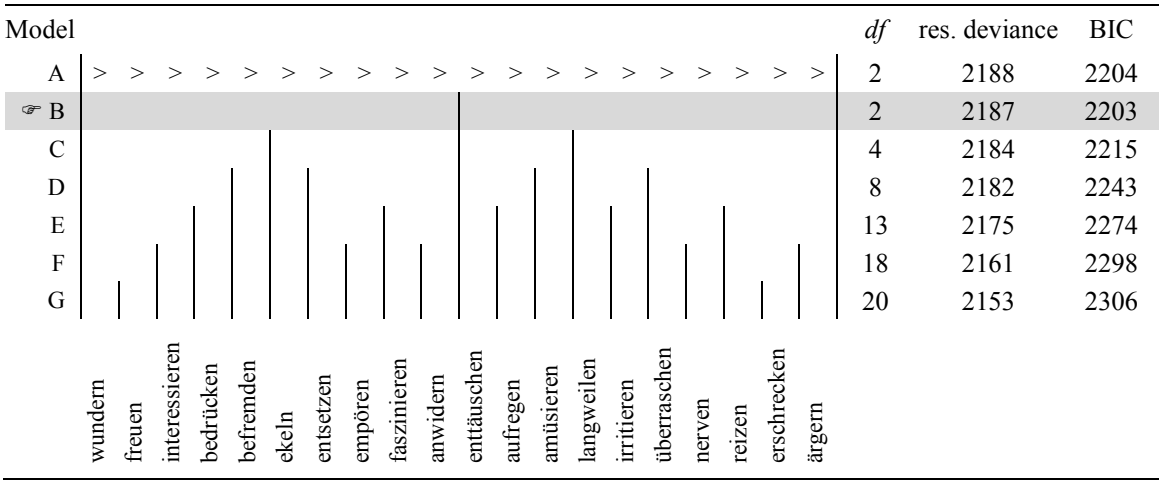
(p_{OS}: likelihood of OS order; α: agentivity)

Figure 5. *Agentivity as explanatory variable for OS*



A visual comparison of the order facts in Figure 5 and the voice facts in Figure 3 suggests that agentivity correlates stronger with word order ($r = -.67$) than with voice ($r = -.32$). The residual deviances of the scalar model (model A) and the two-class model (model B) are very similar with a slight advantage for the binary model; see Figure 6. Note that both models involve the same degrees of freedom (df=2), i.e., the amount of stipulation by assuming that the explanatory variable is a scale or contains two levels (±agentive vs. non-agentive) is the same, correspondingly. The model fit in this case depends on the deviance of the models alone. The models C-G display a lower degree of residual deviance, as is expected for models with more parameters. However, with these more complex models the BIC values do not look as promising. Penalizing the model complexity leads to the conclusion that an assumption of verb classes of any depth is not justified by the amelioration of the model deviance. An analysis of deviance reveals that the improvement of the residual deviance by a model that assumes more than two classes is not significant (model A vs. model C: $\chi^2(2) = 3.1$; $p = .16$, model B vs. model C: $\chi^2(2) = 2.8$; $p = .23$). Hence, there is no reason to assume a model that adds more complexity than the simple models A and B. The class-based model (model A) has a slight advantage in the residual deviance, but there is no way to confirm the statistical significance of this difference because both models have the same degrees of freedom.

Figure 6. *Model fit of logistic regressions on Order*



The linguistic conclusion of the model comparison in Figure 6 is that the word order frequencies are significantly influenced by agentivity – as established by the obtained judgments in the agentivity diagnostics. A simple model assuming a binary distinction between ±agentive vs. non-agentive verbs (model B) reaches the lowest BIC value and hence the maximal fit.

4.4 Discussion

The analyses presented in the two previous sections result in diverging conclusions with regard to our main question, namely whether the gradience in the agentivity judgments of experiencer-object verbs may explain the corpus frequencies for the choice of voice and of argument order. The statistical analysis confirmed the role of agentivity on word order frequencies, but not so for the choice of non-active voice.

In section 4.2 we saw that agentivity weakly correlates with the likelihood of non-active voice, but only a small amount of the variation can be explained by either a scalar or a more fine-grained verb classification based on the agentivity scores (as identified by the cluster analysis in Section 3). The statistical comparison of the fit of alternative

models to the frequencies of the voice alternations revealed that the model supposing idiosyncratic properties of the individual verbs performed best. The fact that the individual influence of verbs is not explained by agentivity does not imply that the likelihood of non-active voice is a property of the lexicon; it can be due to other factors than agentivity that are not considered in our study. This result is in line with observations made by Pijpops and Speelman (2015) for Dutch transitive ~ reflexive psych alternations. In this corpus study, agentivity of the stimulus (measured indirectly through its animacy) is a significant predictor of the occurrence of the transitive vs. reflexive structures in addition to factors such as the topicality of experiencer and stimulus, and, most strongly, the individual verbs. Somewhat differently, Grafmiller (2013, sect. 4.3) shows that in English (written and spoken) corpus data the choice of active vs. passive voice with experiencer-object verbs is significantly determined (again amongst other factors) by the so-called potency of the stimulus (derived from the scale 'animate < event < abstract'). At the same time, there is a strong statistical correlation between the stimulus type and the event type, i.e. a distinction between concrete (animate, event) stimulus types and eventive lexical aspect on the one hand, and abstract stimulus types and stative lexical aspect. Hence, in English the alternation seems to be less determined by lexical idiosyncrasies, which might be related to the fact that argument alternations with psych verbs are regularly expressed through passive voice.

In contrast to English, German provides alternative means to realize the experiencer earlier than the stimulus. Agentivity was shown to inversely correlate with the likelihood of OS order (Section 4.3). Hence, we are in the position to conclude that agentivity has an influence on the argument order with experiencer-object verbs. The question that motivated this study is whether the gradience of the agentivity scale is grammatically

relevant. The model comparison in Figure 6 has shown that the maximal fit on the corpus data is reached by a model assuming a binary distinction between ±agentive and non-agentive verbs. The conclusion is that the predictive power of linguistic models on word order is not increased by assuming a scalar notion of agentivity, even if gradience exists.

5 Conclusion

The development of empirical methods that are based on repeated-observation designs leads to a new paradigm of data that contains gradience, either in terms of likelihoods or other types of scales (e.g., scalar intuitions). This type of data challenges our assumptions about grammar: is grammatical knowledge probabilistic, as advocated for instance in Bresnan (2007), or does gradience come from sources of variation that are irrelevant for grammar? Examining the components of gradience is crucial for our understanding of grammar. As stated in the beginning of this study, this enterprise involves two basic questions: (a) whether particular grammatical distinctions are categorical or gradient; and (b) whether the gradience is grammatically relevant, in particular, whether a scalar definition of grammatical distinctions leads to a better understanding of grammatical phenomena.

The present study examined these possibilities with respect to the concept of agentivity, which is a crucial property of verb meaning that determines several aspects of the respective verbal projections. Diagnostic tests of agentivity (i.e. compatibility with intentional adverbs and embedding under control verbs) yielded scalar results with several degrees of acceptability. Crucially, the gradience obtained in the judgments cannot be an artefact of the experimental procedure, since the averages by verb were strongly correlated in the two diagnostic tests. This suggests that speakers possess a very precise

knowledge of the gradience involved in the contexts at issue and in particular about the part of the variation that is determined by the examined verb. Hence, the diagnostics of agentivity involve gradience and this conclusion opens the question of how this finding is compatible with the theoretical assumption that verbs either allow for an agentive interpretation or not. We have argued that it is possible to create contexts in which the subject of a non-agentive verb may intentionally provoke the event, even if it cannot control event culmination. The gradience in the diagnostics corresponds to this source of variation, presumably to the ease of imagining such a context, which varies between verbs.

Finally, we addressed the question of whether this variation is grammatically relevant, i.e., if it helps us to reach more precise descriptions of the syntactic effects of agentivity. We examined two phenomena that have been considered as effects of agentivity in previous research: (a) frequency of non-active voice; (b) frequency of OS order. In the first study, we found that agentivity is weakly correlated with the likelihood of non-active voice. There is a significant effect of agentivity, but a model reflecting the null hypothesis with each individual verb having its own effect on the likelihood of non-active voice has a stronger explanatory power than agentivity-based models. In the word order study, we found that agentivity is strongly correlated with the likelihood of OS order. A comparison between different models of agentivity revealed that a maximal fit is reached by a model that assumes two verb clusters, namely ±agentive and non-agentive verbs. Summing up, the diagnostics of agentivity lead to a scalar distinction, presumably reflecting the likelihood of individual verbs to appear in the critical contexts; this gradience is empirically confirmed, but it does not increase the power of models that explain the impact of agentivity on syntactic phenomena.

6 Notes

[1] Special thanks are due to Birgit Jänen and Julian Rott for their contribution to data annotation, to Julian Rott for help with the acceptability study, and to Julian Rott and Nico Lehmann for helpful comments and proofreading. This article is part of the project VE 570/1-3 sponsored by the German Research Foundation.

[2] Note that both Grafmiller 2013 and Pijpops and Speelman 2015 infer agentivity on the basis of (scalar) animacy.

[3] See Hartigan and Hartigan 1985 for the mathematics of the dip test. Calculations were made using the R-package diptest, version 0.75-7 (created by M. Maechler).

[4] 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.

[5] The examples illustrating the investigated factors are explained in a word-by-word translation. The case of argument DPs is given by the gloss of the noun/pronoun (independently of the morphological exponence of the case on the noun or the determiner).

[6] Our use of the term subject (S) is in line with the traditional understanding of the term in German grammar, namely that it is always the nominative argument. Crucially, this is not a claim about the syntactic status of non-nominative experiencers.

[7] Since categorical data such as the corpus frequencies are skewed, we calculated the logarithmized odds, i.e. the common logarithm of the odds: $\log_{10}((p_{\text{non-active}})/(1 - p_{\text{non-active}}))$ (see Agresti [2]2007, 31). Log odds *0* results if the likelihood of non-active voice and the likelihood of its complement (active voice) are equal.

7 References

Agresti, Alan. ^{2}2007. *An Introduction to Categorical Data Analysis*. New Jersey: Wiley.

Arad, Maya. 1998. *VP-Structure and the Syntax-Lexicon Interface*. PhD dissertation. London: University College of London.

Bader, Markus, and Jana Häussler. 2010. "Word Order in German: A Corpus Study." *Lingua* 120: 717-762.

Bayer, Joseph. 2004. "Non-nominative Subjects in Comparison." In *Non-nominative Subjects* vol. 1, ed. by Peri Bhaskararao, and Karumuri V. Subbarao, 49-76. Amsterdam: Benjamins.

Belletti, Adriana, and Luigi Rizzi. 1988. "Psych-verbs and θ–theory." *Natural Language and Linguistic Theory* 6: 291-352.

Bornkessel, Ina, Stefan Zysset, Angela D. Friederici, D. Yves von Cramon, and Matthias Schlewewsky. 2005. "'Who Does What to Whom?' The Neural Basis of Argument Hierarchies during Language Comprehension." *NeuroImage* 26: 221-233.

Bresnan, Joan 2007. "Is Syntactic Knowledge Probabilistic? Experiments with the English Dative Alternation." In *Roots: Linguistics in Search of its Evidential Base*, ed. by Sam Featherston, and Wolfgang Sternefeld, 75-96. Berlin: Mouton de Gruyter.

Brunetti, Lisa. 2009. "On the Semantic and Contextual Factors that Determine Topic Selection in Italian and Spanish." *The Linguistic Review* 26: 261-289.

Chafe, Wallace. 1976. "Givenness, Contrastiveness, Definiteness, Subjects, Topics and Point of View." In *Subject and Topic*, ed. by Charles N. Li, 27-55. New York: Academic Press.

Dowty, David. 1991. “Thematic Proto-roles and Argument Selection.” *Language* 67: 547-619.

Engelberg, Stefan. 2015. “Gespaltene Stimulus-Argumente bei Psych-Verben. Quantitative Verteilungsdaten als Indikator für die Dynamik sprachlichen Wissens über Argumentstrukturen.” In *Argumentstruktur – Valenz – Konstruktionen,* ed. by S. Engelberg, Meike Meliss, Kristel Proost, and Edeltraut Winkler, 469-492. Tübingen: Narr.

Fanselow, Gisbert. 2000. “Optimal Exceptions.” In *Lexicon in Focus*, ed. by Barbara Stiebels, and Dieter Wunderlich, 173-209. Berlin: Akademie Verlag.

Grafmiller, Jason. 2013. *The Semantics of Syntactic Choice, an Analysis of English Emotion Verbs*. PhD dissertation. Stanford, CA: Stanford University.

Grewendorf, Günther. 1989. *Ergativity in German*. Dordrecht: Foris.

Grimshaw, Jane. 1990. *Argument Structure*. Cambridge: The MIT Press.

Hartigan, John A., and Pamela M. Hartigan. 1985. “The Dip Test of Unimodality.” *Annals of Statistics* 13.1: 70-84.

Haspelmath, Martin. 2001. “Non-canonical Marking of Core Arguments in European Languages.” In *Non-canonical Marking of Subjects and Objects*, ed. by Alexandra Y. Aikhenvald, Robert M.W. Dixon, and Masayuki Onishi, 53–83. Amsterdam and Philadelphia: Benjamins.

Haupt, Friederike S., Matthias Schlesewsky, Dietmar Roehm, Angela D. Friederici, and Ina Bornkessel-Schlesewsky. 2008. “The Status of Subject-Object Reanalyses in Language Comprehension Architecture.” *Journal of Memory and Language* 59: 54-96.

Hoberg, Ursula. 1981. *Die Wortstellung in der geschriebenen deutschen Gegenwarts-sprache*. München: Hueber.

Holisky, Dee A. 1987. "The Case of the Intransitive Subject in Tsova-Tush (Batsbi)." *Lingua* 71:103-132.

Horvath, Julia, and Tal Siloni. 2011. "Causatives across Components." *Natural Language and Linguistic Theory* 29: 657-704.

Institut für Deutsche Sprache. 1991–2017. *COSMAS I/II Corpus Search, Management and Analysis System*. Mannheim: Institut für Deutsche Sprache. http://www. ids-mannheim.de/cosmas2/.

Institut für Deutsche Sprache. 2010. *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2010-I* (Release vom 02.03.2010). Mannheim: Institut für Deutsche Sprache. http://www.ids-mannheim.de/DeReKo.

Keller, Frank, and Antonella Sorace. 2003. "Gradient Auxiliary Aelection and Impersonal Passivization in German: an Experimental Investigation." *Journal of Linguistics* 39.1: 57–108.

Klein Katharina, and Silvia Kutscher. 2002. "Psych-verbs and Lexical Economy." *Theorie des Lexikons 122.* Düsseldorf: University of Düsseldorf.

Lambrecht, Knut. 1994. *Information Structure and Sentence Form. Topic, Focus and the Mental Representations of Discourse Referents.* New York: Cambridge University Press.

Landau, Idan. 2010. *The Locative Syntax of Experiencers*. Cambridge, MA: MIT Press.

Lehmann, Christian. 1991. "Predicate Classes and PARTICIPATION." In *Partizipation. Das sprachliche Erfassen von Sachverhalten*, Hansjakob Seiler, and Waldfried Premper, 183-239. Tübingen: Narr.

Lenerz, Jürgen. 1977. “Zum Einfluß von ‘Agens’ auf die Wortstellung des Deutschen.” In *Grammatik und interdisziplinäre Bereiche der Linguistik. Akten des 11. Linguistischen Kolloquiums Aachen 1976*, ed. by Hans W. Viethen, Wolf-Dietrich Bald, and Konrad Sprengel, 133-142. Tübingen: Niemeyer.

Levin, Beth, and Malka Rappaport Hovav. 1995. *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Cambridge, MA: MIT Press.

Martin, Fabienne. 2015. “Explaining the Link between Agentivity and Non-culminating Causation.” *Proceedings of SALT* 25: 246-266.

Martin, Fabienne, and Florian Schäfer. to appear. “Sublexical Modality in Defeasible Causative Verbs.” In *Modality across Syntactic Categories*. Oxford. OUP.

Pesetsky, David. 1995. “Zero Syntax: Experiencer and Cascades.” Cambridge, MA: MIT Press.

Pijpops, Dirk, and Dirk Speelman. 2015. “Argument Alternations of the Dutch Psych Verbs. A Corpus Investigation.” *Proceedings of the 6th Conference on Quantitative Investigations in Theoretical Linguistics*.

Primus, Beatrice. 2004. “Protorollen und Verbtyp: Kasusvariaton bei psychischen Verben.” In *Semantische Rollen*, ed. by Rolf Kailuweit, and Martin Hummel, 377-401. Tübingen: Narr.

Primus, Beatrice. 2012. “Animacy, Generalized Semantic Roles, and Differential Object Marking.” In *Case, Word Order, and Prominence. Interacting Cues in Language Production and Comprehension*, ed. by Monique Lamers, Peter de Swart, 65-90. Dordrecht: Springer.

Reinhart, Tanya, and Tal Siloni. 2005. “The Lexicon-syntax Parameter: Reflexivization and Other arity Operations.” *Linguistic Inquiry* 36: 389-436.

Reinhart, Tanya. 2001. *Experiencing Derivations*. SALT lecture, New York.

Reinhart, Tanya. 2002. "The Theta System – an Overview." *Theoretical Linguistics* 28: 229-290.

Roeper, Tom. 1987. "Implicit Arguments and the Head-complement Relation." *Linguistic Inquiry* 18: 267-310.

Scheepers, Christoph, Barbara Hemforth, and Lars Konieczny. 2000. "Linking Syntactic Functions with Thematic Roles: Psych Verbs and the Resolution of Subject-Object Ambiguity." In *German Sentence Processing*, ed. by Barbara. Hemforth, and Lars Konieczny, 95-135. Dordrecht: Kluwer Academic Publishers.

Sorace, Antonella. 2004. "Gradience at the Lexicon-Syntax Interface: Evidence from Auxiliary Selection and Implications for Unaccusativity." In *The Unaccusativity Puzzle*, ed. by A. Alexiadou, E. Anagnostopoulou, and M. Everaert, 243-268. Oxford: OUP.

Talmy, Leonard. 1976. "Semantic Causative Types." In *The Grammar of Causative Constructions,* ed. by M. Shibatani, 43-116. New York: Academic Press.

Temme, Anne, and Elisabeth Verhoeven. 2016. "Verb Class, Case, and Order: A Cross-linguistic Experiment on Non-nominative Experiencers." *Linguistcs* 54.4: 769-813.

Van Valin, Robert D., Jr., and Randy LaPolla. 1997. "Syntax: Structure, Meaning, and Function." Cambridge: Cambridge University Press.

Van Valin, Robert D., Jr., and David P. Wilkins. 1996. "The Case for 'Effector': Case Roles, Agents, and Agency Revisited." In *Grammatical Constructions: their Form and Meaning*, ed. by Masayoshi Shibatani, and Sandra A. Thompson, 289-322. Oxford: Clarendon.

Verhoeven, Elisabeth. 2009. "Subjects, Agents, Experiencers, and Animates in Competition: Modern Greek Argument Order." *Linguistische Berichte* 219: 355-376.

Verhoeven, Elisabeth. 2010. "Agentivity and Stativity in Experiencer Verbs: Implications for a Typology of Predicate Classes." *Linguistic Typology* 14: 213-251.

Verhoeven, Elisabeth. 2014. "Thematic Prominence and Animacy Asymmetries. Evidence from a Cross-linguistic Production Study." *Lingua* 143:129-161.

Verhoeven, Elisabeth. 2015. "Thematic Asymmetries Do Matter! A Corpus Study of Word Order in German." *Journal of Germanic Linguistics* 27.1: 45-104.

Wunderlich, Dieter. 1997. "Cause and the Structure of Verbs." *Linguistic Inquiry* 28: 27-68.

Appendix I: Acceptability data

(n = 32 per verb/test)

Verb	intentional adverb		control verb	
	mean	SE	mean	SE
ärgern	6.19	0.5	5.97	0.54
amüsieren	4.28	0.57	5.16	0.57
anwidern	4.16	0.67	3.56	0.56
aufregen	4.56	0.57	4.56	0.5
bedrücken	2.69	0.49	2.53	0.46
befremden	2.78	0.53	2.5	0.58
ekeln	3	0.61	3.25	0.59
empören	4.44	0.57	3.53	0.59
entsetzen	4.19	0.6	3.12	0.51
enttäuschen	5.03	0.54	4.09	0.62
erschrecken	6.19	0.36	5.56	0.55
faszinieren	4.09	0.63	3.5	0.57
freuen	1.34	0.23	1.25	0.2
interessieren	1.88	0.4	1.81	0.38
irritieren	5.5	0.59	4.88	0.6
langweilen	4.81	0.66	5.41	0.54
nerven	5.94	0.47	6.06	0.59
reizen	5.78	0.49	5.53	0.59
überraschen	5.44	0.54	6.03	0.48
wundern	1.28	0.16	1.31	0.26

Appendix II: Corpus data: Frequencies of voice by verb

odds ratio: $(p_{\text{non-active}})/(1 - p_{\text{non-active}})$; log.: common logarithm of the odds ratio

verb	active	non-active				total	% non-active	odds ratio	log.
		reflexive	passive	stative	total				
bedrücken	225	0	2	8	10	235	4.3	0.04	-3.11
reizen	54	0	3	2	5	59	8.5	0.09	-2.38
langweilen	80	4	1	3	8	88	9.1	0.1	-2.3
nerven	54	3	1	8	12	66	18.2	0.22	-1.5
irritieren	147	0	6	38	44	191	23	0.3	-1.21
faszinieren	301	0	2	134	136	437	31.1	0.45	-0.79

anwidern	140	0	1	65	66	206	32	0.47	-0.75
befremden	90	1	0	44	45	135	33.3	0.5	-0.69
erschrecken	127	42	10	19	71	198	35.9	0.56	-0.58
überraschen	168	0	57	83	140	308	45.5	0.83	-0.18
ärgern	126	138	0	0	138	264	52.3	1.1	0.09
wundern	83	104	0	0	104	187	55.6	1.25	0.23
ekeln	110	158	3	1	162	272	59.6	1.47	0.39
interessieren	106	86	0	74	160	266	60.2	1.51	0.41
aufregen	37	56	0	1	57	94	60.6	1.54	0.43
amüsieren	64	95	0	12	107	171	62.6	1.67	0.51
empören	43	37	0	64	101	144	70.1	2.35	0.85
enttäuschen	36	0	7	110	117	153	76.5	3.25	1.18
entsetzen	16	2	0	77	79	95	83.2	4.94	1.6
freuen	42	331	0	0	331	373	88.7	7.88	2.06
Total	2049	1056	93	744	1893	3942			

Appendix III: Corpus data: Frequencies of Orders by verb

odds ratio: $(p_{OS})/(1 - p_{OS})$; log.: common logarithm of the odds ratio

Verb	SO	OS	total	% OS	odds ratio	log.
reizen	52	2	54	3.7	0.04	-3.26
enttäuschen	32	4	36	11.1	0.12	-2.08
entsetzen	14	2	16	12.5	0.14	-1.95
überraschen	147	21	168	12.5	0.14	-1.95
langweilen	68	12	80	15	0.18	-1.73
aufregen	31	6	37	16.2	0.19	-1.64
amüsieren	53	11	64	17.2	0.21	-1.57
erschrecken	105	22	127	17.3	0.21	-1.56
irritieren	119	28	147	19	0.24	-1.45
ärgern	99	27	126	21.4	0.27	-1.3
nerven	42	12	54	22.2	0.29	-1.25
faszinieren	227	74	301	24.6	0.33	-1.12
befremden	67	23	90	25.6	0.34	-1.07
bedrücken	163	62	225	27.6	0.38	-0.97
ekeln	79	31	110	28.2	0.39	-0.94
wundern	57	26	83	31.3	0.46	-0.78
anwidern	96	44	140	31.4	0.46	-0.78

empören	29	14	43	32.6	0.48	-0.73
freuen	27	15	42	35.7	0.56	-0.59
interessieren	60	46	106	43.4	0.77	-0.27
Total	1567	482	2049			