

Engelberg, Stefan (erscheint 2014): *The argument structure of psych-verbs: A quantitative corpus study on cognitive entrenchment*. – Hans Boas & Alexander Ziem (eds.): *Constructional approaches to argument structure in German*. Boston, Berlin: De Gruyter Mouton.

The argument structure of psych-verbs: A quantitative corpus study on cognitive entrenchment

Stefan Engelberg

Draft, Mannheim, 10.9.2013

1 Introduction¹

1.1 Psych-verbs and argument structure alternations

Psych-verbs exhibit a fairly large variation in argument structure patterns, as the following examples from German show. Many of these verbs allow alternations between stimuli and experiencers in subject position (1a vs. 1b), between nominal and clausal realizations of arguments (1a vs. 1c, 1b vs. 1d), between inanimate and agent-like animate stimuli (1a vs. 1e), between simple stimuli and “split-stimuli” that are spread out over two constituents, a subject NP and a PP (1a vs. 1f), and between explicit and implicit argument realization (1b vs. 1g).

- (1) a. Rebeccas bösertige Bemerkung ärgerte Jamaal.
 Rebecca’s malicious remark.NOM anger-3SG.PST Jamaal.ACC
 ‘Rebecca’s malicious remark angered Jamaal.’
- b. Jamaal ärgerte sich über Rebeccas bösertige Bemerkung.
 Jamaal.NOM anger-3SG.PST RFL over Rebecca’s malicious remark.ACC
 ‘Jamaal was/became angry about Rebecca’s malicious remark.’
- c. Dass Rebecca so eine bösertige Bemerkung gemacht hatte,
 that Rebecca had made such a malicious remark
 ärgerte Jamaal.
 anger-3SG.PST Jamaal.ACC
 ‘That Rebecca had made such a malicious remark angered Jamaal.’
- d. Jamaal ärgerte sich (dar-über),
 Jamaal.NOM anger-3SG.PST RFL there-over
 dass Rebecca so eine bösertige Bemerkung gemacht hatte.
 that Rebecca had made such a malicious remark
 ‘Jamaal was/became angry that Rebecca had made such a malicious remark.’
- e. Rebecca ärgerte Jamaal.
 Rebecca.NOM anger-3SG.PST Jamaal.ACC
 ‘Rebecca angered Jamaal.’

¹ I am grateful to Alexander Kopenig for valuable comments on the statistical analyses and for conducting the cluster analysis presented in section 2.4. The two reviewers also provided a number of helpful comments.

- f. Rebecca ärgerte Jamaal mit ihrer böartigen Bemerkung.
 Rebecca.NOM anger-3SG.PST Jamaal.ACC with her malicious remark
 ‘Rebecca angered Jamaal with her malicious remark.’
- g. Jamaal ärgerte sich.
 Jamaal.NOM anger-3SG.PST RFL
 ‘Jamaal was/became angry.’

Traditionally, psych-verbs have been investigated because the alternation between stimulus and experiencer subjects poses some interesting problems for linking theories (cf., e.g., Belletti & Rizzi 1988, Grimshaw 1990, Dowty 1991). Other argument structure alternations with psych-verbs – of which the examples in (1) only represent a small part – have attracted much less attention. In particular, there has been little discussion of why a particular argument structure is chosen and why the alternations differ so strongly with respect to their frequency of occurrence.

The latter question has of course not been addressed because quantitative investigations in general led a marginal existence in the linguistics of the 80s and 90s with its strong tendency to separate competence and performance, system and use. This is currently changing, and the present article adheres to a conception of linguistics that assumes a strong connection between linguistic knowledge and language use.

1.2 Frequency in usage-based linguistics

While to a certain degree, frequency data has always been discussed in linguistics with respect to language acquisition, language processing and language change, the assumption that frequency of use is an important factor in the cognitive representation of synchronic grammatical structure is a more recent development. Within usage-based linguistics, language is seen “as fluid and dynamic, changing through the interaction of social usage events with the cognitive processes characteristic of the human brain in general” (Bybee and Beckner 2010: 854).² This has consequences for the grammatical system: “[...] language structure comes about through the application of a handful of common mechanisms that recur when human beings use language. The domain-general processes of sequential learning, chunking, categorization, and inference-making, along with the effect of partial or complete repetition, lead to the establishment and conventionalization of the categories and structures we find in languages. This bottom-up and emergentist perspective, we argue, may turn out to be indispensable to our understanding of linguistic processes and structure” (Bybee and Beckner 2010: 853).

This conception of linguistics requires weakening or even giving up on the separation between language use and linguistic structure in favor of a dynamic model of grammar (Diessel 2007: 123f.). Thus, recurrence and co-occurrence of linguistic expressions shape our linguistic system. One of the concepts that is closely attached to the recurrence and co-occurrence of expression is cognitive entrenchment. The entrenchment of a linguistic item or pattern into the cognitive system is strongly influenced by its frequency of occurrence and its frequency of co-occurrence with other entities or patterns. This presumes that speakers have linguistic knowledge that is based on a statistical assessment of the input they are confronted with. Ellis (2002) describes this from a psycholinguistic point of view: “[...] psycholinguistic studies of sentence processing show that fluent adults have a vast statistical knowledge about the behavior of the lexical items of their language. They know the strong cues provided by verbs, in English at least, in the interpretation of syntactic ambiguities. Fluent comprehenders know the relative frequencies with which particular verbs appear in different tenses, in active versus passive and in intransitive versus transitive structures, the typical kinds of subjects and

² The basic ideas of usage-based linguistics are discussed, e.g., in Beckner et al. (2009), Bybee (2010), Bybee and Beckner (2010), and Diessel (2011).

objects that a verb takes, and many other such facts. This knowledge has been acquired through experience with input that exhibits these distributional properties and through knowledge of its semantics” (Ellis 2002: 160).

Similarly, corpus linguists who take a cognitive stance towards language argue: “It is common practice in corpus linguistics to assume that the frequency distribution of tokens and types of linguistic phenomena in corpora have – to put it as generally as possible – some kind of significance. Essentially, more frequently occurring structures are believed to hold a more prominent place, not only in actual discourse but also in the linguistic system, than those occurring less often” (Schmid 2010: 101). Thus, cognitively oriented corpus linguists “try to correlate the frequency of occurrence of linguistic phenomena (as observed in corpora) with their salience or entrenchment in the cognitive system. A corollary of this assumption is that patterns of frequency distributions of lexico-grammatical variants of linguistic units correspond to variable degrees of entrenchment of cognitive processes or representations associated with them” (Schmid 2010: 102).

Entrenchment has become quite a popular concept within cognitive and usage-based linguistics. In this paper, entrenchment will be viewed not so much as the absolute strength of the representation of a linguistic item in memory, but rather as a relative notion. I will investigate how strongly a lexical item is associated with a pattern in which it occurs. Thus, the paper addresses the question of whether and how strongly a verb is cognitively entrenched relative to an argument structure pattern and how strongly an argument structure pattern is entrenched with respect to particular verbs. This might be called relative, or associative entrenchment. It is of course undisputed that the relation between quantitative corpus data and cognitive processes as revealed by psycho- and neurolinguistic experiments is not a straightforward one. However, most linguists using the concept of entrenchment agree that there is a strong connection between quantitative corpus data and cognitive processes (cf. Schmid 2010, Blumenthal-Dramé 2012, Gries 2012b). Since I will not be discussing data from experimental studies in this paper, I shall leave the exact nature of the relationship between data from corpus and experimental studies open.

1.3 Quantitative corpus studies on argument structure

The present article starts from the observation that each verb shows particular frequencies of occurrence with respect to its argument structure patterns and each argument structure pattern seems to attract some verbs more strongly than others. The basic assumption is that the observed quantitative distribution patterns can be accounted for, on the one hand, by numerous diverse linguistic factors and, on the other, by basic functions of the human memory system, such as the entrenchment structures as a consequence of recurrent use. I will assume that argument structure patterns constitute entities in our linguistic memory system that accumulate traces of use that determine their variation and diachronic dynamics.

Corpus-based quantitative studies on argument structure have continually – albeit rather infrequently – been published since the 1990s. A number of early studies collected frequency data in order to explain certain phenomena in language processing. MacDonald (1994) and MacDonald, Perlmutter, and Seidenberg (1994) demonstrate how the frequency of argument structures with particular verbs serves to resolve syntactic ambiguities in argument structure processing.

Within research that highlights the role of frequency in the structure of grammar, Gries and Stefanowitsch’s Collostructional Analysis has revealed distribution patterns of verbs and their argument structure (Stefanowitsch and Gries 2003; Gries and Stefanowitsch 2004, 2010; Gries 2011, 2012a). Having measured the association of argument structure constructions to particular verbs, Gries and Stefanowitsch claim that those verbs most strongly entrenched in the construction are those that reflect the meaning of the construction. For example, the verb showing the strongest association to the ditransitive pattern is *give*, which itself lexicalizes the

transfer meaning that Gries and Stefanowitsch attribute to the ditransitive construction (Stefanowitsch and Gries 2003: 228).

Another topic that has occasionally been addressed in quantitative corpus studies on argument structure is the question of how genre influences the distribution of argument structure. Roland & Jurawsky (1998) and Roland (2001) attribute influences of this sort to the preference of particular kinds of discourse for particular verb senses, which, in turn, are associated with different argument structures. Some more recent studies have also been able to show the extent to which the distribution of argument structure patterns is influenced by register and by the distinction between spoken and written language (Gries 2011, Engelberg et al. 2012). To the degree that discourse frequencies govern the degree of entrenchment of a linguistic entity, the dependency of frequencies on genre raises the question of whether entrenchment is a notion that is to be understood relative to particular communicative situations, such as writing a letter, giving a talk, having a chat on the phone. Since frequencies of particular linguistic structures are based on counting linguistic events and events are always contextualized, entrenchment might turn out to be a context-dependent notion.

Other issues that have been addressed are the mechanisms underlying the extension of a construction to new lexical items (cf., e.g., Boas 2011), the productivity of constructions (e.g., Barðdal 2008), the diachronic development of valencies (Köhler 2005), the distribution of valency frames within the lexicon (Steiner 2011, Duwaerts & Ullmann 2013), the role of frequency in argument structure acquisition (Tomasello 2003; Behrens 2011), and the attraction between lexical elements that fill different slots in argument structure patterns (e.g., Engelberg et al. 2011). Schulte im Walde (2003, 2009) employed frequency data in order to automatically induce verb classes from the distribution of valency frames. Some other investigations assessed argument structure frequencies in order to provide norming data for psychological experiments (e.g., Roland et al. 2000; Gahl, Jurafsky, and Roland 2004).

Despite these studies, most of the issues have only been subjected to very few empirical studies, such that there are still a number of open questions:

- (i) The claim that argument structures are most strongly associated with verbs that share their meaning with the construction has only been checked with respect to very few argument structure constructions, such as the ditransitive transfer construction (Stefanowitsch and Gries 2003) and the *as*-predicative (Gries, Hampe, and Schönefeld 2010). More evidence is needed for the kind of relationship between the meaning of a verb, the meaning of an argument structure, and the frequency with which verb and argument structure pattern co-occur.
- (ii) There is also a lack of cross-linguistic studies in order to establish the extent to which language-specific versus cross-linguistic semantic and grammatical conditions determine the quantitative distribution patterns in the domain of argument structures.
- (iii) A large number of other factors can be expected to determine the frequency of argument realization patterns, such as TAM categories, information structure, or the lexical filling of argument slots. Few of these have so far been investigated.³
- (iv) There have also been only very few quantitative investigations into the influence of genre, register, and medium on the frequencies of argument structure, in particular on a fine-grained level, taking into account the whole array of argument realization patterns a verb has to offer.

1.4 Argument structure: basic concepts

Since the terms ‘argument’ and ‘argument structure’ differ widely in their respective uses denoting different semantic and syntactic concepts, the uses of these and related terms in this article shall be outlined here.

³ Cf. Gries (2011) on the influence of tense and aspect on the distribution of argument structure.

Argument structures are meaning representations in which variables for entities (x,y,z) are related by semantic predicates to each other and – at least in the case of verbs – to some situation variable *e* (an event or state variable), e.g. ASSUME(x,y,e). The argument structure is connected to a lexical predicate by a meaning postulate such that all the entities correspond to the arguments of a lexical predicate. A simple example would be:

- (2) a. ASTONISH(x,y,e) → ASTONISHING_ENTITY(e,x) & ASTONISHED_ENTITY(y,x)
 b. ASTONISH(x,y,e) → STIMULUS(e,x) & EXPERIENCER(y,x)

The semantic predicates specify the verb-specific semantic roles (2a) that can be generalized over to a certain degree (2b). The variables in these argument structures are argument variables, and the arguments are the semantic representations of the expressions that specify these arguments in the sentence (cf. Engelberg 2000).

An argument realization pattern (ARP) is an (empirically observed) mapping of syntactic valency features onto a list of arguments:

- (3) NP^{nom} NP^{acc}
 | |
 x^{ASTONISHING_ENTITY} y^{ASTONISHED_ENTITY}

Argument realization patterns are the entities that are counted in the quantitative verb profile studies described in Section 2.

An argument structure pattern (ASP) is a mapping of syntactic valency features onto an argument structure, where (i) some formal or pragmatic-semantic property of this mapping, (ii) some idiosyncrasy with respect to the lexical specification of the argument variables, (iii) the set of verbs that may occur in this pattern, or (iv) an unexpected quantitative distribution of verbs relative to this pattern prohibits an explanation of the distribution and interpretation of this mapping on the basis of independently necessary linguistic or cognitive rules and principles.

This definition emphasizes the unpredictability of many argument structure phenomena and will render many of these phenomena as ASPs in the above sense. These patterns will of course exhibit unpredictable behavior to different degrees, some being more idiosyncratic than others.

1.5 Aims and structure of the article

This article will explore the quantitative distribution of psych-verbs with respect to argument structure patterns. Its aim is to explore some of the linguistic and non-linguistic factors that determine this distribution.

In the foregoing sections, I outlined the major tenets of usage-based linguistics and its relation to usage frequency, and I reviewed some of the quantitative studies that have been conducted on the relationship between verbs and argument structure. In Section 2, I will describe three studies based on the creation and analysis of verb profiles that have been carried out at the Institute of German Language.⁴ Each of the four subsections of Section 3 is devoted to one of the factors that seem to influence the quantitative distribution. The first three are linguistic factors: functional dependencies between different argument structure patterns (3.1), stylistic properties of text genre (3.2), and cross-linguistic conceptual-semantic properties of verbs (3.3). The fourth factor, the tendency to particular forms of cognitive entrenchment, is dependent on general principles of the memory system (3.4). Section 4 summarizes the results.

⁴ Alexander Koplenig, Kristel Proost, Edeltraud Winkler, and, as a cooperating Partner from the University of Bucharest, Ruxandra Cosma contributed to one or another of these studies. Some other results from these studies have been published in Engelberg et al. (2012) and Cosma & Engelberg (2013).

2 Verb profiles

2.1 Study I: German verbs in newspaper texts

The following studies are based on verb profiles.⁵ A verb profile can be defined as a frequency count of the argument realization patterns of the verb based on a random sample of sentences from text corpora. In the different studies, the samples usually consist of either 100 or 200 sentences. The samples are processed by assigning each sentence to the argument realization pattern it represents. The number of sentences for each argument realization pattern of a verb are counted and subjected to statistical analyses.

Since the verb profiles serve to detect verb-specific idiosyncrasies with respect to argument realization patterns, we adopted a rather generous conception of arguments, also including roles that are usually not considered to be arguments but can be assumed to have a verb-specific distribution, such as comitatives, instruments, and benefactives. No statement about the theoretical status of an argument is connected to the descriptive device used for these studies. The list of arguments was constructed while analyzing the sentences from the samples. This procedure led to between 10 and 80 argument realization patterns per verb. For the verb *ärgern* ‘be angry / anger’, the analysis yields eight roles and an expletive pronoun:

- role 1 = person that experiences the anger [EXP]
- role 2 = inanimate stimulus that triggers the anger [STM-I1]
- role 3 = secondary inanimate stimulus (in split-stimuli constructions, cf. Section 3.3) [STM-I2]
- role 4 = animate stimulus (often with agentive interpretation) [STM-A]
- role 5 = comment on the stimulus (a clause or PP that does not strictly realize the stimulus but comments on it, e.g., a *weil/because*-clause) [CMT]
- role 6 = comitative [COM]
- role 7 = direct/indirect speech [SPE]
- role 8 = result (in a resultative construction) [RES]
- E = expletive reflexive pronoun (no reflexive interpretation)

The argument roles determine the columns of our annotation table while each occurring argument realization pattern determines a row in the table. Table 1 shows an excerpt from the verb profile for *ärgern* ‘be angry / anger’.

Table 1: Verb profile for *ärgern* ‘be angry / anger’ and the realization of its arguments (excerpt).⁶

	role 1 EXP	role 2 STM-I1	role 3 STM-I2	role 4 STM-A	role 5 CMT	role 6 OM	role 7 SPE	role 8 RES	E	Fr
ARP-01	NP-nom						dir_sp		<i>sich</i>	30
ARP-02	NP-nom								<i>sich</i>	24
ARP-03	NP-nom	PP-über							<i>sich</i>	47
ARP-04	NP-akk			NP-nom						13
ARP-05	NP-nom	S-wenn						AP	<i>sich</i>	1
ARP-...										

The frequencies were then subjected to several statistical analyses, such as correlation analyses, cluster analyses, multidimensional scaling, and association measures.

⁵ For a similar but multifactorial approach, cf. Gries and Divjak (2009) and Gries (2010).

⁶ ARP = argument realization pattern; dir_sp = direct speech; E = expletive pronoun; Fr = frequency; S = clause.

The data on psych-verbs discussed in this paper comes from three different studies: (i) German verbs in newspaper texts, (ii) German verbs in corpora representing different text genres, and (iii) German psych-verbs and their Romanian counterparts in newspaper texts. All these studies were explorative, non-hypothesis-driven investigations. The first study comprised a number of verb profiles for psych-verbs and other verbs in order to test the method and to detect basic distribution patterns.

2.2 Study II: German and Romanian verbs contrasted

The second study is a contrastive German-Romanian study on psych-verbs. Its aim was to investigate the extent to which cross-linguistic, language-specific, and verb-idiosyncratic parameters determine the distribution of argument structure. For that purpose, ten German verbs and their closest Romanian counterparts were subjected to verb profiling and statistical analysis. The study was carried out on the basis of newspaper corpora. The verbs investigated are listed in Table 2.

Table 2: Verbs investigated in the contrastive German-Romanian study.

German	Romanian	gloss
<i>(sich) amüsieren</i>	<i>a (se) amuza</i>	‘amuse (oneself)’
<i>(sich) ärgern</i>	<i>a (se) supăra</i>	‘be angry / anger’
<i>(sich) aufregen</i>	<i>a (se) enerva</i>	‘get upset / upset’
<i>deprimieren</i>	<i>a (se) întrista</i>	‘depress’
<i>faszinieren</i>	<i>a fascina</i>	‘fascinate’
<i>(sich) freuen</i>	<i>a (se) bucura</i>	‘be happy / please’
<i>(sich) interessieren</i>	<i>a (se) interesa</i>	‘be interested / interest’
<i>schmerzen</i>	<i>a durea</i>	‘hurt’
<i>überraschen</i>	<i>a surprinde</i>	‘surprise’
<i>(sich) wundern</i>	<i>a (se) mira</i>	‘be astonished / astonish’

Some results from this study are presented in Section 3.3, in particular those that shed light on cross-linguistic lexical factors.

2.3 Study III: German verbs and text genres

The third study arose from the assumption that the general-purpose corpus we used for Study I might conceal register-specific argument structure preferences. In order to check this assumption, we chose 16 verbs from five semantic groups:

- A) **Alternating psych-verbs** (*freuen* ‘become/make happy’, *wundern* ‘be astonished/astonish’, *ärgern* ‘get/make angry’) denote a relation between an experiencer x and a stimulus p (essentially a proposition-like entity).
- B) **Connective verbs** (*widersprechen* ‘contradict’, *erklären* ‘explain’, *verursachen* ‘cause’) denote relations between two proposition-like entities; they also allow the realization of NPs expressing human participants, e.g., *widersprechen* ‘contradict’ expresses a relation between a proposition p , uttered by a participant x , and a proposition q , held by a participant y .

- C) **Directed emotion verbs** (*lieben* ‘love’, *hassen* ‘hate’, *bewundern* ‘admire’) denote an emotion between an animate experiencer x and a target of emotion y (an animate being, an object, or a proposition).
- D) **Perception verbs** (*empfinden* ‘feel/sense’, *fühlen* ‘feel’, *hören* ‘hear’) describe a relation between an animate participant x and the participant (or event/proposition-like entity) y that x experiences or becomes cognitively aware of.
- E) **Action verbs** (*arbeiten* ‘work’, *bauen* ‘build’, *kochen* ‘cook’, *malen* ‘paint’) denote a (mostly) physical action of medium complexity performed by an agent x with respect to an object y .

For each of these verbs, we took samples from six corpora representing five different groups of text genres: a newspaper corpus, a corpus of scientific texts, a corpus of general non-fiction texts, two fiction corpora, and a corpus of spoken German. We created a verb profile for each verb and performed a number of statistical analyses (correlations, cluster analysis, multi-dimensional scaling) (for details cf. Engelberg et al. 2012).

2.4 General observations

With all verbs, the distribution of argument realization patterns is heavily skewed to the right. As shown in Figure 1, (*ärgern* ‘be angry / anger’), there are a few patterns that occur very often and many patterns that occur only once or twice in the samples. This is very similar for all the verbs – not only psych-verbs – we have investigated so far.

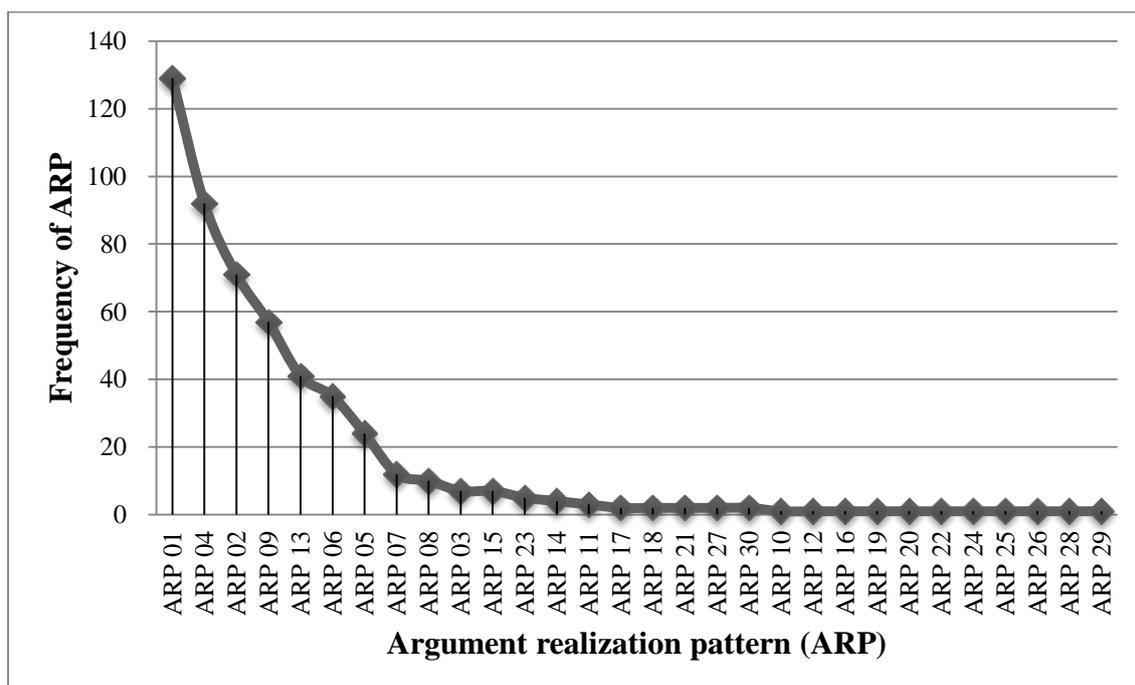


Figure 1: Quantitative Distribution of the 27 argument realization patterns that appear in samples of a total of 518 sentences containing the verb *ärgern* ‘be angry / anger’ (from Study III).

Apart from this general distribution pattern, the single verbs of course differed from each other with respect to the argument realization patterns they preferred. This will be shown in the following for the eleven psych-verbs we investigated, the ten verbs from the contrastive Study II mentioned above, plus the verb *nerven* ‘annoy’. In order to do that, we computed for each pair of the eleven verbs the bivariate (pairwise) correlation coefficients on the basis of the quantitative distribution of the argument realization patterns that occurred with each of the verbs. The higher the correlation coefficient r (between -1 and +1), the more the percentages of patterns for verb A correspond to the percentages of patterns for verb B. The correlation coefficients yield the correlation matrix in Table 3.

Table 3: Correlation matrix for pairs of verbs on the basis of the frequency of argument realization patterns.

	<i>amüsieren</i>	<i>ärgern</i>	<i>aufregen</i>	<i>deprimieren</i>	<i>faszinieren</i>	<i>freuen</i>	<i>interessieren</i>	<i>nerven</i>	<i>schmerzen</i>	<i>überraschen</i>	<i>wundern</i>
<i>amüsieren</i> 'amuse'	1.00										
<i>ärgern</i> 'be angry / anger'	0.71	1.00									
<i>aufregen</i> 'upset / get upset'	0.83	0.85	1.00								
<i>deprimieren</i> 'depress'	0.18	0.25	0.32	1.00							
<i>faszinieren</i> 'fascinate'	0.19	0.16	0.28	0.91	1.00						
<i>freuen</i> 'be happy / please'	0.50	0.76	0.66	0.04	0.02	1.00					
<i>interessieren</i> 'be interested / interest'	0.07	0.09	0.14	0.46	0.44	0.00	1.00				
<i>nerven</i> 'annoy'	0.16	0.17	0.24	0.75	0.74	0.02	0.31	1.00			
<i>schmerzen</i> 'hurt'	0.12	0.10	0.18	0.66	0.61	-0.01	0.25	0.71	1.00		
<i>überraschen</i> 'surprise'	0.13	0.18	0.21	0.77	0.74	-0.01	0.32	0.71	0.53	1.00	
<i>wundern</i> 'be astonished / astonish'	0.46	0.68	0.60	0.07	0.01	0.63	0.01	0.04	0.02	0.02	1.00

On the basis of this correlation matrix, an agglomerative hierarchical cluster analysis was carried out that measured the similarity between the eleven verbs with respect to the quantitative distribution of their argument realization patterns and grouped verbs within the binary-branching cluster tree in Figure 2.⁷ The more similar the correlation values of two verbs are, the closer the two verbs are linked in the tree and the more to the left the branching node is situated, the higher the similarity.

⁷ This analysis used the complete linkage algorithm with the correlation coefficients as measures of similarity. It is important to note that the order in the dendrogram from top to bottom is arbitrary. For example, the distance matrix (cf. Table 3) reveals that the fact that *aufregen* is arranged right after *freuen* in Figure 2 does not imply that *aufregen* and *freuen* are more similar than *ärgern* and *freuen*. I thank one anonymous reviewer for pointing that out.

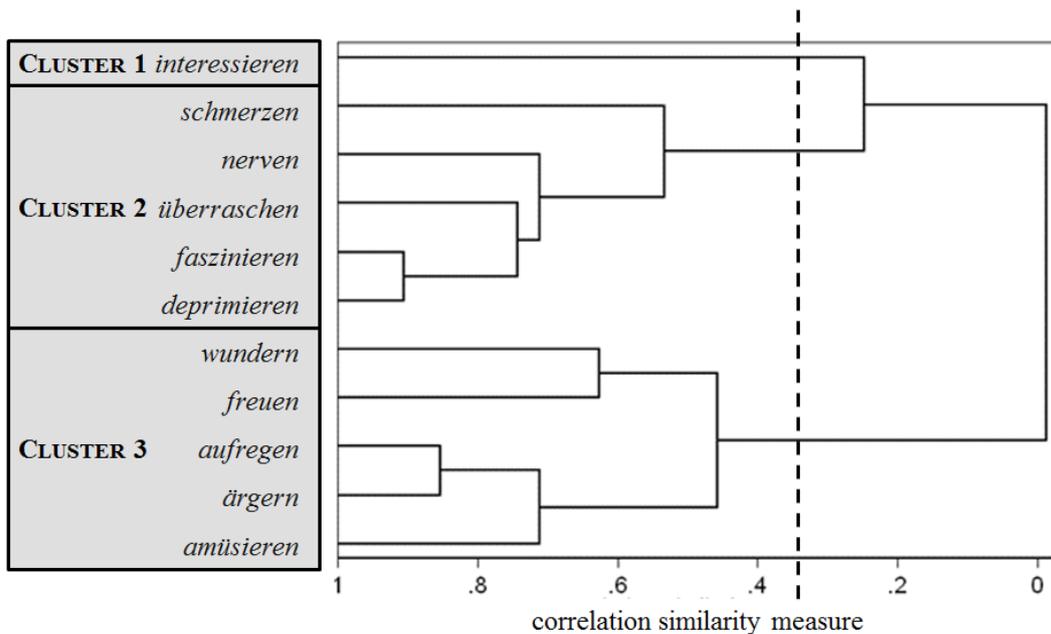


Figure 2: Cluster analysis on the basis of the correlation matrix in Table 3.

If we cut through the tree in its lower third, the tree gives rise to two observations. Firstly, the tree reveals three main clusters that correspond to the three main types of argument structures that characterize the three classes.⁸ The two main classes are represented by Clusters 2 und 3. Cluster 3 comprises those psych-verbs that allow an alternation between a pattern with the stimulus as subject and the experiencer as direct object and a pattern with the experiencer as subject and the stimulus as a PP headed by the preposition *über* ‘over’ (2). Cluster 2 assembles those verbs that are restricted to the first of these patterns (3). The verb *interessieren* in Cluster 1 finally is representative of those verbs that express a sort of mental focusing on some state of affairs. These verbs do allow an alternation similar to that in Cluster 3 but use the preposition *für* ‘for’ (4).

- (2) a. Seine dumme Bemerkung ärgerte sie.
his stupid remark.NOM angered her.ACC
‘His stupid remark made her angry.’
b. Sie ärgerte sich über seine dumme Bemerkung.
she.NOM angered REFL over his stupid remark
‘She got angry at his stupid remark.’
- (3) a. Seine dumme Bemerkung überraschte sie.
‘His stupid remark surprised her.’
b. *Sie überraschte sich über seine dumme Bemerkung.
‘She got surprised at his stupid remark.’
- (4) a. Seine dumme Bemerkung interessierte sie.
‘Her stupid remark interested her.’
b. Sie interessierte sich für seine dumme Bemerkung.
‘She was interested in his stupid remark.’

Secondly, the correlation matrix shows that psych-verbs are a very heterogeneous group with respect to the quantitative distribution of argument realization patterns. While some pairs

⁸ A further analysis of the clustering reveals that a three-cluster solution is the best clustering of the data, as indicated by the Calinski-Harabasz pseudo-F index ($F = 7.27$), closely followed by a two-cluster solution ($F = 7.22$).

of verbs show a moderate to high similarity with correlation coefficients of up to 0.85, other pairs yield correlation coefficients of around zero.⁹

3 Distribution factors

3.1 Functional dependencies between argument realization patterns

As we have seen in (2) through (4), some psych-verbs allow an alternation between an external realization (i.e., as subject) of the stimulus argument and an external realization of the experiencer argument. This alternation has been used as crucial data in theories on argument linking and verbal aspect. Sometimes it was assumed that the stimulus-as-subject variant – in contrast to the experiencer-as-subject variant – has to be interpreted as causative. Linking of the stimulus to the subject position was then attributed to its causative nature (e.g., Grimshaw 1990). Other approaches assumed that both psych-verb variants were basically semantically equivalent (Dowty 1991).

The frequency data from Study II shows that the stimulus-as-subject and the experiencer-as-subject variant are not evenly distributed:

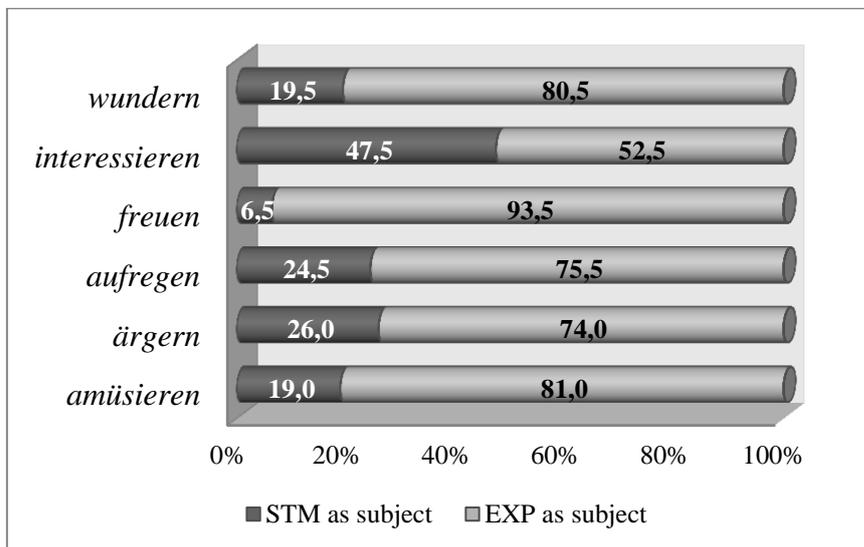


Figure 3: Proportion of experiencer-as-subject and stimulus-as-subject examples in the German samples (active voice).

All together, less than a quarter of the example sentences of the sample show the stimulus as subject; the other examples realize the experiencer in subject position. In the following, it will be shown how the frequencies of two other phenomena, namely sentential subjects and passive constructions, are related to this data.

Stimulus arguments can not only be realized as NPs but also as sentential complements. The complement sentence as a sentential subject (5) or as a sentential object (6) can appear preverbally or at the end of the sentence. As a sentential object, the preverbal position requires the occurrence of a prepositional correlate (*darüber*).

⁹ One anonymous reviewer asked the reasonable question why raw frequencies were correlated instead of log frequencies. Since the data contains many “zero” observations, a log-transformation would have lead to many missing values because $\log(0) = -\infty$. In principle, this could be avoided by fudging the whole data set (for example by adding 1 to the count before the log-transformation). However, as O’Hara & Kotze (2010) show, this is seldom a good strategy. Nevertheless, after doing a log-transformation and re-running the cluster analysis, the results remained almost identical. The only exception is that the Calinski-Harabsz pseudo-F index (cf. Footnote 8) now favors a two-cluster solution ($F = 6.50$) instead of the three-cluster solution presented in the text, with Cluster 1 and Cluster 2 (cf. Figure 2) as one distinct cluster and Cluster 3 as the other one. A three-cluster solution receives the second highest value ($F = 4.47$).

- (5) a. Dass er kommen wollte, freute sie.
 that he wanted to come please-3SG.PST her.ACC
 b. Es freute sie, dass er kommen wollte.
 it please-3SG.PST her.ACC that he wanted to come
 ‘It pleased her that he wanted to come.’
- (6) a. Sie freute sich (darüber), dass er kommen wollte.
 she.NOM please-3SG.PST RFL over it that he wanted to come
 b. Dar-über, dass er kommen wollte, freute sie sich.
 there-over, that he wanted to come please-3SG.PST she.NOM RFL
 ‘It pleased her that he wanted to come.’

In general, it seems to be assumed that complement clauses are realized more frequently as objects than as subjects or at least that they play a more important role in language use and the language system. This impression is reinforced by the fact that both large typological studies on complement sentences, Noonan (1985) and Cristofaro (2003), do not take sentential subjects into consideration at all.

The tendency of subjects to attract NPs that have animate referents and the restrictions of the subject roles of many verbs to agents explain why many verbs do not select sentential subjects. However, sentential subjects are not as marginal as their treatment in linguistics might suggest. Salkoff (2002:132f.) lists 700 psych-verbs in English that allow sentential subjects. A cursory look at his verb list suggests that the German equivalents of most of the English verbs also allow a sentential subject. Yet, the data from Study II seems to show that, even with psych-verbs, sentential objects are preferred to sentential subjects. In our samples, sentential objects occur twice as often as sentential subjects. One might speculate whether this dispreference for sentential subjects is due to processing differences that might go back to the conflict between the postverbal position as the preferred position for complement sentences (cf. Dryer 1980) and the preverbal position as the preferred position for subjects.

However, citing the higher numbers of object sentences in our samples does not take into consideration the different preferences for the stimulus-as-subject and experiencer-as-subject variants (Figure 3). If we compute the proportion of subject sentences of all stimulus-as-subject variants and the proportion of object sentences of all experiencer-as-subject variants, the picture changes (Figure 4).

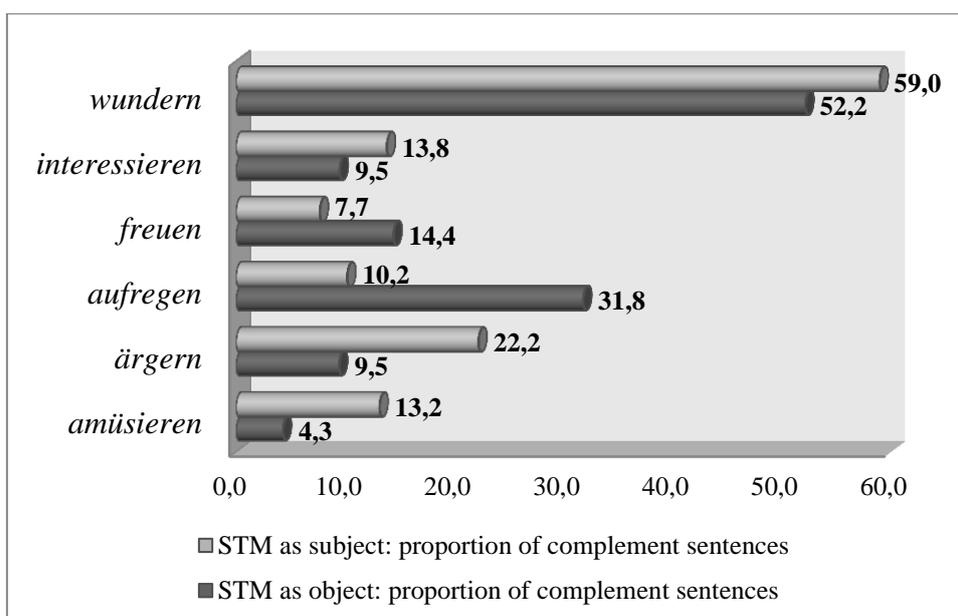


Figure 4: Proportion of complement sentences of all stimulus-as-subject versus stimulus-as-object sentences (only for those verbs that allow both variants).

conditioning in his Boeing during a trip to Egypt in 1982 (while the journalists, traveling in a more modern plane, felt comfortably cool).’

[M01/SEP.66530 Mannheimer Morgen, 07.09.2001; Sparsamkeit ist vorgeschrieben]

- b. **Du** wachst morgens auf, bist eigentlich guter Laune, es ist ein schöner heller Tag –

und dann musst **du** dich hinsetzen und dich

and then must-3SG.PRS you RFL sit down and RFL

über irgendetwas ärgern.

over something anger-INF

‘**You** wake up in the morning, you are actually in a good mood, it is a beautiful bright day – and then **you** have to sit down and get angry about something.’

[M00/APR.08355 Mannheimer Morgen, 08.04.2000; Über die fehlenden Reize der Tagespolitik]

- c. Von Fernseh-Werbespots fühlen sich danach **zwei Drittel** der Verbraucher gestört.

Ein Drittel ärgert sich über Anzeigenwerbung in Zeitschriften.

One third anger-3SG.PRS RFL over ads in journals

‘**Two thirds** of consumers feel annoyed by TV commercials. **One third** are angry about ads in journals.’

[F05/507.31913 Frankfurter Allgemeine, 26.07.2005; Verbraucher ärgern sich über Werbung]

In summary, the psych-verb alternation is relevant for information structure; it serves to promote an argument to subject position in order to pick up the theme of the preceding discourse.¹¹

This section has shown that the frequency of one argument realization pattern often depends on the frequency of another pattern in a meaningful way. The frequency of complement sentences depends on the frequency of the two basic argument realization patterns for psych-verbs, which in turn depends on general preferences for particular semantic role configurations. The frequency of passive sentences and that of active experiencer-as-subject sentences depend on each other since they carry the same functional load.

3.2 Influences of register

As described in Section 2.3, we investigated the distribution of argument realization patterns across register and medium. Some of the results were as follows (cf. Engelberg et al. 2012):

- While some verbs hardly show any cross-corpus differences with respect to the distribution of argument realization patterns, other verbs differ widely across corpora, that is, across register and medium.
- Often particular stylistic properties attached to registers account for the distribution of argument realization patterns, that is, the tendency to employ many passives and few addressee datives in scientific texts is the result of a tendency for impersonal descriptions in scientific texts.
- Sometimes, particular verbs show a strong association with a particular pattern in a particular register. For example, there is a high proportion of impersonal passives with *arbeiten* ‘work’ in scientific texts due to expressions such as *Über das Thema wurde viel gearbeitet* (‘That topic has been worked on extensively’).
- Some types of argument realization patterns show a strong tendency towards low cross-corpus correlations, for example addressee and benefactive datives or patterns involving direct speech.

Among the classes of verbs we investigated, there were two groups of psych-verbs, three “alternating psych-verbs” (*freuen* ‘become/make happy’, *wundern* ‘be astonished/astonish’, *ärgern* ‘get/make angry’), and three “directed emotion verbs” (*lieben* ‘love’, *hassen* ‘hate’,

¹¹ However, this hypothesis is not corroborated by quantitative data and calls for a more thorough corpus study.

bewundern ‘admire’). This is obviously too small a basis to yield representative results for the whole class of psych-verbs. However, I will sketch some results here, not published in Engelberg et al. (2012), which might give some preliminary answers to the question of how register, verb, verb class, and language-specific factors interact with respect to the distribution of argument realization patterns.

There is one striking difference between the two subclasses of psych-verbs. The three directed emotion verbs show hardly any cross-register variance. We compared the frequencies of the argument realization patterns for each verb in the six corpora. This yielded 15 correlation coefficients for each verb and the pairwise comparison of the six frequency lists. The correlation coefficients for the verbs *lieben* and *hassen* were almost perfect (0.99 or 1.00); the coefficients for *bewundern* were only slightly lower (10 coefficients > 0.95 and 5 slightly below 0.90). That is, no matter what corpus/register is chosen, the verbs always show almost the same distribution of argument realization patterns. The picture for the alternating psych-verbs is different: *freuen* exhibits correlation coefficients of between 0.36 and 0.89, *ärgern* between 0.64 and 0.94, and *wundern* between 0.21 and 0.88. These verbs only show a moderate cross-register correspondence with respect to argument realization frequencies. With *wundern* and *freuen*, the lowest coefficients show up when the spoken language data is compared with the written corpora.¹² *Freuen* shows noticeable distribution differences in the corpora of scientific texts and of spoken language. But even with respect to all other corpus pairs, the three alternating psych-verbs show less correspondence between frequencies than the directed emotion verbs.

In order to reveal the interaction between register and other factors, I will now take a closer look at the alternating psych-verbs and three argument structure patterns they occur in: (i) the pattern where the experiencer is realized as subject and the stimulus as a PP headed by *über* ‘over’ (9a), (ii) the introduction of direct speech by psych-verbs (9b), and (iii) the realization of an inanimate stimulus as subject (9c):

- (9) a. Sie ärgerte / freute / wunderte sich über ihn.
 she anger- / please- / astonish-3SG.PST RFL over him
 ‘She was angry at him / happy about him / astonished about him.’
- b. „Das habe ich nicht erwartet,“ ärgerte / freute / wunderte sie sich.
 „I didn’t expect that,“ anger- / please- / astonish-3SG.PST her RFL
 ‘„I didn’t expect that,“ she said with anger / pleasure / astonishment.’
- c. Das / Dass er das tat / Sein Verhalten ärgerte / freute / wunderte sie.
 that / that he did that / his behavior anger- / please- / astonish-3SG.PST her
 ‘That / That he did that / His behavior angered / pleased / astonished her.’

Figure 5 exhibits the proportion of the argument realization patterns corresponding to pattern (i) (experiencer as subject, stimulus as PP-*über*).¹³

¹² Cf. the multi-dimensional scaling in Engelberg et al. (2012) that allows us to map the differences between corpora for verbs and verbs groups.

¹³ An anonymous reviewer criticized the fact that Figures 5-7 present proportional data without any indication of significance. However, in accordance with Alexander Koplenig, who assisted with the statistical analyses of the data, this was done on purpose: essentially, statistical significance testing quantifies how likely it is that a pattern that is found in a sample will also be found in the population. The most crucial assumption in this context is that the sample is taken randomly from the population; otherwise it would not be representative of the population, so a significance test would not be valid. Because it is very doubtful that the sample in this case (which consists of sentences taken from six different corpora representing different groups of text genres, cf. Section 2.3) is unbiased, for practical and legal reasons (for example, a secret conversation between two diplomats will most likely not be found in a corpus of spoken language), it is more adequate to present the data without significance values.

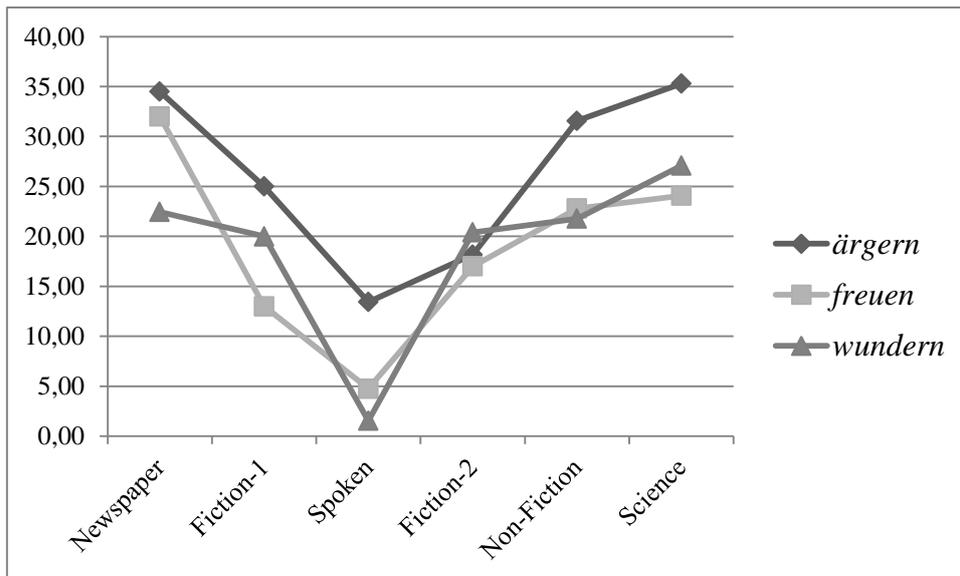


Figure 5: Proportion of argument realization patterns with experiencer as subject and stimulus as PP-über in the three verb samples.

Figure 5 shows that, regardless of the verb, the pattern is not preferred in spoken language and occurs most often in newspaper texts and – to a lesser degree – in other non-fictional registers. Yet, regardless of register restrictions, there is a stronger overall preference of *ärgern* 'be angry / anger' for this pattern relative to the other two psych-verbs.

The distribution of pattern (ii) (psych-verbs introducing direct speech) also shows how register-specific and verb-specific preferences interact (Figure 6).

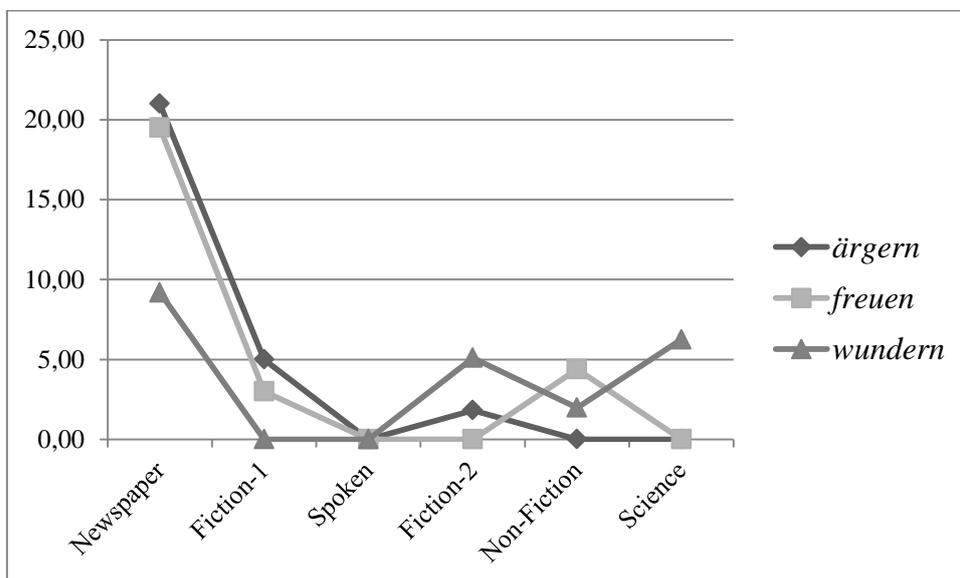


Figure 6: Proportion of argument realization patterns in which the psych-verb introduces direct speech.

What can be seen in Figure 6 is that psych-verbs are not used in spoken language to introduce direct speech. In fiction, and in scientific and other non-fiction texts, psych-verbs occur with direct speech only rarely, while in newspaper texts this usage is widespread. However, as we have seen with respect to pattern (i), all verbs display this tendency, but to a different degree. The proportion of examples in newspaper texts with *ärgern* and *freuen* is twice as high as with *wundern*. A third factor comes into play if we look at newspaper data cross-linguistically. Our contrastive German-Romanian study showed that German employs psych-verbs to introduce direct speech more than six times as often as Romanian does (Cosma

& Engelberg 2013), and an ongoing German-Spanish study shows that this use of psych-verbs is even more rarely attested in Spanish newspaper texts.¹⁴

Pattern (iii) (inanimate stimulus as subject) points to another kind of interaction between register- and verb-specific peculiarities.

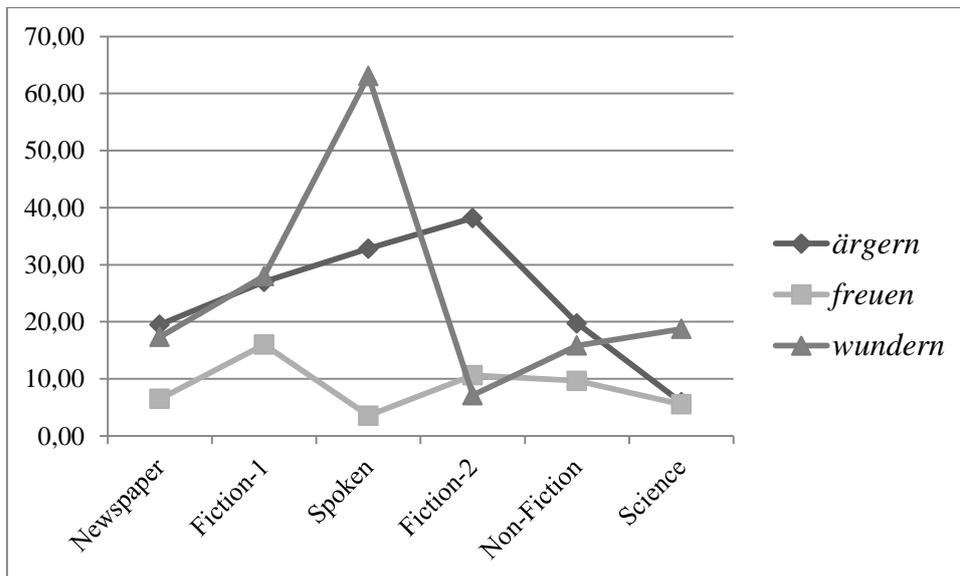


Figure 7: Proportion of argument realization patterns with an inanimate stimulus as subject.

Since all active sentences with these three verbs have to realize either the stimulus or the experiencer as subject, the complement of the percentages in Figure 7 indicates the proportion of experiencers as subject. That is, Figure 7 reflects the fact that – as we have seen in Section 3.1 (Figure 3) – experiencer as subjects are more frequent than stimuli as subjects (approx. 75.0% versus 25.0%). The data also confirms a verb-specific result from the investigation presented above, namely that *freuen* is even less associated with the stimulus-as-subject variant than other psych-verbs. As Figure 7 shows, this also holds for spoken language and other written registers besides newspapers. What is particularly striking is that *wundern* exhibits an extremely high number of stimuli as subjects in spoken language. This is mainly due to examples such as the following:

- (10) a. ja das war äh an der Grenze zu Belgien also von hier zirka fünfundachtzig Kilometer ungefähr **das hat mich auch gewundert** ich dachte eigentlich da würden mehr deutsch sprechen
 ‘yes that was at the border with Belgium, that is about eighty-five kilometers from here, **that astonished me, too**, actually I thought more people would speak German there’
 [E:\IDS\KorporaGS\Dh_IV\PRM1_IV.TextGrid]
- b. **das hat mich gewundert** nein das hat mich nicht gestört nein wir saßen alle am Tisch haben gegrillt und haben uns halt unterhalten
 ‘that astonished me, no, that didn’t bother me, no, we all sat at the table, barbecued, and talked’
 [E:\IDS\KorporaGS\Dh_IV\COB1_IV.TextGrid]
- c. ja **das hat mich auch unwahrscheinlich gewundert** ja
 ‘yes that astonished me a great deal, too, yes’
 [E:\IDS\KorporaGS\Dh_IV\AUG1_IV.TextGrid]

In spoken discourse, the almost formulaic expression *das hat mich gewundert* is used frequently and serves to express astonishment about a state of affairs that is at stake in the

¹⁴ This study is currently being carried out by Stefan Engelberg, Meike Meliss, and Paloma Sánchez Hernández.

conversation. Since this presupposes that the state of affairs, which constitutes the stimulus of *wundern*, has already been introduced into the discourse, the stimulus as topic usually occurs in subject position.

3.3 Cross-linguistic conceptual-semantic properties of verbs

One of the aims of the contrastive German-Romanian study was to find out whether German verbs and their Romanian equivalents show cross-linguistically similar preferences for particular argument realization patterns.

In each of the two languages, we observed that (i) argument realization patterns were distributed very unevenly for each verb (cf. Figure 1) and that (ii) even within a class of semantically similar verbs like psych-verbs, these verbs differed widely with respect to the distribution of argument realization patterns. Thus, we also expected there to be quite a large number of verb-specific idiosyncratic differences between each German verb and its Romanian equivalent. However, this assumption turned out to be wrong. This will be illustrated with respect to three phenomena: split stimuli, sentential subjects and experiencers as subjects.

Stimulus arguments are basically propositional arguments (for evidence cf. Cosma & Engelberg 2013). This allows them to take various shapes, among them different kinds of complement sentences and NPs denoting propositional entities. Another way to reflect the propositional nature of the stimulus is to split up the argument into an NP and a PP constituent. The two basic variants of this construction are illustrated in (11):

- (11) a. Im Lager der gedemütigten Asiaten, [die]_{EXP} [der Doppel-Vizeweltmeister]_{STM_part_1} [mit verfeinerten und neuen Schlagtechniken]_{STM_part_2} überrascht hatte, erteilte Trainer Liu Guoliang dem Triumphator einen Ritterschlag.
 ‘In the camp of the humiliated Asians whom [the two-time runner-up in the world championships]_{STM_part_1} had surprised [with refined and new hitting techniques]_{STM_part_2} the champion received the accolade from coach Liu Golang.’
 [Mannheimer Morgen, 25.11.2005]
- b. [Ihr aktuelles Album „Brother, Sister, Bores“]_{STM_part_1} fasziniert [mit zerbrechlichen Pianopassagen]_{STM_part_2} und schlägt mit kantigen Rockriffs feine Macken in die Gehörgänge.
 ‘[Their current Album “Brother, Sister, Bores”]_{STM_part_1} fascinates [with fragile piano passages]_{STM_part_2} and chisels fine scratches into the ear canals with edgy rock riffs.’
 [Hamburger Morgenpost, 8.9.2006]
- c. Das Japanische klingt rhythmisch, das Philippinische melodisch,
 [am Polnischen]_{STM_part_1} faszinier-ten [Gal]_{EXP}
 at the Polish fascinate-3PL.PST Gal.ACC
 [„die hochfrequenten Zischlaute“]_{STM_part_2}.
 the highly frequent sibilants.NOM
 ‘Japanese sounds rhythmical, Filipino melodic, Polish fascinated Gal with highly frequent sibilants.’
 [Berliner Zeitung, 12.12.2003]

Other prepositions, for example, *durch* ‘through’ instead of *mit* ‘with’ or *bei* ‘by’ instead of *an* ‘at’, also occur with slightly different shades of meaning. Romanian shows pretty much the same array of split stimulus patterns. Looking at the preferences of German verbs and their Romanian counterparts for split stimuli, the following picture emerges:

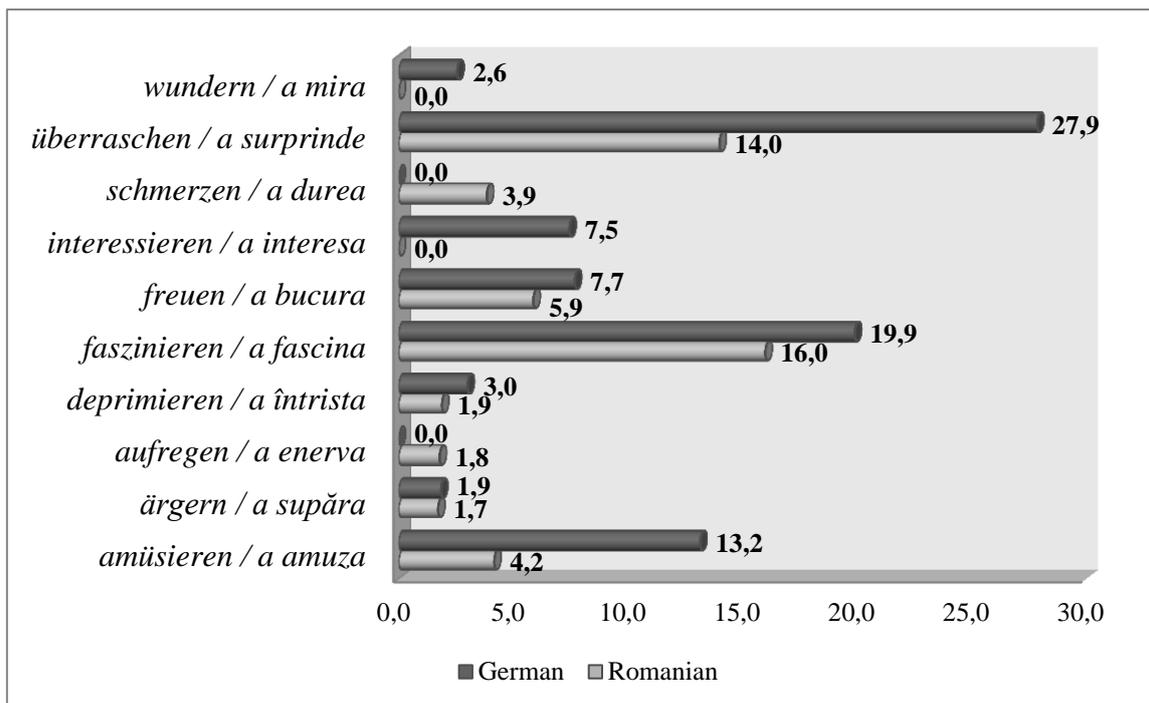


Figure 8: Proportion of examples exhibiting split stimuli in the German and Romanian verb profiles.

Correlating the two paired vectors of frequency data for German and Romanian yields a high correlation coefficient of 0.86. In particular, those verbs expressing surprise and fascination are strongly associated with the split stimulus pattern in both languages. A closer look at the data shows that *überraschen* ‘surprise’ tends to realize the variant exemplified in (11a) and *faszinieren* ‘fascinate’ the variant in (11c).

The stimulus can also be realized as a sentential subject. This is not a preferred option. Only 7.1% of the German sentences and 10.6% of the Romanian sentences exhibit sentential subjects. Again, the tendency to select sentential subjects is very similar for German verbs and their respective Romanian equivalents (Figure 9). Only the verb expressing interest contradicts this generalization. Excluding *interessieren / a interesa*, the correlation coefficient is 0.88; including it, the correlation coefficient drops to 0.37. Looking more closely at the data, it can be seen that this difference is mainly due to a large number of indirect interrogative clauses functioning as the sentential subject for the Romanian verb *a interesa*, which is the only verb in the Romanian samples that shows this pattern. It serves – to a much higher degree than its German counterpart – as a kind of interrogative verb.

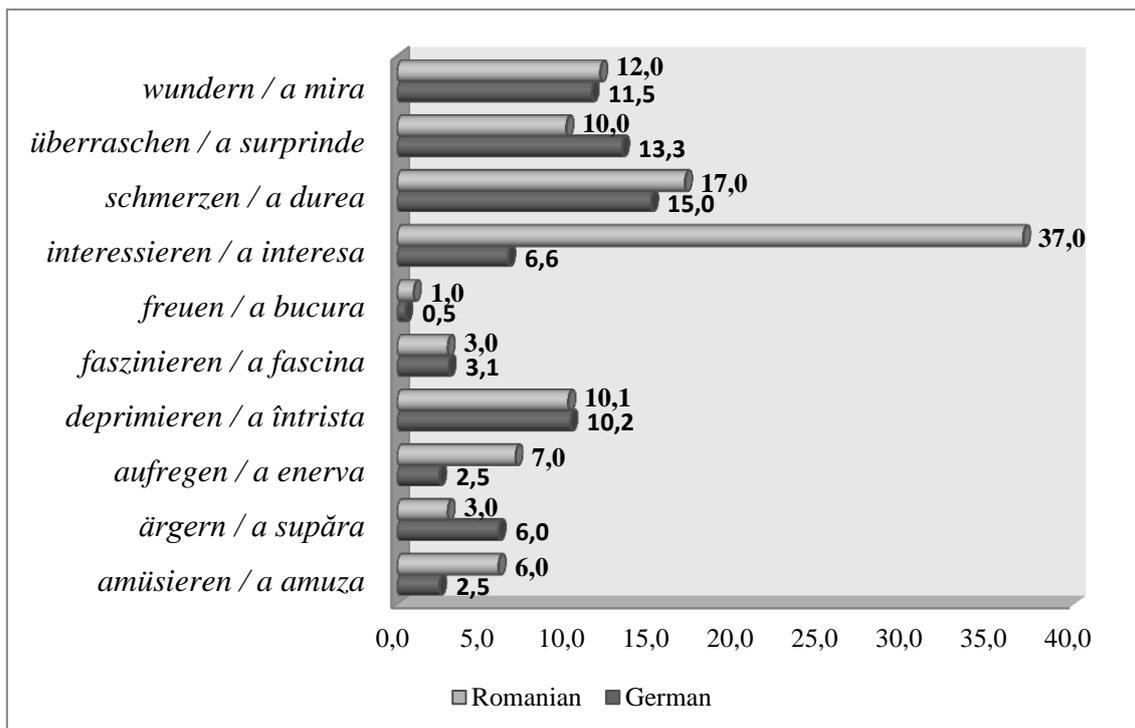


Figure 9: Proportion of examples exhibiting sentential subjects in the German and Romanian verb profiles.

A comparable observation can be made for the tendency of verbs to select inanimate versus animate, agent-like stimuli (e.g., *it surprised me* versus *he surprised me*). German verbs and their Romanian equivalents have similar affinities towards one or the other variant, except for one verb: *a supăra* ‘be/get/make angry’ shows a much stronger tendency towards animate stimuli in Romanian (79.0% of all stimulus-as-subject sentences) than its German counterpart *ärgern* (29.0%). This is due to the fact that *a supăra* has agentive readings not available in German as ‘to hinder / disturb’ and ‘to afflict / plague’.

The argument realization patterns for each psych-verb can be exhaustively classified into experiencer-as-subject and stimulus-as-subject variants (ex. 2 vs. 3). As Figure 3 has shown, German is characterized by a strong tendency towards experiencer-as-subject patterns. Romanian does not show this tendency: averaged over all samples, experiencer-as-subject and stimulus-as-subject variants are evenly distributed (49.9 % experiencer as subject, 50.1% stimulus as subject):

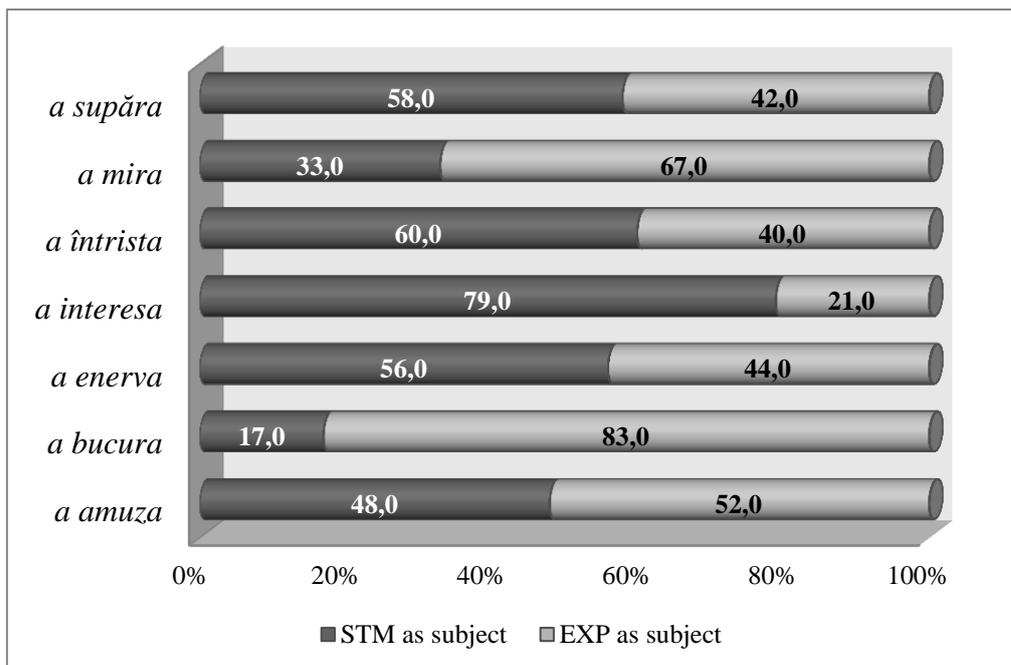


Figure 10: Proportion of experimenter-as-subject and stimulus-as-subject examples in the Romanian samples (active voice).

However, even though both languages show very different preferences for the two types of argument realization patterns, the individual preferences of the German verbs and their Romanian counterparts are very similar. The correlation coefficient for the two rows of frequencies for the German and Romanian verbs is 0.94. Of all verbs, the Romanian verb *a interesa* and its German counterpart *interessieren* ‘interest’ show by far the strongest tendency towards stimulus subjects while *a bucura* / *freuen* ‘be/get happy’ has the strongest affinity to experimenter subjects.

We have seen in Section 2.4 that psych-verbs in a given language have very different preferences for particular argument realization patterns. However, cross-linguistically, verbs and their counterparts often show a very similar association strength with respect to particular patterns. However, semantic correspondence between two verbs of different languages only holds to a certain degree. In particular, a verb may have a function, a reading or a particular shade of meaning which is not available for its cross-linguistic counterpart. This often occurs in particular constructions, as we have seen for the Romanian *a interesa* ‘interest’ and its selection of indirect interrogatives as subject and *a supăra* ‘be/get/make angry’ and its tendency to combine with animate stimulus subjects. Apart from that, all patterns we looked at cross-linguistically exhibited clear positive correlations with respect to the verb’s affinity to particular argument realization patterns.¹⁵

Thus, German verbs and their Romanian counterparts exhibit similar preferences for argument realization patterns in spite of strong language-internal distribution differences between verbs and in spite of cross-linguistic differences with respect to the preference for particular argument realization patterns. This points to a strong influence of cross-linguistic verbal concepts. Particular concepts like FASCINATION, INTEREST, or SURPRISE seem to attract particular argument realization patterns independent of the particular language.¹⁶

¹⁵ The only pattern not clearly corresponding to this observation was the use of psych-verbs for the introduction of direct speech. However, since German employs this function very often while the Romanian samples produced only a handful of examples, we dispensed with a correlation analysis.

¹⁶ “Cross-linguistic” is, of course, not to be understood as “universal”. On the basis of two languages, that would be a bold claim. One reviewer remarked that the observed similarity between German and Romanian might as well be due to the fact that both languages are genetically related or that they are geographically not very distant.

3.4 Entrenchment

Even though the linguistic factors discussed in Sections 3.1, 3.2, and 3.3 explain some of the properties of the frequency distributions, the strong adherence of an argument structure pattern to one or very few verbs is determined by non-linguistic factors, too. One of these factors is entrenchment. The entrenchment of patterns is probably one of the general – and not specifically linguistic – concepts that explain the organization and processes of our memory. The degree of entrenchment of an item within a pattern shall be understood here as the degree of association of this item with the pattern; this is – at least partly – determined by the frequency of events in which this pattern occurs. The cognitive entrenchment of argument structure patterns is facilitated by their strong association with very few verbs. These verbs seem to function as “memory anchors” that help to ground the argument structure pattern in our memory system. I will close this article by measuring the strength with which these anchor verbs ground their argument structure patterns.

Which measure adequately captures these kinds of associations is the subject of some debate (cf. Schmid 2010: 125f.).¹⁷ One of the measures that is assumed to reveal cognitive entrenchment is ΔP , which is usually employed in psychological learning and conditioning theory (Ellis 2006: 10) and applied to corpus linguistics by Baayen (2011).¹⁸ ΔP measures the strength with which a particular outcome can be expected as a reaction to a particular cue. It

However, considering the rather moderate genetic relatedness of the two languages, it seems unlikely to me that preferences concerning the realization of verbal concepts can be traced back to the common origin of both languages. Language contact might be taken into consideration, in particular the moderate lexical influence of German on Romanian due to the presence of a large German minority in Romania since medieval times. However, the verbs under investigation are not involved in borrowing processes between the two languages. The possibility that the concepts expressed by the Romanian verbs or particular usage preferences might have been influenced by the corresponding German verbs cannot of course be completely excluded. If there is an influence on the similar behavior of German and Romanian with respect to argument structure preferences, I would rather suspect that the meaning and use of the verbs might have been shaped by the European culture both languages participate in. As Reichmann (2012) has shown, the common cultural ancestry of European languages, shaped in particular by Latin, its scriptures and the translations of Latin texts into many languages, had a large influence on the lexicon of European languages, not necessarily as a result of words being borrowed from Latin or other languages but mainly due to the influence on the semantic structures of words. However, it will be difficult to show empirically whether this had an influence on the distribution of the argument structure distribution in our samples.

¹⁷ Schmid (2010) reminds us of the work that still has to be done in this respect: “[...] so far we have understood neither the nature of frequency itself nor its relation to entrenchment, let alone come up with a convincing way of capturing either one of them or the relation between them in quantitative terms. This remains true in spite of the indisputable advantages of quantitative methods such as their predictive power, the possibility to falsify models by means of repeat analysis and their enormous capacity when it comes to coming to grips with highly multivariate datasets. Essentially, this failure is caused by the following complications. Firstly, frequency of occurrence is a much less objective measure than most proponents of quantitative (cognitive) linguistics seem to realize. The assessment of frequency scores depends not only on what researchers retrieve and count as valid tokens, but also on how they calculate frequency. Even if they show awareness of the need to distinguish absolute from relative frequency (as of course most practitioners do), then it is still unclear how the two interact with each other, since absolute frequency may not be as irrelevant as most corpus linguists think. Secondly, advanced statistical techniques, which take absolute frequencies into consideration in order to gauge the significance of observed relative frequencies, have the problem of determining the reference scores required for the tests and run the risk of obscuring different combinations of absolute and relative frequency of occurrence. Thirdly, even if we accept the plausibility of the general claim that frequency of processing, and thus of occurrence in discourse, correlates with strength of entrenchment, we are still underinformed about the relation between cotext-free and cotextual entrenchment. This is particularly true of the large bulk of cases showing a medium range of association of lexemes and construction.” (Schmid 2010: 125f.)

¹⁸ It is defined as $P(\text{Outcome}|\text{Cue}) - P(\text{Outcome}|\sim\text{Cue})$. In this case it is the conditional probability of observing the verb given the ASP [a/(a+b)] minus the conditional probability of observing the verb given another ASP [c/(c+d)].

yields values between -1 and 1.¹⁹ In our case, an argument structure pattern (ASP) serves as the cue and the verb as the outcome. Four cases are distinguished with respect to the co-occurrence of argument structure pattern and verb:

Table 4: ΔP .

		OUTCOME	
		verb present	verb not present
CUE	ASP present	a	b
	ASP not present	c	d

In our basic verb profiles, the absolute corpus frequency of verbs and combinations of verbs and argument realization patterns is not captured. In order to apply ΔP to our data, we compute the absolute frequency of verbs and extrapolate the frequencies for verb-ARP combinations to a corpus of 1,000,000,000 running words.²⁰ The resulting numbers serve as input for the computation of ΔP according to Table 4. In the following, the ΔP values for four argument structure patterns that occur with psych-verbs are shown: (i) split stimuli with *mit* ‘with’ (Figure 11), (ii) sentential subjects headed by *dass* ‘that’ (Figure 12), (iii) direct speech introduced by psych-verbs (Figure 13), and (iv) animate stimuli as subjects (Figure 14).

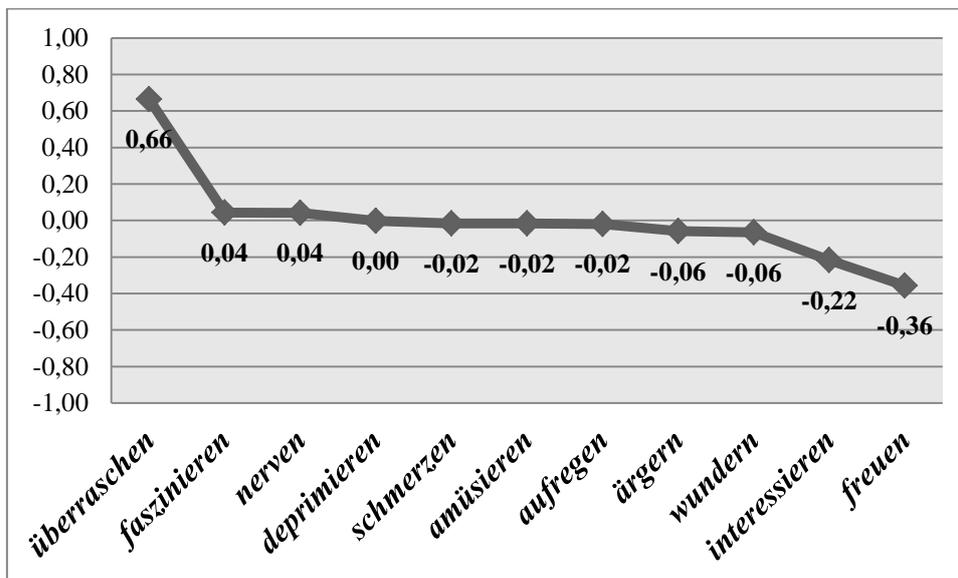


Figure 11: ΔP values for split stimuli with *mit* ‘with’.

¹⁹ Baayen (2011) compared five different measures for collexeme strength (among them Stefanowitsch & Gries’s standard collexeme measure and ΔP) and found positive correlations between all measures.

²⁰ Of course, to calculate the ΔP values, the extrapolation to 1 billion tokens is not necessary. It is done merely to make the values for different verbs with strongly differing corpus frequencies intuitively more comparable. Further analyses (not reported here) show that the results (especially the rankings presented in this section) do not change if we use the raw frequencies instead of the extrapolated values.

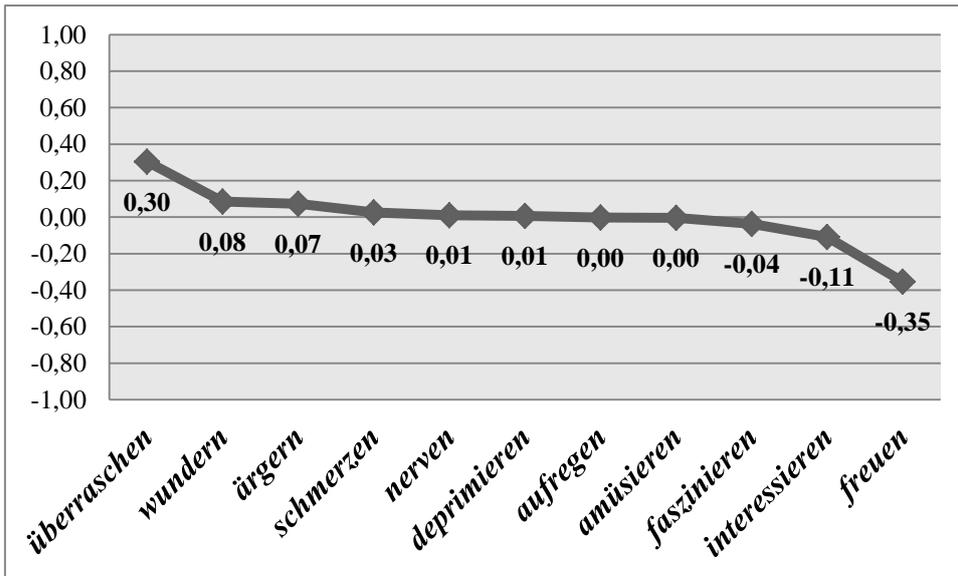


Figure 12: ΔP values for sentential subjects headed by *dass* 'that'.

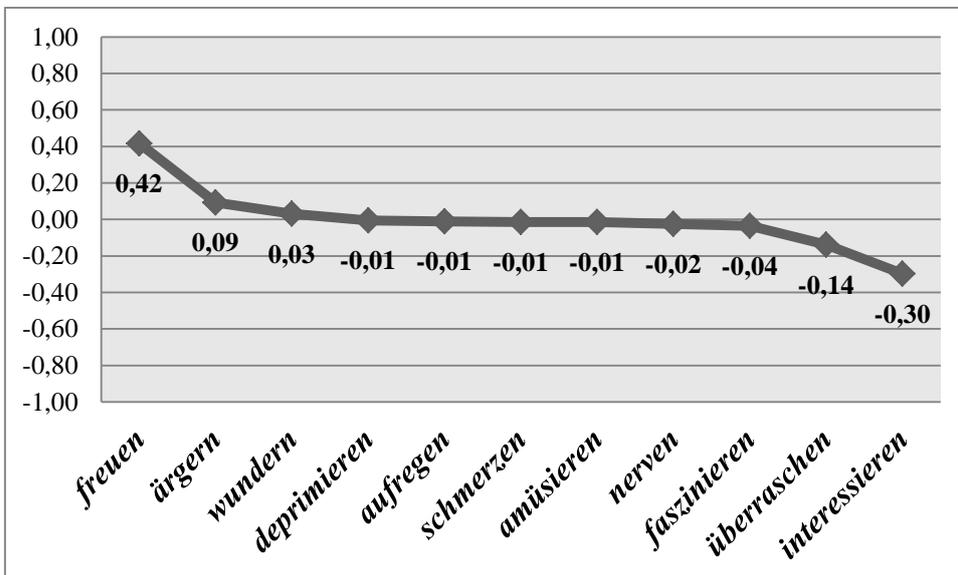


Figure 13: ΔP values for direct speech introduced by psych-verbs.

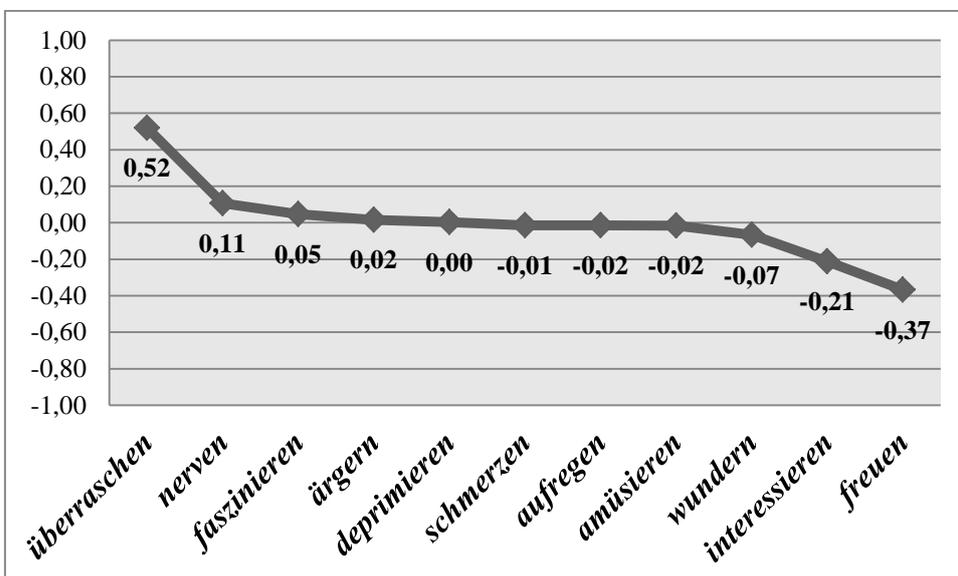


Figure 14: ΔP values for animate stimuli as subjects.

A strong caveat is in order here. Since our investigation on psych-verbs is limited to eleven verbs and their verb profiles, the following results reflect a small part of linguistic reality in which only eleven verbs and their range of argument realization patterns exist. The extension of this study to more verbs and their profiles will probably modify the results found so far. Yet, we do expect that it will always be the case that argument structure patterns not only show stronger associations with some verbs than with others but also that we will not get a linear distribution of ΔP values. There will always be very many verbs that show an unobtrusive connection to a particular argument structure pattern and very few verbs that have significantly higher ΔP values than all others. The ΔP values for a particular pattern not only show that few verbs are entrenched very strongly with a pattern but also that there are some verbs that quite strongly repel the argument structure pattern.

Stefanowitsch and Gries (2003: 228) analyse the ditransitive construction with its basic transfer meaning and find that *give* shows the strongest association to this construction. Thus, the verb most strongly associated with the construction is the verb that also shares its basic meaning with the meaning of the construction. Of all four phenomena, the one represented in Figure 13 probably lends most to a Gries/Stefanowitsch type of explanation. If the function of introducing direct speech by a psych-verb is to convey the emotion somebody displayed while making an utterance, then emotion verbs like *freuen* ‘be happy / please’ or *ärgern* ‘be angry / anger’ should indeed show a stronger affinity to the construction than cognition verbs like *interessieren* ‘be interested / interest’. A different explanation might hold for Figure 11: the strong attachment of *überraschen* ‘surprise’ to the split stimulus pattern with *mit* ‘with’ is partly due to the fact that *überraschen* has a not infrequent usage variant similar to verbs of transfer, meaning ‘give something (as a present) to somebody and thereby surprise him’.

Semantic explanations for the other two phenomena are not as easy to come by. Although emotions can be triggered by all kinds of states of affairs, they are most often triggered by what other people do (cf. Ben Ze’ev 2001). One should, therefore, expect that in particular psych-verbs expressing strong emotions should be attracted to an animate stimulus pattern. This is not the case, though (cf. Figure 14). The two verbs expressing prototypical emotions, namely, *ärgern* ‘be angry / anger’ and *freuen* ‘be happy / please’ do not show any particular association to the pattern with animate stimulus subjects. Similarly, it is hard to imagine what semantic property might render *überraschen* ‘surprise’ in Figure 12 particularly suited to realize its stimulus as sentential subject.²¹

This brief discussion suggests that the general quantitative distribution patterns as represented by the four curves in Figure 11 through 14 are probably more similar than the reasons why particular verbs show an above-average association to each of the argument structure patterns.

4. Conclusion

The article tried to shed some light on the linguistic and non-linguistic factors that determine the quantitative distribution of psych-verbs with respect to argument structure patterns. On the basis of verb profile analyses, it explored the following four factors:

- (i) Functional dependencies between different argument structure patterns cause a certain trade-off between these patterns; the promotion of experiencers by the standard psych-

²¹ One of the reviewers suggested that information structure might play a role in the strong association of sentential stimuli with *überraschen*. Verbs expressing surprise might attract stimuli with a topic-comment structure such as sentential subjects more often than other psych-verbs, since surprise often involves learning something new about something that is already known. This is different from emotions such as anger or happiness which often have more immediate experiences as stimuli.

verb alternation and by passive formation correlated negatively, and the frequency of sentential complements depended on the general distribution of stimulus-as-subject and stimulus-as-experiencer patterns.

- (ii) Different genres show particular preferences for particular kinds of argument structure patterns; these preferences interact with the strength of genre-independent verb-specific affinities to these patterns.
- (iii) Cross-linguistically, it can be observed that verbs and their counterparts in another language often exhibit similar associations to particular patterns: the German-Romanian study showed that this holds even then when the general language-specific preferences for these patterns differ; thus, cross-linguistic conceptual-semantic properties seem to influence the distribution of argument structure patterns.
- (iv) Finally, general principles of the memory system influence the distribution; the cognitive entrenchment of each argument structure pattern seems to be facilitated by its strong association with very few verbs. It is not yet completely clear in how far semantic properties of verbs guide this entrenchment with particular argument structure patterns.

The empirical basis for investigations of this sort is still very thin. Only a couple dozen of verbs have been subjected to verb profile analyses so far. We are currently extending the verb profile analyses to more verbs. Furthermore, since the studies so far have indicated that there are probably more factors that strongly influence the distribution of argument structure patterns, these will be included into our verb profiles in order to facilitate multi-factorial analyses. This concerns factors such as word order, information structure, mode, pronominalization, and others.

References

- Baayen, Harald R. 2011. Corpus linguistics and naive discriminative learning. *Brazilian Journal of Applied Linguistics* 11. 295-328.
- Barðdal, Jóhanna. 2008. *Productivity. Evidence from Case and Argument Structure in Icelandic*. Amsterdam, Philadelphia: Benjamins.
- Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann. 2009. Language Is a Complex Adaptive System: Position Paper. *Language Learning* 59 (Suppl. 1). 1-26.
- Behrens, Heike. 2011. Grammatik und Lexikon im Spracherwerb: Konstruktionsprozesse. In Stefan Engelberg, Anke Holler & Kristel Proost (eds.), *Sprachliches Wissen zwischen Lexikon und Grammatik*, 375-396. Berlin & New York: de Gruyter.
- Belletti, Adriana & Luigi Rizzi. 1988. Psych-Verbs and θ -Theory. *Natural Language and Linguistic Theory* 6. 291-352.
- Ben-Ze'ev, Aaron. 2001. *The Subtlety of Emotions*. Cambridge, MA & London: MIT Press.
- Blumenthal-Dramé, Alice, 2012. *Entrenchment in Usage-based Theories. What Corpus Data Do and Do not Reveal about the Mind*. Berlin, Boston: De Gruyter.
- Boas, Hans C. 2011. Coercion and leaking argument structures in Construction Grammar. *Linguistics* 49(6). 1271-1303.
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge et al.: Cambridge University Press.
- Bybee, Joan L. & Clay Beckner. 2010. Usage-based theory. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, 827-855. Oxford: Oxford University Press.
- Cosma, Ruxandra & Stefan Engelberg. 2013. Subjektsätze als alternative Valenzen im Deutschen und Rumänischen. In Ruxandra Cosma, Stefan Engelberg, Susan Schlotthauer, Speranța Stanescu & Gisela Zifonun (eds.), *Komplexe Prädikationen als Argumente. Kontrastive Untersuchungen zum Deutschen, Rumänischen und Englischen*. Berlin: Akademie-Verlag.
- Cristofaro, Sonia. 2003. *Subordination*. Oxford, New York: Oxford University Press.
- Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25. 108-127.
- Diessel, Holger. 2011. Review of "Language, usage and cognition. By Joan Bybee. Cambridge: Cambridge University Press, 2010. Pp. ix, 262". *Language* 87(4). 830-844.
- Dowty, David R. 1991. Thematic Proto-Roles and Argument Selection. *Language* 67. 547-619.

- Dryer, Matthew Sygne. 1980. The Positional Tendencies of Sentential Noun Phrases in Universal Grammar. *Canadian Journal of Linguistics* 25. 123-195.
- Duwaerts, Tim & Gereon Ullmann. 2013. Quantitative Untersuchung zur Valenz deutscher Substantive. In Reinhard Köhler & Gabriel Altmann (eds.), *Issues in Quantitative Linguistics 3, Dedicated to Karl Heinz Best on Occasion of his 70th Birthday*, 392-403. Lüdenscheid: RAM-Verlag.
- Ellis, Nich C. 2002. Frequency effects in language processing. A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition* 24. 123-183.
- Ellis, Nick C. 2006. Language acquisition as rational contingency learning. *Applied Linguistics* 27(1). 1-24.
- Engelberg, Stefan. 2000. *Verben, Ereignisse und das Lexikon*. Tübingen: Niemeyer.
- Engelberg, Stefan, Svenja König, Kristel Proost & Edeltraud Winkler. 2011. Argumentstrukturmuster als Konstruktionen? Identität - Verwandtschaft - Idiosynkrasien. In Stefan Engelberg, Anke Holler & Kristel Proost (eds.), *Sprachliches Wissen zwischen Lexikon und Grammatik*, 71-112. Berlin, New York: de Gruyter.
- Engelberg, Stefan, Alexander Koplenig, Kristel Proost & Edeltraud Winkler. 2012. Argument structure and text genre: cross-corpus evaluation of the distributional characteristics of argument structure realizations. *Lexicographica* 28. 13-48.
- Gahl, Susanne, Dan Jurawsky & Douglas Roland. 2004. Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, & Computers* 36(3). 432-443.
- Gries, Stefan Th. 2010. Behavioral Profiles. A fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon* 5(3). 323-346.
- Gries, Stefan Th. 2011. Corpus data in usage-based linguistics. What's the right degree of granularity for the analysis of argument structure constructions? In Mario Brdar, Stefan Th. Gries & Milena Žic Fuchs (eds.), *Cognitive Linguistics. Convergence and Expansion*, 237-256. Amsterdam, Philadelphia: Benjamins.
- Gries, Stefan Th. 2012a. Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics. Some necessary clarifications. *Studies in Language* 11(3). 477-510.
- Gries, Stefan Th. 2012b. Corpus linguistics, theoretical linguistics, and cognitive/psycholinguistics: Towards more and more fruitful exchanges. In Joybrato Mukherjee & Magnus Huber (eds.), *Corpus Linguistics and Variation in English*, 41-63. Theory and Description. Amsterdam, New York: Rodopi.
- Gries, Stefan Th. & Dagmar Divjak. 2009. Behavioral profiles: A corpus-based approach to cognitive semantic analysis. In Vyvyan Evans & Stéphanie Pourcel (eds.), *New Directions in Cognitive Linguistics*, 57-75. Amsterdam, Philadelphia: Benjamins.
- Gries, Stefan Th., Beate Hampe & Doris Schönefeld. 2010. Converging evidence II: more on the association of verbs and constructions. In John Newman & Sally Rice (eds.), *Experimental and empirical methods in the study of conceptual structure, discourse, and language*, 59-71. Stanford: CSLI.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collocation analysis. A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1). 97-129.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2010. Cluster Analysis and the Identification of Collexeme Classes. In John Newman & Sally Rice (eds.), *Empirical and Experimental Methods in Cognitive/Functional Research*, 59-71. Stanford: CSLI.
- Grimshaw, Jane. 1990. *Argument Structure*. Cambridge, MA & London: MIT Press.
- Köhler, Reinhard. 2005. Quantitative Untersuchungen zur Valenz deutscher Verben. *Glottometrics* 9. 13-20.
- MacDonald, Maryellen C. 1994. Probabilistic Constraints and Syntactic Ambiguity Resolution. *Language and Cognitive Processes* 9(2). 157-201.
- MacDonald, Maryellen C., Neal J. Pearlmutter & Mark S. Seidenberg. 1994. Syntactic ambiguity resolution as lexical ambiguity resolution. In Charles Clifton, Jr., Lynn Frazier & Keith Rayner (eds.), *Perspectives on Sentence Processing*, 123-153. Hillsdale, NJ: Lawrence Erlbaum.
- Noonan, Michael. 1985. Complementation. In Timothy Shopen (ed.), *Language Typology and Syntactic Description. Vol. II: Complex Constructions*, 42-140. Cambridge et al.: Cambridge University Press.
- O'Hara, Robert. B., & Kotze, D. Johan. 2010. Do not log-transform count data. *Methods in Ecology and Evolution*, 1(2), 118-112.
- Reichmann, Oskar. 2012. *Historische Lexikographie. Ideen, Verwirklichungen, Reflexionen an Beispiele des Deutschen, Niederländischen und Englischen*. Berlin, Boston: De Gruyter.
- Roland, Douglas William. 2001. *Verb Sense and Verb Subcategorization Probabilities*. Ph.D. University of Colorado.
- Roland, Douglas & Daniel Jurafsky. 1998. How verb subcategorization frequencies are affected by corpus choice. *COLING-ACL '98: 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Aug. 10-14, 1998, Montreal, Quebec, Canada*, 1122-1128. Montreal.
- Roland, Douglas, Daniel Jurafsky, Lise Menn, Susanne Gahl, Elizabeth Elder & Chris Riddoch. 2000. Verb Subcategorization Frequency Differences between Business-News and Balanced Corpora: The Role of Verb Sense. *Comparing Corpora. A workshop held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics, 7th October 2000, Hong Kong*, 28-34. Hongkong.

- Salkoff, Morris. 2002. Verbs with a Sentential Subject. A Lexical Examination of a Sub-Set of Psych Verbs. *Linguisticæ Investigationes* 25(1). 97-147.
- Schmid, Hans-Jörg. 2010. Does frequency in text instantiate entrenchment in the cognitive system? In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, 101-133. Berlin & New York: de Gruyter.
- Schulte im Walde, Sabine. 2003. *Experiments on the Automatic Induction of German Semantic Verb Classes*. PhD. Stuttgart: Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung.
- Schulte im Walde, Sabine. 2009. The induction of verb frames and verb classes from corpora. In Anke Lüdeling, & Merja Kytö (eds.), *Corpus Linguistics. An International Handbook*, Teilband 2, 952-971. Berlin & New York: Mouton de Gruyter.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209-243.
- Steiner, Petra. 2011. Diversification of English Valency Patterns. In Emmerich Kelih, Victor Levickij & Yuliya Matskulyak (eds.), *Issues in Quantitative Linguistics 2, Dedicated to Reinhard Köhler on Occasion of his 60th Birthday*, 148-169. Lüdenscheid: RAM-Verlag.
- Tomasello, Michael. 2003. *Constructing a Language. A Usage-Based Theory of Language Acquisition*. Cambridge, MA & London: Harvard University Press.