DE GRUYTER MOUTON GCLA 2017; 5: 143–176

Volodymyr Dekalo and Beate Hampe

Networks of meanings: Complementing collostructional analysis by cluster and network analyses

https://doi.org/10.1515/gcla-2017-0011

Abstract: This paper contributes to the development of collostructional analysis by taking up one of the long-standing issues in collostructional analysis, viz. the assessment and interpretation of the collexeme lists created. It uses the DWDS corpus documenting 20th century written German to study two modal-verb constructions which contain either *vermögen* or *bekommen* as the first verb plus an infinitive. When expressing 'possibility'/'capability', these constructions compete with the ubiquitous and strongly grammaticalized *können*-construction. To compare their constructional semantics, the paper combines collostructional analyses with a close investigation of the verbal meanings characterizing their V_{inf} slots. To classify the collexemes occurring in these slots comprehensively, the semantic information provided about each V_{inf} collexeme by "GermaNet" is utilized in three different ways: A manual, context-sensitive classification of the top collexemes into the 15 generic verb classes provided by GermaNet is followed by cluster analyses over the semantic-relatedness values given out by the SemRel tool of GermaNet for the pairwise comparison of all possible collexeme combinations. This is complemented by a network analysis identifying collexemes that are salient or "central", i.e. located on the largest number of "shortest paths" connecting any two collexemes. The implications of the results for the functional differential between the two constructions as well as the methodology itself are discussed.

Keywords: Usage-based construction grammar, collostruction analysis, cluster analysis, network analysis, German modal-verb constructions, grammaticalization

Volodymyr Dekalo: University of Hannover.
E-mail: volodymyr.dekalo@germanistik.uni-hannover.de
Beate Hampe: Erfurt University. E-mail: beate.hampe@uni-erfurt.de

1 Introduction

In usage-based construction grammar, syntactic constructions are viewed as complex as well as highly (though not necessarily fully) schematic symbolic units distilled by language users from their experience of real usage events (e.g. Bybee 2013; Bybee & Beckner 2010). In line with the assumption of a lexicon-syntax cline, syntactic constructions have been hypothesized to resemble other, less complex and non-schematic symbolic units (like lexemes) in that they, too, may exhibit polysemy (e.g. Goldberg 1992) and also enter into relations of (near) synonymy, i.e. compete with other constructions in the realisation of a given function. Regarding the latter, the principle of "No Synonymy" (e.g. Goldberg 1995: 67) exerted a strong guiding influence on recent cognitive-functional research aiming at explicating functional differences between competing constructions, especially those partaking in argument-structure alternations (e.g. Goldberg 2002).

Regarding the corpus-based investigation of the relations, or associations, between words and syntactic constructions, the family of quantitative methods referred to as "Collostructional Analysis" (for a survey, see Stefanowitsch 2013) has for some time established itself as a kind of methodological standard. In line with the tenets of usage-based approaches to syntactic structure, the underlying assumption of all collostructional analyses is that knowledge about the lexemes most closely associated with the open slots of a syntactic construction gives access to (aspects of) the constructional semantics. Of the three collostructional methods, two are of relevance to this study, called "simple" and "distinctive" collexeme analysis (Stefanowitsch & Gries 2003; Gries & Stefanowitsch 2004a,b). They provide the means for the synchronic analysis of (i) the strength of the associations between a specific syntactic construction and the lexical items occurring in one of its slots (ex 1.a) as well as (ii) the relations between the lexical realisations of (functionally parallel) open slots in two or more functionally similar/competing constructions, allowing for the assessment of the typicality of these realisations for any of the constructions under comparison (ex 2). To identify constructions or to study the semantics of constructions as parts of larger constructional networks, collostructional methods have also been applied in conjunction (e.g. Hampe 2011, 2014; Schönefeld 2015; Wulff, 2006).

(1) ...that a colleague... should <u>think nothing of *abandoning*</u> his wife... (the *think-nothing-of* VERB-*ing* construction, cf. Stefanowitsch & Gries 2003: 222)

(2) John <u>will</u> *send* Mary a book. vs. John <u>is going to</u> *send* Mary a book. (*will*-future vs. *be-going-to* future, Gries & Stefanowitsch 2004a: 113)

Collexeme analyses have frequently been applied to determine the semantic potential of (predominantly English) syntactic constructions, especially argument-structure constructions, determining the patterns of clausal syntax (ex 1), but also used to study tense, aspect and modality constructions, such as the future constructions illustrated in (2) (see Stefanowitsch 2013: 302). In all of these studies, the syntactic slot investigated for lexical variety has been that of the lexical verb.[1]

Applying collostructional analysis to the study of constructions that express syntactic categories of the verb presupposes that information about the lexical realisations of their main-verb slot is relevant to the determination of the functional range of the tense-, voice-, or modality constructions they occur with. Given that a true serial-verb construction (V_{lex} + $V_{lex\ infinitive}$) whose first element has not undergone any semantic bleaching poses greater syntagmatic restrictions on the second verb slot than a strongly grammaticalized auxiliary or modal would do, the determination of the most productive *verb classes* is of particular importance. The lexical variety found in the second verb slot (as reflected both by the type frequency of the collexemes it attracts and the number and types of the verb classes involved) indicates the productivity of the construction, hence also the level of grammaticalization achieved (for a survey, see Heine & Narrog 2010). Here, this is applied to the study of the V + V_{inf} constructions (potentially) competing with the strongly grammaticalized and overwhelmingly frequent modal construction with *können*.

Though the pros and cons of a number of association measures employed in association statistics have recently been the topic of considerable debate (Ellis & Ferreiro-Junior 2009; Gries 2015; Schmid & Küchenhoff 2013, 2015), the present paper is not concerned with the choice of the most adequate association measure, but with another long-standing issue in collostructional analysis, viz. the assessment and interpretation of its results, the collexeme lists created. While early collostructional analyses have used traditional linguistic analysis to group the top collexemes into semantic classes, sometimes making reference to pre-existing work on word classes (e.g. Wulff 2006 classifies verbs on the basis of Levin 1993), some scholars have employed further quantitative methods to explore the semantic regularities exhibited by collexeme lists: Most notably, these were either cluster analyses (e.g. Gries & Stefanowitsch 2010) or network analyses (e.g. Ellis et al. 2014, for a survey, see Gries & Ellis, 2015).

1 See Hilpert (2008) for an extensive collostructional analysis of constructions expressing future tense in a range of Germanic languages.

This paper reports selected results from a much larger collostructional analysis of four constructions in written German which are potential competitors of the extremely frequent and highly grammaticalized *können*-construction (ex 3.a) in that they (can) also express 'possibility/capability' (ex 3b–e):[2]

(3) a. In unserer Welt *kann* ein Mensch oder ein Buch nicht mehr eine so starke Erschütterung <u>hervorrufen</u> wie früher. (DWDS: Die Zeit, 16.12.1999)
b. Das Volk aber will nicht die allgemeine Wehrpflicht, folglich schaffen sie ein Heer, das für eine tätige und aktive Politik genügende Dienste zu <u>leisten</u> *vermag* [...]. (DWDS: Berliner Tageblatt, 05.03.1907)
c. Werden wir *verstehen*, diese Welt zu einer sicheren Welt zu <u>machen</u>? (DWDS: Archiv der Gegenwart, 1985)
d. Leistungssport am Wochenende kann den Verfall nicht aufhalten, die Kollegen *wissen* nichts Neues zu <u>erzählen</u>. (DWDS: Die Zeit, 1999)
e. Ansonsten *bekamen* zunächst nur die Ring-Angestellten die Folgen des Missmanagements zu <u>spüren</u>. (DWDS: Der Spiegel, 06.10.1980)

The constructions with *können*, *vermögen*, *verstehen*, *wissen* and *bekommen*, all associated with the schema [V_{finite} $V_{infinitive}$], do not express the modality of 'possibility'/'capability' to the same degree. Rather there seems to be a cline of meanings from the most strongly grammaticalized uses of *können* down to the construction with *bekommen*, which is polyfunctional in that it retains both its uses as a full lexical verb in serial-verb combinations with other lexical verbs (e.g. *etwas zu essen bekommen*, 'receive sth. to eat') and has developed more grammaticalized uses like those in (3.e). The fact that the former presents the source of the grammaticalization process motivating the latter makes the two uses hard to distinguish, i.e. simultaneously relevant, in many cases.

Apart from reporting the results of the collostruction analysis, section 3 also shows how the combination of cluster and network methods can assist in the interpretation of the collexeme lists, and thus also in the determination of the functional differential between pairs or sets of competing constructions, by (i) improving the determination of the crucial verb classes involved, and by (ii) identifying specific exemplars that are likely to be of particular semantic salience in the categories presented by the modal constructions investigated. The potential of this methodology for cognitive, usage-based construction grammar

2 The case studies in this paper report data and analyses from the first author's Ph.D. project at the Linguistics Department of the University of Erfurt, supervised by the second author.

in general and for an understanding of the modal constructions studied here will be assessed in section 4.

2 Methods

The German corpus chosen for the investigation of the modal constructions with *vermögen, verstehen, wissen* and *bekommen* listed in (3.b–d) was the “DWDS-Kernkorpus des 20. Jahrhunderts” containing about 100 million words of written German and balancing texts from literary, academic/scientific, journalistic and non-literary, popular writing (for a survey, see Geyken 2007). A simple collexeme analysis was used to determine which words are attracted most strongly to the V_{inf} slot of the respective modal constructions.[3] As the collostruction analysis assumes these infinitives to provide access to vital aspects of the constructional semantics, cases of verbal ellipsis were manually recovered such that examples of finite verbs occurring with several coordinated verbal infinitives in the same clause (ex 4) were treated as more than one token of the construction in accordance with the number of coordinated infinitives.

(4) […], daß Portugals Kommunisten eine große Kraft verkörpern, die niemand zu <u>übersehen</u> oder zu <u>übergehen</u> *vermag*. (DWDS: Neues Deutschland, 15.09.1977)

In line with usage-based models – and assuming that a balanced corpus approximates a downsized sample of an (average) experience at least very roughly – a simple collexeme analysis assesses the typicality of a lexeme (as a *type*) for a given construction in terms of various relevant *token* frequencies in the corpus chosen. Table 1.a illustrates which kinds of frequencies are needed in the present study for each single verb occurring in the V_{inf} slot of the *vermögen*-construction (henceforth also ‘*vermögen*-cx’). Apart from the construction-frequency of *leisten* (i.e. *leisten* inside the construction, henceforth also ‘cx-frequency’), these are the corpus frequencies of the construction at issue (here all tokens of *vermögen* occurring with verbal infinitives: 9.379) as well as the corpus frequency of infinitival *leisten* (cells A and C = 5.628) – thus counting in occurrances of *leisten* in other constructions with verbal infinitives – and the frequency of all infinitival lexical verbs in the corpus as the overall corpus size (here 1.744.053).

3 We thank Stefan Th. Gries for kindly providing his R-script *Coll.analysis 3.5* (Gries 2014), which was used for all collostruction analyses, both simple and distinctive.

Table 1.a: Input to a simple collexeme analysis for one lexical verb in the VERMÖGEN construction.

	leisten	**¬ *leisten***	**total**
vermögen-cx	**A 127**	B (= 9.379 – A)	**9.379**
¬ vermögen-cx	C (= 5.682 – A)	D (= 1.744.053 – (A + B + C))	C + D
total	**5.628**	B + D	**1.744.053**

Table 1.b: Input to a distinctive collexeme analysis comparing verbs in the V_{inf} slot of the *vermögen*-cx and the *bekommen*-cx.

V_{inf}	**freq of V_{inf} in the *vermögen*-cx**	**freq of V_{inf} in the *bekommen*-cx**	**total**
leisten	127	0	127
spüren	5	172	177
beurteilen	41	0	41
fressen	1	10	11
[...]	[...]	[...]	[...]
total (= construction freq)	**9,379**	**1,259**	**10,638**

For each verb occurring in the V_{inf}-slots of one of the two constructions, the distribution illustrated in Table 1.a was statistically assessed by means of the Fisher-Yates Exact test.[4] As has become customary, the strength of the association between a lexical item and the construction it occurs in, its “collostruction strength” is given as the negative logarithm to the basis of 10 of the p-value yielded by this assessment. A higher rank of a collexeme in the collexeme list is taken to indicate a higher degree of typicality for the construction in terms of actual usage.[5]

4 We continue to use this measure despite debate in the collostruction literature about its use as an association measure because of experimental evidence supporting its value in predicting actual speaker behavior and because the resulting collexeme rankings correlate closely with rankings created by the application of ΔP construction->verb (for extensive discussion see, e.g., Ellis & Ferreiro-Junior 2009; Gries 2015; Gries et al. 2005, 2010; Schmid & Küchenhoff 2013, Küchenhoff & Schmid 2015).

5 *Collostruction strength*: –log (Fisher exact, 10): coll. str. > 3: p < .001; coll. str. > 2: p < .01; coll. str. > 1.301: p < .05

While two separate simple collexeme analyses were employed to determine the functional core of the constructions with *vermögen* und *bekommen*, a (pairwise) distinctive collexeme analysis was additionally carried out in order to determine those verb types that are most characteristic of one of the two constructions in the direct comparison, hence indicative of areas of functional divergence between the constructions compared.[6] If high-ranked simple collexemes are also distinctive for one of the constructions in this comparison, the functional cores of the constructions under comparison are assumed to differ.

A distinctive collexeme analysis requires fewer input frequencies than a simple one, in our case the respective corpus frequencies of the two modal constructions to be compared, as well as those of the verbs occurring in the V_{inf} slots of the two constructions (Table 1.b).

As implied by the preceding discussion, collexeme analyses are *not* sense-specific. While a sense-specific determination of the cx-frequencies of the verbs occurring in the respective V_{inf} slot would 'only' require a very high analysis- and coding effort, the retrieval of the overall corpus frequencies of these verbs additionally required for the simple collexeme analysis cannot be done in a sense-specific way, at least not given standard corpus annotations. A collexeme analysis thus always incurs a partial loss of relevant semantic information. This problem appears enlarged here because the V_{inf} slot of a highly grammaticalized modal, aspectual, or tense construction is semantically far less restricted than the predicator slot of an argument-structure construction (but see our discussion in section 3).

In order to determine the functional core of each of the constructions under investigation, the functional differences between them as well as the potential role of single exemplars in this, the collexeme lists were further analysed as follows: In a first (more traditional) step, each token of each significantly attracted collexeme was categorized as a member of one of 15 more generic verb classes. Although the verb rankings created by the collexeme analysis are not sensitive to verbal polysemy, this step brings verbal polysemy back in, at least in the minimal form of a manual check for the particular sense that every single token of a given top collexeme actually instantiates in its context of occurrence in the corpus. In order to avoid most of the reliability issues arising in the context of a manual coding for semantic class, the coding was done by the first author following the classification and illustrative examples provided by "GermaNet" (Hamp & Feldweg 1997), the German version of "WordNet" (Miller 1990), whose

6 In the original study, a *multiple* distinctive collexeme analysis is used to compare all four constructions simultaneously.

semantic hierarchy was originally inspired by the verb classes discussed in Levin (1993).[7] The core functionality of each of the constructions was thus assessed with considerable precision – and hence to some extent also the functional differences between the two constructions, viz. in terms of the number, kinds and dominance of the verb classes *actually* represented by all of the tokens instantiating the top collexemes in each of the rankings.[8] This close qualitative inspection was then complemented by a comparison of the simple collexeme rankings with the distinctive ones – primarily in terms of verb types again, i.e. disregarding issues of verbal polysemy for the reasons expounded above.

The second step implemented a methodology developed by Ellis and colleagues for English constructions, utilizing the functionality of WordNet (Ellis et al. 2014): To this end, all collexemes of each construction were paired with the help of the R-function `combinations ()`, exhausting all possible combinations for undirected pairs. In order to assess the semantic similarity of the verbs making up any given pair, the list with all verb pairs was fed into the semantic-relatedness tool provided by GermaNet,[9] yielding a similarity measure for each verb pair derived from information about the relative positions of the collexemes of our constructions within the overall taxonomy of GermaNet. In accordance with preceding analyses, the measure chosen was *PATH* (Pedersen 2004). It calculates the relatedness between two synset nodes (here the meanings of two collexemes) as a function of the distance between these two nodes and the longest possible of all "shortest" paths (between any two nodes) in the entire taxonomy.[10] In doing so, the semantic-relatedness tool takes into account verbal polysemy and gives out a large number of measurements, comparing all combinations of all of the verbal senses recorded in GermaNet for each of the paired collexemes to be compared. However, the collexeme pairs fed into the tool are not marked for verbal polysemy. Following Ellis and colleagues (2014), this problem was dealt with by including the highest of all of the similarity measures offered for each verb-pair in the further analysis.

7 We thank the University of Tübingen (especially Prof. Dr. Erhardt W. Hinrichs) for granting us an Academic Research License to use GermaNet.

8 We are aware that the (currently unfeasible) inclusion of sense-specific information in the collostruction analysis itself would have yielded different collexeme rankings.

9 GermaNet Semantic Relatedness API 9.1 (see: http://www.sfs.uni-tuebingen.de/ GermaNet/tools).

10 Specifically, the function creates a similarity value <1 by computing the difference in length between the longest of all "short paths" and the length of the path between the nodes to be compared and dividing that difference by the length of the longest "short path" (see the GermaNet manual for further details).

Going beyond Ellis and colleagues, but inspired by other studies (e.g. Gries & Stefanowitsch 2010), the aforementioned list of pairwise similarity measures for all of the collexemes of each of the constructions investigated was, in a third step, furthermore explored by a *hierarchical agglomerative cluster analysis* employing the Ward's-method as a clustering algorithm. The cluster analysis groups together the most similar collexemes in a bottom-up fashion, thus suggesting fine-grained classifications. As verbal polysemy is not represented in the input to the cluster analysis, each collexeme can appear in the analysis only once, i.e. belong to only one cluster, no matter how polysemous it actually is.

Fourthly, and again following Ellis et al. (2014), these similarity measures provided the input to a *network analysis*, visualizing how the semantic relations in GermaNet as a lexical/semantic hierarchy apply to the collexemes (as lexical types) of the two constructions investigated. This is done rather indirectly again, viz. via the similarity measures obtained in step 2: The network software includes any pairs that exhibit a similarity value above a certain threshold in the network to be built – i.e. links them by an *edge* in the resulting visualization.[11] In order to create any clearly visible (i.e. intelligible) network structure, we had to diverge from Ellis and colleagues and define a much higher threshold for including related elements in the net, viz. similarity values of 0.9 instead of 0.5, because the networks for the constructions under investigation here (especially the *vermögen*-cx) turned out to be much denser. This is perhaps not too surprising considering that the lexical-verb slots of modal-verb constructions are less restricted than the predicator slots of the argument structure constructions investigated by Ellis and colleagues.

As both the cluster and the network method take similarity values as their input, some consideration about what differentiates those methods is required. Crucially, both the network analysis and the cluster analysis allow the analyst to view and explore relevant substructures of the semantic taxonomy provided by GermaNet, i.e. visualize parts of the relations in the paradigmatic network "behind" the collexeme list, but they perspectivize "collexeme relatedness" differently.

The cluster analysis differs from the manual determination of verb class carried out as step 1 of our analysis primarily with respect to *scope*. While the close, context-sensitive inspection of each corpus token can only be applied to few top collexemes, the cluster analysis presents a large-scale attempt at a semantic classification, involving *all* significantly attracted collexemes of the constructions. The classifications yielded by these two steps should still converge considerably because the similarity measures reflecting relative distances

11 We used a different software than Ellis and colleagues, viz. Gephi 0.9.1 (Bastian, Heymann & Jacomy 2009) between lexical nodes should correlate highly with the manual grouping of collexemes into the 15 GermaNet classes. The manual classification is needed to interpret the classifications suggested by the cluster analysis because it creates reliable information about which of the collexemes are likely to occur as members of a particular verb class. This is important because polysemous verbs can occur in several verb classes.

While likewise grouping similar verbal meanings together in sub-networks, or so-called "communities",[12] the network analysis takes a small subset of the input data for the cluster analysis – viz. only collexemes from pairs with similarity values above 0.9 – and highlights those of the remaining verb types whose meanings are "central", i.e. linked to a large amount of other meanings in the hierarchy of GermaNet. In network science, the determination of the "centrality" of a node in a network can technically be done in a variety of ways. We followed preceding work by Ellis and colleagues (cf. e.g. 2016: 86), who view central nodes as "hubs through which most paths pass", and likewise chose the *betweenness-centrality* measure to determine centrality. This score "measures the extent to which a node lies on paths between other nodes" (Newman 2010: 185). In the networks created here, the relative centrality of a collexeme is visually reflected not only by its size (the bigger the node, the more central), but also by its position.

As Ellis and colleagues (2016: 82–88) furthermore suggest, it must be assumed relatively more generic collexemes are likely to be central in the manner described, as it is the hyperonyms that are connected to (and lie on many paths between) many other verbal meanings. We assess the role of these "central nodes" more cautiously than Ellis and colleagues (ibid.: 86), who regard them as "best exemplars" of (and thus also as prototypical for) a category, providing an "idealized central description", a summary of its "most representative attributes". However, we agree that (the meanings of) collexemes presenting central nodes may be especially *salient* in the overall paradigmatic network associated with the V_{inf} slot of the respective constructions.

3 Results and discussion

Table 2 surveys the overall number of tokens that the entire investigation yielded, here we report in more detail only the results for the constructions with *bekommen*

12 "A community within a ... network is often informally defined as a group of nodes with dense connections to the other nodes in the group and sparser connections to other nodes posited to belong to a different community." (Ellis et al. 2016: 87)

and *vermögen* (henceforth called *vermögen*-cx and *bekommen*-cx), i.e. the two constructions which have turned out to be most dramatically different.

Table 2: Observed frequencies of four modal constructions competing with the *können*-cx in the *DWDS-Kernkorpus*.

	***bekommen*-cx**	***verstehen*-cx**	***wissen*-cx**	***vermögen*-cx**
DWDS-Kernkorpus	1,259	2,645	4,804	9,379

The functional differences between the two constructions under consideration are already indicated by the results of the two simple collexeme analyses. While the *bekommen*-cx significantly attracts 24 verbal collexemes (see Table 3.a), the overall number of significantly attracted verbs in the construction with *vermögen*, viz. 451, exceeds this number by far (see Table 2.b for the top 24 collexemes).

To further elaborate on the huge difference in the type frequency of the respective V_{inf} slots, Table 4 provides the results of step 1 of the analysis, surveying the semantic classes actually instantiated by the collexemes of the two constructions in their contexts of use. For clarity, we use the WordNet class labels, but provide the German labels actually used in GermaNet in parenthesis. Underlining marks out collexemes that instantiate senses from more than one of the 15 GermaNet classes and thus occur in more than one verb class. The metaphorical polysemy links between the perception-, cognition- and communication senses of many lexemes are well-known. Though the cluster and network methods cannot adequately reflect this, it is clear that highly polysemous verbs of an intermediate semantic specificity tend to appear in the top ranks of the collexeme lists precisely because of their potential to express concepts from a number of different domains.

The fact that 12 of the 15 GermaNet classes are represented in the top ranks of the collexeme list of the *vermögen*-cx further confirms that the semantic range of the V_{inf} slot in the *vermögen*-cx is much wider than that of the *bekommen*-cx, with only little overlap between the two. Both collexeme lists exhibit classes of perception and cognition verbs, though with different members and at different points in the rankings. Cognition verbs are among the two strongest collexeme groups of the *vermögen*-cx; perception verbs are marginal in its top ranks but present the strongest collexeme group of the *bekommen*-cx. Considering its relatively high type frequency in the latter, perception verbs must be determined as crucial to the modal use of the *bekommen*-cx. It is probably no co-incidence either

Table 3.a: All V_{inf} collexemes of the constructions with *bekommen* (coll. str. > 3: $p < .001$; > 2: $p < .01$; > 1.301: $p < .05$).

No	collexeme	corpus freq of V_{inf}	cx-freq of V_{inf}	coll str
1	*hören*	6414	286	409.11
2	*sehen*	19003	369	400.80
3	*spüren*	1195	172	333.58
4	*essen*	2661	92	118.92
5	*fassen*	2902	45	42.97
6	*tun*	19869	87	38.83
7	*lesen*	4542	47	36.98
8	*fühlen*	1940	32	31.59
9	*trinken*	1733	20	17.06
10	*fressen*	397	10	12.16
11	*verspüren*	176	7	10.08
12	*kosten*	858	10	8.92
13	*kaufen*	2622	14	7.86
14	*schlucken*	222	4	4.63
15	*packen*	397	4	3.66
16	*merken*	1287	5	2.58
17	*verkosten*	5	1	2.44
18	*mieten*	168	2	2.17
19	*plätten*	11	1	2.10
20	*ausfressen*	12	1	2.06
21	*rauchen*	214	2	1.97
22	*riechen*	265	2	1.79
23	*lästern*	32	1	1.64
24	*schmecken*	379	2	1.51

that the few cognition verbs in the *bekommen*-cx are metaphorical perception or activity verbs (*hören, fassen*) rather than verbs like *erfassen* or *beurteilen*, which are primarily cognition verbs and typical of the corresponding collexeme class of the *vermögen*-cx.

The remaining collexemes of the *bekommen*-cx that form a notable verb class are all consumption verbs. Here, *bekommen* is clearly used as a lexical verb meaning ‘receiving’. The same goes for the combinations with the few remaining collexemes from the classes of possession verbs, location verbs, and change verbs. As indicated above, these uses of *bekommen* with an infinitive present a serial-verb construction with two lexical verbs, which implicates rather than expresses the modal meaning. The *bekommen*-cx in contrast exhibits a large variety of collexeme classes uniting verbal meanings from concrete and abstract (including social) domains.

Table 3.b: Collexemes 1–24 of the construction with *vermögen* (coll. str. > 3: p < .001; > 2: p < .01; > 1.301: p < .05).

No	collexeme	corpus freq of V_{inf}	cx-freq of V_{inf}	coll str
1	*leisten*	5628	127	39.01
2	*fassen*	2902	86	34.94
3	*durchsetzen*	2240	72	31.55
4	*bieten*	2488	74	30.29
5	*lösen*	3843	91	29.70
6	*erkennen*	8770	143	29.02
7	*hervorbringen*	310	31	28.36
8	*folgen*	3995	86	25.35
9	*erfassen*	1742	53	22.40
10	*ändern*	3077	68	20.84
11	*fesseln*	311	24	19.48
12	*abgewinnen*	224	20	17.61
13	*beurteilen*	1362	41	17.37
14	*standhalten*	151	17	16.80
15	*durchdringen*	513	26	16.54
16	*ertragen*	1245	37	15.59
17	*widerstehen*	462	24	15.56
18	*geben*	20140	201	15.24
19	*ausüben*	2269	49	14.89
20	*unterscheiden*	3752	64	14.35
21	*eindringen*	1018	32	14.26
22	*bewältigen*	749	27	13.54
23	*angeben*	1612	38	12.87
24	*behaupten*	2449	47	12.50

The distinctive collexemes rankings also show the functional differential between the two constructions to be considerable (see Table 5). Given the high type frequency of the *vermögen*-cx, it is especially noteworthy that the twenty distinctive collexemes of the much "smaller" *bekommen*-cx are *all* also to be found on its simple collexeme list, determining the functional core of the construction to be simultaneously distinctive for it in the direct comparison with the *vermögen*-cx. Vice versa, nine of the top 20 collexemes of the *vermögen*-cx (viz. *geben, erkennen, leisten, folgen, bieten, durchsetzen, finden, lösen, ändern, unterscheiden, erklären*) are also distinctive for it in that comparison. Quite obviously, it is primarily the social, competition and cognition verbs that characterize the core of the *vermögen*-cx as distinct from that of the *bekommen*-cx. The few perception verbs further down the collexeme list of the *vermögen*-cx notwithstanding (e.g., 93: *wahrnehmen*; 122: *erblicken*), the particular specialty of the *bekommen*-cx in this

Table 4: GermaNet verb classes instantiating den V_{inf} slot der *vermögen*-cx und *bekommen*-cx.

***vermögen*-cx: 1,478 of 9,379 tokens**	***bekommen*-cx: 1,216 of 1,259 tokens**
8 COGNITION (‘KOGNITION’): *<u>fassen</u>, <u>lösen</u>, <u>erkennen</u>, <u>folgen</u>, erfassen, abgewinnen, beurteilen, <u>unterscheiden</u>*	9 PERCEPTION (‘PERZEPTION’): *<u>hören</u>, sehen, spüren, fühlen, verspüren, kosten, merken, riechen, schmecken*
8 SOCIAL VERBS (‘GESELLSCHAFT’): *leisten, durchsetzen, bieten, <u>lösen</u>, <u>folgen</u>, ausüben, bewältigen, <u>behaupten</u>*	5 CONSUMPTION (‘VERBRAUCH’): *essen, trinken, fressen, schlucken, rauchen*
4 COMMUNICATION (‘KOMMUNIKATION’): *<u>fassen</u>, hervorbringen, angeben, <u>behaupten</u>*	3 COGNITION (‘KOGNITION’): *<u>hören</u>, <u>fassen</u>, lesen*
3 STATIVE VERBS (‘ALLGEMEIN’): *<u>folgen</u>, standhalten, widerstehen*	2 POSSESSION (‘BESITZ’): *mieten, kaufen*
3 CONTACT (‘KONTAKT’): *<u>fassen</u>, <u>lösen</u>, <u>fesseln</u>*	2 CONTACT (‘KONTAKT’) *<u>fassen</u>, packen*
2 LOCATION (‘LOKATION’): *<u>folgen</u>, durchdringen*	1 CHANGE (‘VERÄNDERUNG’) *tun*
2 CHANGE (‘VERÄNDERUNG’): *<u>lösen</u>, ändern*	
2 EMOTION (‘GEFÜHL’): *<u>fesseln</u>, ertragen*	
2 CREATION (‘SCHÖPFUNG’): *hervorbringen, <u>geben</u>*	
2 COMPETITION (‘KONKURRENZ’): *durchsetzen, eindringen*	
2 POSSESSION (‘BESITZ’): *<u>fassen</u>, <u>geben</u>*	
2 PERCEPTION (‘PERZEPTION’): *<u>erkennen</u>, <u>unterscheiden</u>*	

comparison are verbs of perception and consumption. The distinctiveness of the latter does not come unexpected, as the lexical meaning of *bekommen* (‘receive’) goes well with the notion of consumption and the serial-verb use depends on that. With respect to the grammaticalization of this meaning to a modal meaning (‘possibility’/‘ability’), the results of the first step of our analysis suggest that this process has so far really reached only one verb class, viz. verbs of perception, and also extends to include a few polysemous cognition verbs whose senses are derived metaphorically from the former or from other concrete domains.

Step 2 of our procedure resulted in the creation of 101,475 collexeme pairs for the construction with *vermögen* and 276 collexeme pairs for the construction with *bekommen*. Applying the measure PATH, each pair was given the highest similarity value of all values suggested by GermaNet for this pair.

Table 5: Results of the pairwise distinctive collexeme analysis (coll. str. > 3: p < .001; coll. str. > 2: p < .01; coll. str. > 1.301: p < .05).

No	V_{inf} dist. collexemes of the *vermögen*-cx	obs freq in the cx	coll str	No	V_{inf} dist. collexemes of the *bekommen*-cx	obs freq in the cx	coll str
1	*geben*	201	11.11	1	*sehen*	286	281.26
2	*erkennen*	143	7.88	2	*hören*	369	249.60
3	*werden*	141	7.77	3	*spüren*	172	155.23
4	*leisten*	127	6.99	4	*essen*	92	80.26
5	*machen*	113	6.22	5	*tun*	45	42.85
6	*bringen*	110	6.05	6	*lesen*	87	27.71
7	*halten*	98	5.39	7	*fühlen*	47	20.62
8	*sagen*	118	5.31	8	*trinken*	32	15.50
9	*folgen*	86	4.72	9	*kaufen*	20	11.02
10	*bieten*	74	4.06	10	*fassen*	10	10.99
11	*durchsetzen*	72	3.95	11	*kosten*	7	9.28
12	*finden*	72	3.95	12	*fressen*	10	8.29
13	*lösen*	91	3.93	13	*verspüren*	14	6.49
14	*ändern*	68	3.73	14	*merken*	4	4.64
15	*unterscheiden*	64	3.51	15	*schlucken*	4	3.05
16	*erklären*	62	3.40	16	*mieten*	5	1.85
17	*sein*	62	3.40	17	*rauchen*	1	1.85
18	*erreichen*	57	3.13	18	*riechen*	2	1.85
19	*aufnehmen*	56	3.07	19	*schmecken*	1	1.85
20	*denken*	54	2.96	20	*packen*	1	1.82

Complementing the results yielded by step 1, comprehensive classifications of the collexeme rankings were suggested by the two cluster analyses created in step 3. Apart from corroborating the preceding observations by creating a classification in line with them, the cluster analysis of the *bekommen*-cx (see Figure 1) primarily provided us with a small-scale test case to prepare for the analysis of the much bigger cluster of the *vermögen*-cx.

Most notably, the dendrogram divides the 24 collexemes of the *bekommen*-cx into two main clusters at a distance height of 0.51. The first of these two major clusters is very homogenous in that (apart from the change verb *plätten*) it contains only the consumption verbs *rauchen*, *fressen*, *essen*, *schlucken*, *trinken*, which are united in a subcluster defined at a distance height of about 0.1. The second and bigger cluster, defined at a distance height of about 0.4, is more heterogeneous, uniting a number of smaller subclusters. The largest of these, defined at a distance height of about 0.2 and perfectly homogeneous, is larger than the first major cluster and made up of the perception verbs *verspüren*, *fühlen*, *spüren*, *sehen*, *riechen*, *hören*, *merken*, *verkosten*, *kosten* and *schmecken*. The rest of the second major cluster is not only smaller but also heterogeneous, with the two communication verbs *lästern* and *lesen* set apart from a few collexemes from other classes, such as verbs of change and possession. In sum, the cluster analysis foregrounds verbs of consumption and perception, thus suggesting classification in accordance with the results of step one of our analysis.

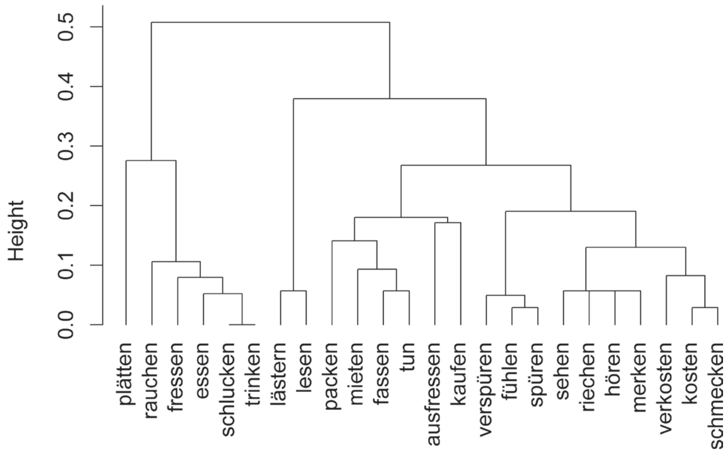

Figure 1: Dendrogram for the collexemes of the *bekommen*-cx.

Given that the *vermögen*-cx exhibits 451 significantly attracted collexemes, we can report only the major results of the cluster analysis here (see Appendix, Figures 4–13 for parts of the dendrogram). Aiming at the identification of the GermaNet verb classes uniting all or most of the members of a cluster (in one of their senses), a first manual inspection of the clusters of the dendrogram informed the decision to cut it at a distance height of 0.55.[13] This created a total of 20 relatively homogeneous major clusters. Table 6 lists these and the The GermaNet classes reflected by semantically homogeneous clusters or subclusters in the order of their appearance in the dendrogram.

Concerning overlap with the results yielded by step 1, it can firstly be stated that eight of the nine (largely) homogeneous major clusters unite verbs that can instantiate the same GermaNet classes as the tokens of the high-ranked collexemes do in the corpus. In the order of decreasing cluster size, these are: change verbs

13 The R-function `cutree (tree, h = 0.55)` was used to cut the dendrogram.

Table 6: GermaNet verb classes represented by (sub)clusters of the dendrogram for the *vermögen*-cx (Clusters marked by asterisks are provided in the Appendix, Figures 4–13).

No	No of members	Distance height uniting the cluster	Character of cluster	GermaNet classes represented
1	7	0.185	*homogeneous*	Location verbs
2	37	0.501	*homogenous*	Communication verbs
3	13	0.424	*homogeneous*	Location verbs
4	21	0.458	*homogeneous*	Location verbs
***..5**	23	0.449	*largely homogeneous*	Cognition verbs 1 communication verb
6	39	0.506	*heterogeneous*	Cognition verbs Perception verbs
7	15	0.375	*largely homogeneous*	Competition verbs, 2 social verbs
8	7	0.117	*homogeneous*	Social verbs
***9**	41	0.482	*heterogeneous*	Social verbs Competition verbs Change verbs Stative verbs Emotion verbs
***10**	6	0.281	*heterogeneous*	Cognition verbs Stative verbs
***11**	9	0.258	*homogeneous*	Verbs of bodily function*
12	7	0.326	*heterogeneous*	Location verbs Cognition verbs
***13**	14	0.509	*heterogeneous*	Cognition verbs Verbs of natural phenomena*
14	13	0.194	*homogeneous*	Emotion verbs
***15**	48	0.474	*highly heterogeneous*	Contact verbs Possession verbs Stative verbs Verbs of natural phenomena* Location verbs 5 verbs from 5 further classes
***16**	19	0.479	*heterogeneous*	Cognition verbs Social verbs

Table 6 (continued)

No	No of members	Distance height uniting the cluster	Character of cluster	GermaNet classes represented
***17**	22	0.338	*largely homogeneous*	Change verbs 6 verbs from 5 further classes
***..18**	61	0.446	*largely homogeneous*	Change verbs 2 verbs of bodily function
***19**	22	0.432	*homogeneous*	Creation verbs
***20**	27	0.474	*highly heterogeneous*	Emotion verbs Creation verbs Location verbs Change verbs Social verbs

(clusters 17, 18), communication verbs (cluster 2), location verbs (clusters 1, 3, 4), cognition verbs (cluster 5) and creation verbs (cluster 19), competition verbs (cluster 7) and emotion verbs (cluster 14), as well as social verbs (cluster 8). All of these classes are additionally represented by subclusters of the major heterogeneous clusters; this is the case with the components of clusters 6, 9, 10, 12, 15, 16 and 20. Secondly, the remaining verb classes found to be instantiated by tokens of top-ranked collexemes can be identified in the dendrogram, too, but likewise present subclusters of heterogeneous major clusters. These are: stative verbs (component of clusters 9, 10), contact verbs (component of cluster 15), possession verbs (component of cluster 15) and perception verbs (component of cluster 6).

The results of the cluster analysis furthermore suggest that the list of GermaNet classes derived in step 1 from the manual analysis of all tokens of the top 24 collexemes was not complete. Two more verb classes need to be added, which increases the number of verb classes from 12 to 14 (of a total of 15) GermaNet classes: One homogeneous major cluster (cluster 11) unites verbs of bodily function; and two subclusters (components of 13 and 15) are composed of verbs of natural phenomena.

The fact that so many of the 15 GermaNet classes are reflected by partially non-adjacent (sub)clusters scattered over the entire dendrogram suggests that, in line with the explorative character of the clustering method, its classificatory results require some further consideration. We suspect that the level of granularity or specificity at which classifications are created that are maximally informative, esp. with a view to constructional productivity, may well be higher than that defined by the 15 GermaNet classes (and thus may only become visible in the cluster analysis at levels defined by distance heights much lower than 0.55).

The issue is relevant as it is well-known in usage-based construction grammar that the productivity of a construction does not exclusively depend on type frequency and statistical pre-emption, but also on the degree of overall category coverage by exemplars and their similarity to central exemplars (or a prototype schematizing over these) – so that new exemplars can be added locally on the basis of strong analogies with members of a relatively dense cloud of existing exemplars (cf. e.g. Abbot-Smith & Tomasello 2006; Suttle & Goldberg 2011; Zeschel 2010). We thus checked whether (sub)clusters containing verbs belonging to one of 7 different GermaNet classes actually expressed meanings that are hyponymically related to those given by the GermaNet classes themselves and found this to be the case. We also found that the rank-highest collexeme in any given cluster instantiates a relatively more generic meaning and presents a good approximation to the meaning of the subcluster (see Table 7 and Appendix: Figures 4–13).

The bottom line is that the V_{inf} slot in the *vermögen*-cx is practically unrestricted with respect to the GermaNet verb classes that its collexemes can instantiate. Apart from confirming the centrality of the verb classes resulting from step 1 and completing the inventory of relevant verb classes at this level of specificity, the classificatory results achieved by the cluster analysis also suggest that, on its way to higher productivity, constructional slots open up local and highly specific pockets of productivity, which might be supported by slightly more generic collexemes of higher ranks whose meaning equals or approximates that of the entire group.

What the network analysis adds to this picture in step 4 is a visualization of the interconnectedness of the network of (paradigmatically related) meanings "behind" the collexeme list (Figures 2; 3, see also Appendix, Figure 14, for an enlarged image of the "community" encircled in Figure 3).

The communities making up the network for the *bekommen*-cx are largely in accordance with the results of the cluster analysis, again separating verbs of consumption and perception verbs from a more heterogeneous smaller verb group. The same cannot be said about the vast and dense network of the *vermögen*-cx. There, the cut-off similarity value of 0.9 has created communities that are fragments of semantic classes at best and thus of doubtful classificatory value.[14]

More relevant to the goals of step 4 were the collexemes determined as "central" by the network analysis.[15] In contrast to the cluster analysis, this

14 The community in Fig. 3, for instance, turns out to be largely composed of two verb sets from subclusters in the dendrogram: 4 verbs of natural/light phenomena: *erhellen, beleuchten, bestrahlen, scheinen*; as well as 3 change verbs related to natural development: *ausreifen, keimen, auswachsen* (see also Appendix, Fig. 14).

15 The much higher centrality values of the lexemes of the *bekommen*-cx are caused by the dramatically increased network size and density.

Table 7: Dendrogram of the *vermögen*-cx: Semantically coherent (sub)clusters of major clusters.

<table>
<tr><th>GermaNet class</th><th>(Part of) Cluster</th><th>Distance height</th><th>Sense uniting all collexemes</th><th>Collexemes forming the cluster (collexeme no ≤ 100 or no of rank-highest collexeme)</th></tr>
<tr><td>Stative verbs</td><td>no 9</td><td>0.05</td><td>protect</td><td>schirmen, bewahren (101), schützen</td></tr>
<tr><td></td><td>no 9 (adjacent)</td><td>0.20</td><td>support</td><td>fußen, stützen, hinweghelfen, anhaben, tragen</td></tr>
<tr><td></td><td>no 10</td><td>0.00</td><td>endure</td><td>überbrücken, überstehen (91)</td></tr>
<tr><td></td><td>no 10 (adjacent)</td><td>0.28</td><td>resist</td><td>standhalten (14)|widerstehen (17), aushalten, durchhalten</td></tr>
<tr><td>Verbs of natural phenomena</td><td>no 13</td><td>0.11</td><td>light up objects/ spaces</td><td>hineinleuchten, scheinen, bestrahlen, beleuchten, erhellen (104)</td></tr>
<tr><td></td><td>no 15</td><td>0.19</td><td>emit or reflect light</td><td>aufglänzen, spiegeln (73), aufhellen (87), blitzen</td></tr>
<tr><td>Cognition verbs</td><td>no 6</td><td>0.15</td><td>understand</td><td>nachvollziehen, begreifen (82), fassen (2), versetzen|hineindenken|hineinversetzen, einfühlen, erfassen (9), erschauen</td></tr>
<tr><td></td><td>no 13</td><td>0.33</td><td>solve problems/ arguments</td><td>nachweisen, entkräften, widerlegen, entwirren, lösen (5), entschlüsseln, entziffern, enträtseln (59), klären</td></tr>
<tr><td></td><td>no 16</td><td>0.12</td><td>conceptualize counterfactual situations</td><td>erlügen, prägen, vorausdenken, imaginieren|vorstellen (72), konzipieren, schmieden, ausdenken (100), erdenken</td></tr>
<tr><td>Contact verbs</td><td>no 15</td><td>0.07</td><td>separate (parts of) objects</td><td>abtrennen, abzupfen, abstreifen, abstoßen, loswinden, loslösen (134)|losmachen</td></tr>
<tr><td></td><td>no 15 (non-adj.)</td><td>0.08</td><td>use hands to get into contact</td><td>hineintasten, hintasten, kitzeln, festhalten, halten (84)</td></tr>
</table>

Table 7 (continued)

GermaNet class	(Part of) Cluster	Distance height	Sense uniting all collexemes	Collexemes forming the cluster (collexeme no ≤ 100 or no of rank-highest collexeme)
Verbs of creation	no 19	0.04	copy from an original	*nachbilden* (126), *nachzeichnen, nachmalen, zeichnen*
	no 19 (adjacent)	0.05	produce	*zusammenzimmern, verweben, ritzen, synthetisieren* (81), *induzieren, hervorzaubern* (74), *erzeugen* (41)|*herstellen*
Verbs of change	no 17	0.18	develop	*entwickeln, ausbilden, bilden, auskeimen* (161), *reifen, auswachsen, gestalten*
	no 18	0.10	destroy	*brechen, durchbrechen* (60), *zertrümmern, durchnagen, auslöschen, zerstören*
	no 20	0.06	cause or activate	*erwecken, wecken* (35), *einhauchen, anrichten, auslösen* (29)|*hervorrufen* (51)
Emotion verbs	no 14		cause or express positive emotion	*erwärmen, begeistern|fortreißen, aufstören, hinwegtrösten, trösten, empfinden* (76), *sänftigen, erspüren, locken, weglocken, nachempfinden|nachfühlen*
	no 20	0.11	exert emotional influence	*fesseln* (11), *bestricken, betören, bannen* (37), *umwerfen*
	no 20 (non-adj.)	0.16	instigate emotional response	*stacheln, aufbringen* (26), *zürnen, aufheitern, erschüttern* (83)

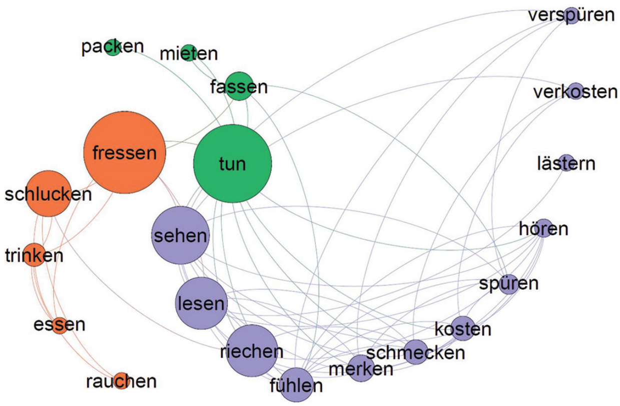

Figure 2: Collexeme network of the *bekommen*-cx.

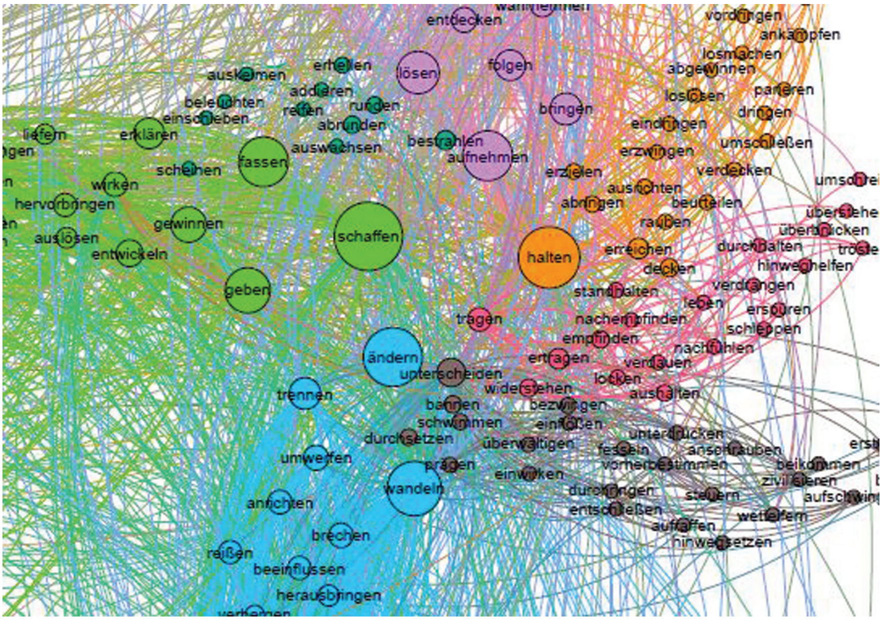

Figure 3: Central part of the collexeme network of the *vermögen*-cx, one “community” highlighted.

perspective favours tightly interlinked, i.e. relatively generic meanings (see Table 8). The 8 most “central” nodes of the network for the *bekommen*-cx are given by *fressen* and *schlucken* from the consumption-verb community; *sehen, lesen, riechen* und *fühlen* from the perception-verb community as well as *tun* and *fassen* from the remaining part of the network. The network analysis thus highlights the class of perception verbs. 5 of those 8 collexemes are also in the top 10 collexeme ranks (2: *sehen*, 5: *fassen*, 6: *tun*, 7: *lesen*, 8: *fühlen*).

The 15 most "central" nodes in the network of the *vermögen*-cx are all not only very generic, but also highly polysemous (see Table 9). Some of these (esp. *halten, fassen, geben, bringen*) can even be characterized as *light verbs* – in the sense that their meaning depends to a large extent on the argument structures they occur in. Since it has little diagnostic value that the 15 most central collexemes belong to 5 of 7 different, extremely heterogeneous communities, we consulted the classification resulting from the cluster analysis carried out in step 3 to check the nature of the clusters these verbs belong to. We found that the 15 most central nodes of the network reflect 8 of the 14 GermaNet classes instantiated by the collexemes of the construction. Apart from that, the network analysis also emphasizes highly generic cognition verbs, which belong to 3 different "communities" of the network (and also appeared in 4 different major clusters in the cluster analysis. The network analysis thus confirms the semantic versatility of the V_{inf} slot previously observed.

However, only 6 of the verbs with the highest 20 betweenness-centrality values are also among the top 20 collexemes, and still only half of them are found in the top 100 collexeme ranks. The precise relation between the "centrality" ranking created on the basis of the betweenness-centrality values and the collexeme ranking created on the basis of token frequencies in usage is thus far from clear.

4 Concluding remarks and issues for further research

Applying three methods in conjunction to explore the results of the collostruction analysis in depth allowed us to analyse the collexeme lists in a much more detailed way than was possible before. While the cluster analysis created a comprehensive and fine-grained semantic classification, the network analysis identified semantically salient collexemes, coding for meanings that are "central" in the hierarchy of paradigmatically related meanings 'behind' the collexeme list.

Concerning the modal constructions investigated, the main result is that only the *vermögen*-cx appears as a fully grammaticalized and highly productive modal construction expressing 'possibility/capability'. As a modal construction, the *bekommen*-cx in contrast was shown to be of a limited productivity only. Not only does its V_{inf} slot exhibit a very low type frequency, it also underlies strong

Table 8: Top betweenness-centrality values for collexemes of the *vermögen*-cx and the *bekommen*-cx.

***vermögen*-cx**			***bekommen*-cx**		
collexeme	**betweenness-centrality value**	**collexeme rank**	**collexeme**	**betweenness-centrality value**	**collexeme rank**
schaffen	5246.93	353	*fressen*	41.24	10
halten	4581.54	84	*tun*	38.74	6
ändern	4341.19	10	*sehen*	25.98	2
wandeln	3901.94	189	*lesen*	22.40	7
fassen	3463.82	2	*riechen*	22.26	22
aufnehmen	3461.97	49	*schlucken*	18.60	14
geben	3023.77	18	*fühlen*	11.14	8
lösen	2745.89	5	*fassen*	7.54	5
gewinnen	2075.89	124	*merken*	6.47	16
bringen	1739.19	187	*schmecken*	5.13	24
trennen	1717.64	156	*kosten*	5.06	12
erklären	1637.78	102	*trinken*	4.22	9
folgen	1609.54	8	*spüren*	2.09	3
unterscheiden	1409.80	20	*hören*	1.13	1
entwickeln	1253.87	297	*essen*	0.00	4
umwerfen	1088.91	449	*lästern*	0.00	23
entdecken	1067.76	64	*mieten*	0.00	18
wahrnehmen	1051.80	93	*packen*	0.00	15
brechen	1031.87	139	*rauchen*	0.00	21
wirken	1017.84	157	*verkosten*	0.00	17

semantic restrictions, essentially narrowing down V_{inf} verbs to verbs of perception. At the same time, high-ranked collexemes from the class of consumption verbs identified the serial-verb use as central in combinations of *bekommen* with infinitives. The level of grammaticalization exhibited by the *bekommen*-cx was thus determined as very low.

Concerning the methodology employed, the first issue relates to consequences of the loss of polysemy-related information for the cluster and network analyses, which were applied to the collexeme lists to answer strictly meaning-related questions. Obviously, the bottom-up classification performed by the cluster analysis is more precise the less polysemous and the more specific the collexemes are, because this reduces information loss. The cluster analysis thus brings to attention small, semantically coherent classes which may play an important role in driving the productivity of a construction (but may be entirely overlooked in the

Table 9: Collexemes functioning as central nodes in the network analysis of the *vermögen*-cx.

GermaNet verb class	Central nodes (betweenness-centrality value)	No of senses in GermaNet	Cluster no in Cluster Analysis
Social verbs	*schaffen* (5246.93)	7	9
Contact verbs	*halten* (4581.54)	26	15
Change verbs	*ändern* (4341.19)	3	17
	wandeln (3901.94)	4	18
Cognition verbs	*fassen* (3463.82)	11	6
	geben (3023.77)	10	5
	lösen (2745.89)	9	13
	trennen (1717.64)	9	5
	folgen (1609.54)	6	15
	unterscheiden (1409.80)	5	5
Perception verbs	*aufnehmen* (3461.97)	18	6
Possession verbs	*gewinnen* (2075.89)	18	6
Creation verbs	*bringen* (1739.19)	11	19
	entwickeln (1253.87)	10	17
Communication verbs	*erklären* (1637.78)	4	2

manual analysis of collexeme rankings, as their members will not occur in the top collexeme ranks).

The network analysis, however, focusses on collexemes with tightly interconnected meanings. Given that these tend to be generic and polysemous, multiple occurrences in the semantic hierarchy (which should further increase an item's centrality in the network) are reduced to one. We therefore concede that network analyses are more accurate the more semantically restricted a constructional slot is. This was clearly not the case with the *vermögen*-cx, given the highly polysemous nature of the collexemes appearing as central nodes in its network and the large number of verb classes actually and potentially represented by them.

The second crucial question revolves around whether and to what degree the two notions of "centrality" as determined by the collexeme ranking and by the network analysis do correspond. The question is of some importance, as both are known to favour more generic meanings and both have been discussed in the literature as pointing to prototypical (hence particularly salient, or even acquisitionally “path-breaking”) exemplars presenting the category core of the constructional slot under investigation. At the same time, the overlap between collexemes in the top positions of the respective rankings was found to be partial. Further research must determine whether these results are due to the extent of imprecision incurred by the loss of polysemy information involved in the quantitative methods or due to the fact that a central network position does not correspond *directly* to an association measure based on token frequencies *in principle*.

References

Abboth-Smith, Kirsten, and Michael Tomasello. 2006. Exemplar learning and schematization in a usage-based account of acquisition. *The Linguistic Review* 23. 275–290.

Bastian, Mathieu, Sebastien Heymann & Mathieu Jacomy. 2009. Gephi: An open source software for exploring and manipulating networks. *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media,* 361–362.

Bybee, Joan L. 2013, Usage-based theory and exemplar representations of constructions. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 49–69. Oxford: Oxford University Press.

Bybee, Joan L. & Clay Beckner. 2010. Usage-based theory. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis,* 827–856. Oxford: Oxford University Press.

Ellis, Nick C., Matthew Brook O’Donnell & Ute Römer. 2014. The processing of verb-argument constructions is sensitive to form, function, frequency, contingency, and prototypicality. *Cognitive Linguistics* 25(1). 55–98.

Ellis, Nick C. & Fernando Ferreiro-Junior. 2009. Constructions and their acquisition: islands and the distinctiveness of their occupancy. *Annual review of Cognitive Linguistics* 7. 187–220.

Ellis, Nick C., Ute Römer & Matthew Brook O’Donnell. 2016. *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*. West Sussex: Wiley.

Goldberg, Adele E. 1992. Inherent semantics of argument structure: The case of the English ditransitive construction. *Cognitive Linguistics* 3(1). 37–74.

Goldberg, Adele E. 1995. *Constructions. A Construction Grammar Approach to Argument-Structure*. Chicago, London: The University of Chicago Press.

Goldberg, Adele E. 2002. Surface generalizations: An alternative to alternations. *Cognitive Linguistics* 13(4). 327–356.

Geyken, Alexander 2007. The DWDS corpus: A reference corpus for the German language of the 20th century. In Christiane Fellbaum (ed.), *Idioms and collocations. Research in corpus and discourse*, 23–41. London, New York: Continuum.

Gries, Stefan Th. 2014. *Coll.analysis 3.5. A script for R to compute perform collostructional analyses*.

Gries, Stefan Th. 2015. More (old and new) misunderstandings of collostructional analysis: On Schmid and Küchenhoff (2013). *Cognitive Linguistics* 26(3). 505–536.
Gries, Stefan Th., & Ellis, Nick C. 2015. Statistical measures for usage-based linguistics. *Language Learning* 65(Supplement 1). 1–28.
Gries, Stefan Th., Beate Hampe & Doris Schönefeld. 2005. Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16(4). 635–676.
Gries, Stefan Th., Beate Hampe & Doris Schönefeld. 2010. Converging evidence II: More on the association of verbs and constructions. In Sally Rice & John Newman (eds.), *Empirical and experimental methods in Cognitive/Functional Research*, 59–72. Stanford: CSLI Publications.
Gries, Stefan Th. & Anatol Stefanowitsch. 2004a. Co-varying collexemes in the *into*-causative. Michel Achard & Suzanne Kemmer. *Language, culture and mind*, 225–236. Stanford: CSLI Publications.
Gries, Stefan Th. & Anatol Stefanowitsch. 2004b. Extending collostructional 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 Sally Rice & John Newman (eds.), *Empirical and Experimental Methods in Cognitive/Functional Research,* 73–90. Stanford: CSLI Publications.
Hampe, Beate. 2011. Discovering constructions by means of collostruction analysis: The English denominative construction. *Cognitive Linguistics* 22(2): 211–245.
Hampe, Beate. 2014. More on the *as*-predicative: Granularity issues in the description of construction networks. In Susanne Flach & Martin Hilpert (eds.), *Yearbook of the German Cognitive Linguistics Association 2013,* 207–234. Berlin, Boston: de Gruyter.
Heine, Bernd & Heiko Narrog. 2010. Grammaticalization and linguistic analysis. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, 401–424. Oxford, New York: Oxford University Press.
Hilpert, Martin. 2008. *Germanic Future Constructions. A Usage-based Approach to Language Change*. Amsterdam, Philadelphia: John Benjamins.
Küchenhoff, Helmut & Hans-Jörg Schmid. 2015. Reply to “More (old and new) misunderstandings of collostructional analysis” by Stefan Th. Gries. *Cognitive Linguistics* 26(3). 537–547.
Levin, Beth. 1993. *English Verb Classes and Alternations. A Preliminary Investigation.* Chicago, London: The University of Chicago Press.
Newman, Mark. 2010. *Networks: An introduction.* Oxford: Oxford University Press.
Pedersen, Ted, Siddharth Patwardhan & Jason Michelizzi. 2004. WordNet: Similarity – Measuring the relatedness of concepts. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 38–41. Boston: Association for Computational Linguistics.
Schmid, Hans-Jörg, & Helmut Küchenhoff. 2013. Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24(3). 531–577.
Schönefeld, Doris. 2015. A constructional analysis of *un*-participle constructions. *Cognitive Linguistics* 26(3). 423–466.
Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.

Stefanowitsch, Anatol. 2013. Collostructional Analysis. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 290–306. Oxford, New York: Oxford University Press.

Suttle, Laura & Adele E. Goldberg. 2011. The partial productivity of constructions as induction. *Linguistics* 49(6). 1237–1269

Wulff, Stefanie. 2006. *Go*-V vs. *go-and*-V in English: A case of constructional synonymy? In Stefan T. Gries & Anatol Stefanowitsch (eds.), *Corpora in Cognitive Linguistics*, 101–126. Berlin, New York: Mouton de Gruyter.

Zeschel, Arne. 2010. Exemplars and analogy: Semantic extension in constructional networks. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches*, 201–221, Berlin, New York: Mouton de Gruyter.

Appendix

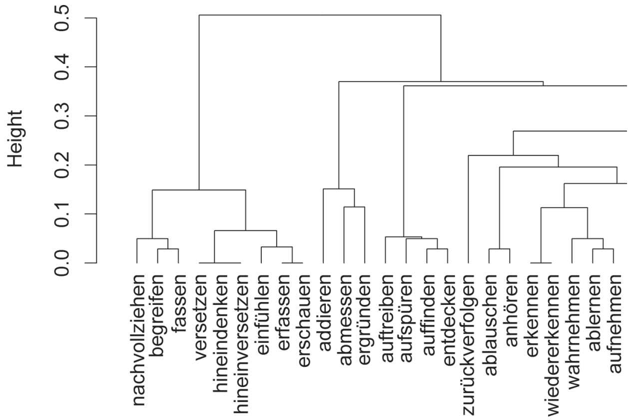

Figure 4: Dendrogram of the *vermögen*-cx: Part of cluster no 6.

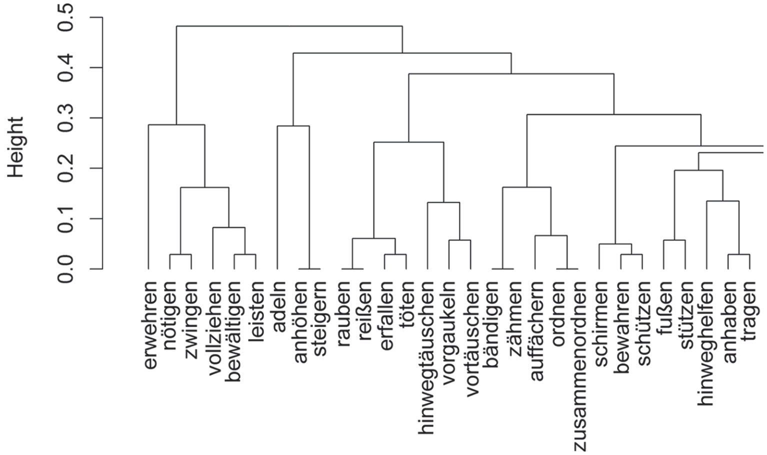

Figure 5: Dendrogram of the *vermögen*-cx: Part of cluster no 9.

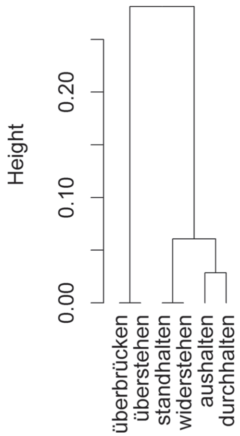

Figure 6: Dendrogram of the *vermögen*-cx: Cluster no 10.

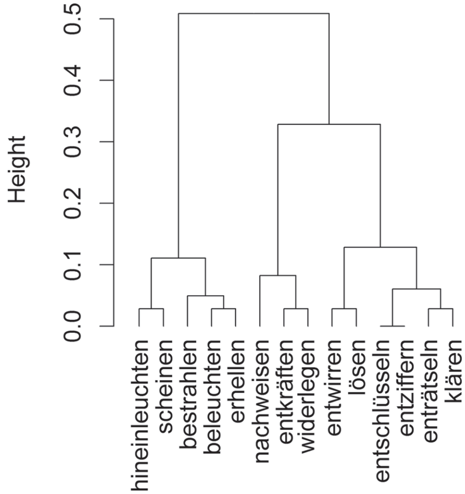

Figure 7: Dendrogram of the *vermögen*-cx: Cluster no 13.

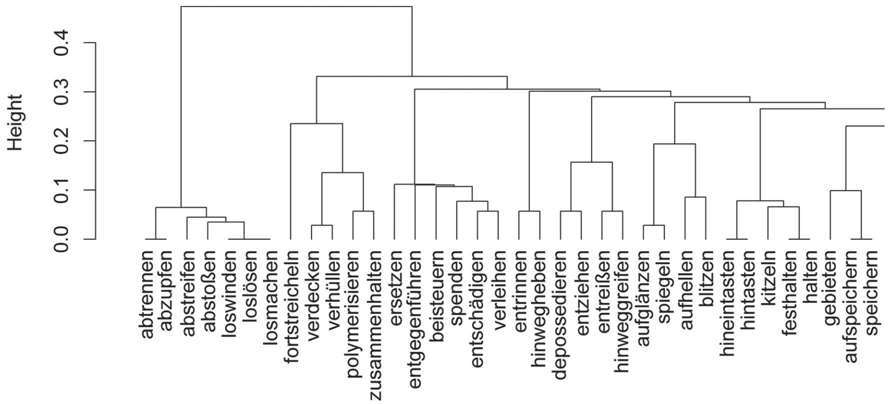

Figure 8: Dendrogram of the *vermögen*-cx: Beginning of cluster no 15.

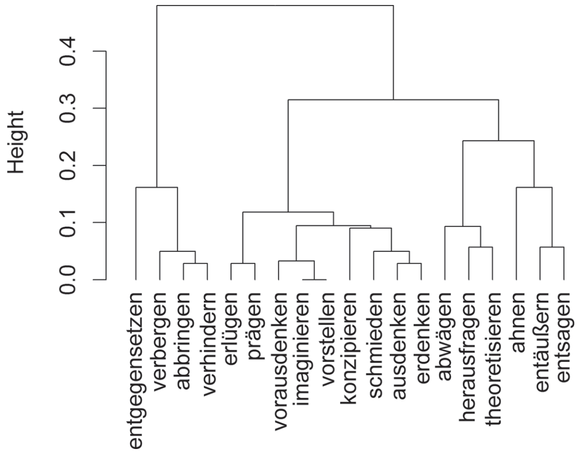

Figure 9: Dendrogram of the *vermögen*-cx: Cluster no 16.

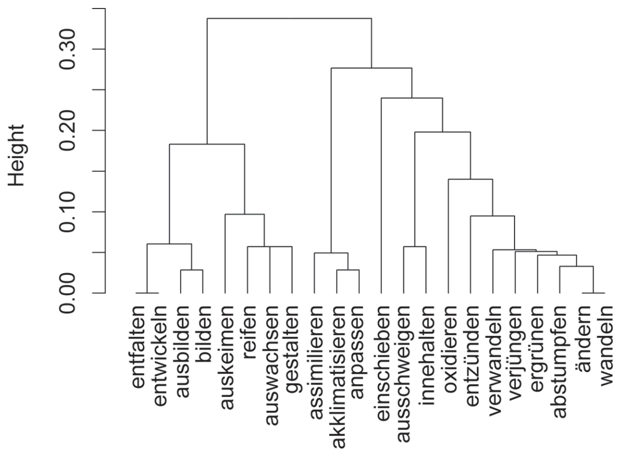

Figure 10: Dendrogram of the *vermögen*-cx: Cluster no 17.

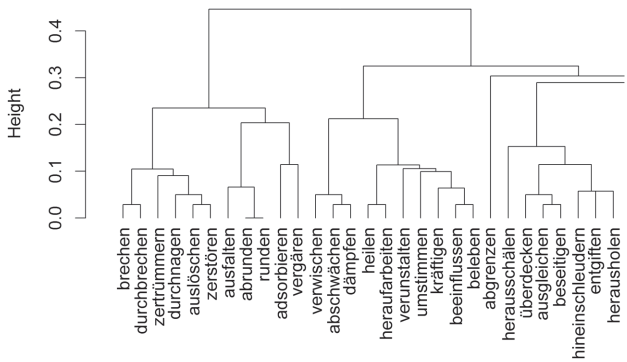

Figure 11: Dendrogram of the *vermögen*-cx: First part of cluster no 18.

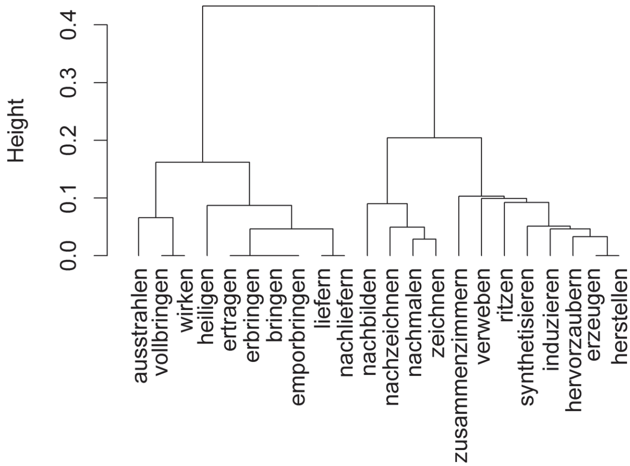

Figure 12: Dendrogram of the *vermögen*-cx: Cluster no 19.

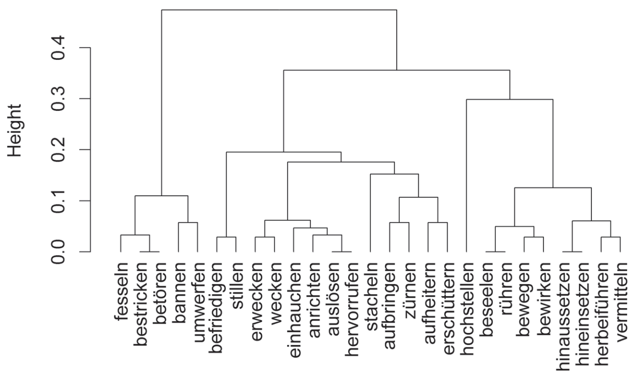

Figure 13: Dendrogram of the *vermögen*-cx: Cluster no 20.

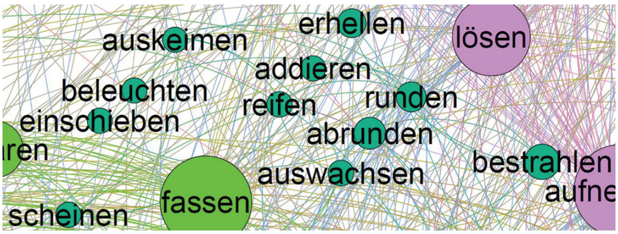

Figure 14: Enlarged image of the community encircled in Figure 3.