

Binding scale dynamics

Divjak, Dagmar

DOI:

[10.1515/9783110607963](https://doi.org/10.1515/9783110607963)

License:

Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Divjak, D 2018, Binding scale dynamics: fact or fiction? in D Van Olmen, T Mortelmans & F Brisard (eds), *Aspects of Linguistic Variation*. Trends in Linguistics. Studies and Monographs, De Gruyter, Berlin, pp. 9-42. <https://doi.org/10.1515/9783110607963>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Dagmar Divjak
Binding scale dynamics

Fact or fiction?


Abstract: This paper contributes to current debates in linguistic theory and methodology by focusing on discreteness versus continuity in linguistic description as well as on the importance of structure versus use for understanding mental representations of language phenomena. It does so through a case study on the Polish [finite verb + infinitive] construction, henceforth [Vfin Vinf]. Within a Cognitive Linguistic framework, Divjak (2007) proposed a structurally underpinned Binding Scale encompassing eight levels of looser to tighter integration, with verbs expressing modality, intention, attempt, result and phase representing the most integrated type of [Vfin Vinf] constructions. Cognitive Linguistics aims to give a usage-based account of the complex system that language is, grounded in general cognitive principles. But at which level of abstraction should we pitch the linguistic description of a system such as the [Vfin Vinf] system to find such motivating principles at work? In this paper, I assess the distance between usage and structure by investigating whether the proposed Binding Scale can be reliably distinguished in judgments of usage events through statistical unsupervised learning. By experimenting with the type of abstraction that needs to be imposed on acceptability ratings to arrive at a meaningful classification, conclusions can be drawn about the social or mental nature of this structure.

Keywords: structure, use, discreteness, continuity, cluster analysis, Polish, Binding Scale, complementation

1 The structure versus usage debate

During most of the 20th century, the classical Saussurean distinction between Langue and Parole dominated mainstream linguistic theory. Generativists took

Dagmar Divjak, University of Sheffield, School of Languages and Cultures, Jessop West, 1 Upper Hanover Street, Sheffield, S3 7RA, United Kingdom, d.divjak@sheffield.ac.uk

Open Access. © 2018 Dagmar Divjak, published by De Gruyter.  This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
<https://doi.org/10.1515/9783110607963-002>

the distinction between *Langue* and *Parole* on board, accepting there to be structural facts and usage facts that are in principle independent of each other and can be described in complete isolation from each other. Once performance errors are declared irrelevant to competence, it suffices to describe facts about structure or competence, to the neglect of use or performance. As an added bonus, allowing linguists to study an idealized version of language greatly simplified linguistic analysis.

Cognitive and functional approaches have been challenging this view for the past four decades, stressing the usage-based nature of structure. Within the functional-cognitive camp, this has led to a focus on usage facts to the extent that now structure is largely ignored. A radical usage-based approach would seem to do away with the notion of system altogether, indeed (Geeraerts 2010: 258). Yet, “accounts of language usage, language acquisition and language change are impossible without an assumption about what it is that is being used, acquired, or subjected to change. And more moderate functionalists and cognitive functionalists recognize both structural facts and usage facts as genuine facts central to the understanding of language” (Boye and Engberg-Pedersen 2010: vii).

Much cognitive and functional writing does not concern itself with characterizing the precise relationship between usage and structure. Usage is observable, but where is the structure? Geeraerts (2010: 237) suggests “a dialectal relationship between Structure and Use: individual usage events are realizations of an existing systemic structure, but at the same time, it is only through the individual usage events that changes might be introduced into the structure”. Boye and Harder (2007: 572) agree that “language is indeed based on actual, attested usage, but that it rises above attested instances in providing the speaker not only with actual usage tokens but also with a structured potential that is distilled out of previous usage”.

Structure plays no doubt a role in linguistic description and theorizing but the question that I want to pose here is whether speakers distil and store structure out of use. And if they do, how similar is the structure stored by speakers to the structure proposed by linguists?

2 The role of abstraction in linguistic description and representation

On a methodological level, the discussion about the relationship between structure and usage resurfaces as the ongoing debate about the choice for continuity

or discreteness in linguistic analysis (for a first book-length treatment, see Fuchs and Victorri 1994). In the following two sections, I will discuss the role of abstraction in linguistic description (Section 2.1) and in linguistic representation (Section 2.2).

2.1 The role of abstraction in linguistic description

Separating *Langue* from *Parole* and declaring the former to be the object of linguistic study allowed Saussureans to focus on the “neat and tidy” side of linguistics and to describe language structure independently of language use in terms of clean paradigmatic and syntagmatic relations. This discrete frame of description marginalized phenomena falling outside the realm of such an approach, a trend that was further supported by the Chomskyan focus on syntax and preference for algebraic formalizations.

Nevertheless, there have always been dissidents, denouncing the reductionism inherent in discrete models. The past few decades have witnessed a surge in explicitly continuous models, both for analysis and for representation, couched in functionally oriented frameworks. Langacker (2006) remarks that all (linguistic) models are metaphorical, and all metaphors are potentially misleading. Although, generally speaking, formalists tend towards metaphors involving discreteness while functionalists favor those based on continuity, even functionalist metaphors based on continuity such as the network model have been (rightly) criticized for being too discrete. The network model, for example, remains too discrete in the identification of sub-meanings and fails to capture the continuous dispersal of phenomena (Janda 2009: 111).

What is it that is discrete or continuous? Is continuity or discreteness a property of a (certain type of) phenomenon (see Fuchs and Victorri 1994 for semantic phenomena) or merely a characterization of the model capturing the phenomenon? The choice for continuity or discreteness comes into play in all domains of linguistic analysis (as well as outside of linguistics) and at multiple levels. Whether something is discrete or continuous is subject to construal (Langacker 2006: 114): a linguistic phenomenon is typically so complex that both discrete and continuous descriptions are appropriate, for different aspects of it. Thus, even if a phenomenon is gradual in nature, we could well gain insights from thinking about it in discrete terms, and vice versa.

Langacker (2006: 114–126) discusses a variety of ways in which phenomena can be viewed discretely or continuously. On the one hand, there are the *discretization techniques* of, first, all-or-nothing responses to gradient input and, second, zooming in to yield a higher resolution and see more detail. Discreteness can be

imposed through all-or-nothing responses to gradient input since the placement of the boundary is arbitrary and implies discontinuity where there is none. Another critical factor for discreteness is specificity, i.e., whether a phenomenon is viewed in coarse-grained or fine-grained detail. Something that appears continuous can be rendered discrete by “zooming in” to examine it at a higher resolution, where differences between individual items become visible.

On the other hand, there are *continuity-imposing measures* such as schematization and summation. Schematization ensures that two experiences become equivalent at a certain level, so that comparing them registers identity rather than disparity and thus facilitates recognition: if we apprehended everything in full, fine-grained detail, we could not build up a coherent view of the world, since every experience would be unique. Summation too yields continuous properties. Grammaticality judgments, for example, are intrinsically continuous, with deviance being the cumulative result of multiple factors. It is only when the sum of these individual factors passes a certain threshold that a clear-cut judgment of ill-formedness emerges. But any particular cut-off point is arbitrary, since the judgments are gradient. At the same time, the continuity is derivative rather than primitive, since it represents the cumulative result of numerous individual assessments.

2.2 The role of abstraction in linguistic representation

The problem of continuity versus discreteness also poses itself on a representational level. What kind of linguistic information is encoded? Structure or usage? Rules or facts? Or is the former derived from the latter?

Since rules are not “given” in the input, if they “exist”, they must be inferred from input. If we see syntactic knowledge in terms of rules, we must postulate either a rich body of innate linguistic knowledge or a sophisticated grammar induction device. There are problems with both the generativist approach, postulating a Universal Grammar, as well as with the emergentist approach, searching for a powerful grammar induction device.

Recently, proposals have been put forward that favour storage of facts, i.e., minimally different, partially overlapping exemplars. Researchers disagree as to what then happens to these exemplars. Do exemplars remain stored in clouds that (have a prototype structure? and) are efficiently searched when activated (cf. Bybee 2013) or do such rote-learned formulas form templates that gradually develop into distinct low-level schemas? In low-level schemas, none of the slots is tied to specific lexical items, as a result of storage-efficient data compression in

long-term memory (Dąbrowska 2000). Unlike the abstract rules of formal linguistics, usage-based schemas are derived from actual expressions and have the same structure as their instantiations. According to Langacker (1991: 133 and elsewhere), the function of higher level schemas in the linguistic system is primarily an organizational one.

Human beings purportedly excel at observing patterns in the speech stream (Saffran, Aslin and Newport 1996; Gomez and Gerken 1999) and abstract distributionally defined categories from input. But does pattern detection (need to) yield anything like a linguist's grammar? Distributional analysis has also proven relevant in the context of computational modeling. Redington and Chater (1997, 1998) show that distributional analysis yields relevant patterns at low and high levels of abstraction. Yet, they point out that the study of distributional information and semantics from a psychological perspective is in its infancy (Redington and Chater 1998: 183). Although the cognitive system is sensitive to features of the input, determining empirically whether infants actually exploit particular sources of distributional information to build their grammatical knowledge from the ground up remains an open question. This raises the issue of cognitive reality for results of distributional linguistic analysis.

The following survey-based study on Binding Scale dynamics in Polish is a case in point. It explores what level of granularity is ideal for describing the Binding Scale. What kind of picture emerges at a lower level of abstraction, with more detail about variation? Data for this study stems from a large survey of verbs that combine with an infinitive in Polish. Before presenting details on the measuring instrument (Section 3.1) and the data collection (Section 3.2), I will briefly introduce the [V_{fin} V_{inf}] phenomenon and its relevance to the issues outlined in Sections 1 and 2.

3 The V_{fin} V_{inf} system: diagnostics and data

Polish has more than 20,000 verbs but very few take an infinitive. Culling verbs that combine with an infinitive from the 100,000-word corpus-based dictionary *Inny Słownik* (Bańko 2000) yielded 95 such verbs (a list is provided in Appendix 1). Descriptions of the [V_{fin} V_{inf}] system are few and far between and this comes as no surprise. The [V_{fin} V_{inf}] construction is exceptional within any verbal system: usually, one verb is enough to form a full-fledged clause or sentence, as in the example *I came across a problem*. Such events are called simplex events. Sometimes, more than one verb will be used in one clause or sentence, as in *I decided to solve the problem*, with the finite verb *decided* and the infinitive [*to*]

solve. Although less than 1% of all verbs combine with an infinitive, some of the members of this category are highly frequent, such as modals or auxiliary verbs. Moreover, not all [V_{fin} V_{inf}]s are created equal: a distributional analysis shows that different finite verbs entertain links of different strength with their infinitives (Divjak 2007). In Sections 3.1.1 to 3.1.3, I will describe the set of three diagnostic tests that make it possible to differentiate between the different degrees of integration between the two verbs in a [V_{fin} V_{inf}] construction.

3.1 Diagnostic tests

The three diagnostic tests, initially proposed in Divjak (2007) (to which I refer for details and references), reveal the degree to which the two verbs or events are structurally integrated. They measure the cognitive status of the infinitive clause and the degree of integration between finite verb and infinitive by referring to the functions verbs typically fulfil. Verbs express events that have participants and this is captured in their argument structure. This observation forms the basis for the thing-test in Section 3.1.1 and for the *that*-test in Section 3.1.2. Events also take place at a certain moment in time (and space), which forms the verbs' temporal event structure. This is exploited in the time-test in Section 3.1.3.

3.1.1 The thing-test

The first diagnostic, the “thing”-test, reveals the conceptual status of the infinitive seen from the point of view of the finite verb. Very briefly, in Cognitive Grammar, nouns and verbs instantiate diverging kinds of predication (Langacker 1987: Ch. 4, 5, 6): verbs represent relational predications whereas nouns represent non-relational predications. Furthermore, nouns and verbs differ in terms of the type of entities they designate and the sort of scanning required to capture the entities they depict. Nouns are symbolic structures whose semantic poles profile things, i.e., scenes that are conceived as being unrelated to time and are scanned summarily, as a whole. Verbs profile processes or series of component states distributed through a continuous span of conceived time and are scanned sequentially, frame by frame. Infinitives are intermediary between nouns and verbs as they profile atemporal relations. Therefore, the conceptualization type typical of the (finite) verb can be determined by tracking whether the verb combines with both things and relations or only with one of them.

The question thus becomes: does a specific finite verb need an infinitive or can it do with a noun? In (1) and (2), this question is explored with pro-structures,

i.e., pro-nouns to refer to things and a pro-verb to refer to actions. If the pro-verb *do something* subsumes under the pro-nominal question *something* for a particular (lemma of a given) verb, then the verb referred to by *do something* is in essence conceptualized as a thing, despite its relational appearance as a verb.

- (1) *He planned to travel to Warsaw.*
 what?
 to do what?
- (2) *He had to travel to Warsaw.*
 **what?*
 to do what?

The verb *plan* from (1) expresses a process, i.e., it is a relational entity, and combines with infinitives, i.e., entities that, just like processes, have their own relational profile, albeit an atemporal relational profile. Yet, the question *what (does he plan) to do?* is not strictly necessary. One could also ask *what (does he plan)?* and receive as response *to travel to Warsaw*. At a more abstract, non-lexicalized level, the action expressed by the infinitive is thus reified, i.e., conceptualized as a thing. In other words, the thing-test shows that verbs like *plan* do not need another relational profile as offered by the infinitive: the infinitive can be the answer to a pro-nominal question. Thus, conceptually, *plan* treats the infinitive as any other non-relational entity it combines with. One could say that a verb like *plan* evokes conceptualization of the conceived scene expressed by the infinitive like any non-relational thing in that position, more precisely, like a direct object. The infinitival relation is thereby presented as a thing, i.e., as an entity that is scanned as a unitary whole and is made conceptually subordinate to the process expressed by *plan*.

The situation is quite different with a finite verb like *have* in (2), which exemplifies the second scenario. The infinitive that follows this verb cannot be captured by the pro-noun *what*, belonging to the argument structure of the finite verb. The question *what (did he have) to do?* remains required to obtain *to travel to Warsaw* as answer. This indicates that, with certain verbs, the infinitival relational profile cannot be backgrounded or made conceptually subordinate to that of the finite verb. The finite verb necessarily evokes the idea of another verbal relation, albeit an atemporal relation.

3.1.2 The *that*-test

Apart from differences in the “cognitive status” of the infinitive, [V_{fin} V_{inf}] patterns also differ in how “close” the second verb needs to be to the finite verb. Closeness can be judged spatially (i.e., within sentence boundaries) as well as temporally and sheds light on the strength and independence of the (finite) verb and the event it expresses.

Closeness within sentence boundaries can be determined by rephrasing the infinitive clause as a *that*-clause. Some verbs that combine with an infinitive are restricted to the [V_{fin} V_{inf}] pattern while other verbs can link to the second verb using a *that*-construction, without causing the finite verb to change its meaning. The verb *promise* can introduce *that*-complement clauses and can use these complement constructions to express the infinitival content alternatively: (3a) can be (partially) paraphrased using the pattern of (3b). Unlike *promise*, *try* does not occur with a *that*-complement clause at all, as illustrated in (4a) and (4b).

- (3) a. *She promised to tell him the truth.*
 b. *that she would tell him the truth.*
- (4) a. *She tried to tell him the truth.*
 b. **that she would tell him the truth.*

Complementation has been described in terms of conceptual subordination and dependence (Langacker 1991: 440–442). Viewing the subordinate clause as a main clause participant implies conceptual distancing that encourages summary scanning of the component states if not their reification. In other words, construing the second verb’s content as a full-fledged complement clause equals imposing a nominal construal on the second verb and the elements that depend on it and detaching that structure conceptually from the finite verb. Compare here Wierzbicka’s (1988: 132–141) and Givón’s (2001: Ch. 12) analysis of *that*-complementation in English.

Verbs that do not allow *that*-complementation and are instead restricted to combinations with infinitives share morphological and syntactic information and strict co-reference rules apply. Such verbs depend to a higher degree on the infinitive than those finite verbs that combine with an infinitive as well as with a full-fledged complement clause. Although the latter constructions also consist of two events, both events exist to a certain extent independently of one another and the infinitive event can be made subordinate to the finite verb event.

3.1.3 The time-test

The (im)possibility of modifying both verbs in a [Vfin Vinf] structure with conflicting time adverbials or adverbial expressions of time shows how the different verbs that combine with an infinitive deal with the co-temporality requirement. This provides a second measure for the degree of integration between the finite verb and the infinitive, a measure that is moreover independent of the verb's argument structure and conceptual subordination of one event to the other.

The verb *ask* could be used in a construction that locates the finite verb and the infinitive in two different and not necessarily tightly sequential moments in time. The verb *manage* demands overlap in or tight sequentiality of time. This requirement is illustrated in (5) and (6).

- (5) a. *He asked her to buy a ticket.*
 b. *Yesterday he asked her to buy a ticket tomorrow.*
- (6) a. *He managed to buy a ticket.*
 b. **Yesterday he managed to buy a ticket tomorrow*

Temporal distancing does not imply conceptual subordination. Inserting conflicting temporal specifications is a way to measure the degree of distance or integration between the two verbs in [Vfin Vinf] structures, independent from their argument structure. The occurrence of temporal distance between two events merely entails their conceptual distance. The two events take place at two different moments in time. They are construed as distinct (though related) events (Wierzbicka 1975: 497–499; Lakoff and Johnson 1980: 131; Langacker 1991: 299 fn. 11).

3.2 A theoretically supported Binding Scale

The grammaticality of using each of the verbs that combines with an infinitive in each of the three diagnostic tests can be used to build a Binding Scale, a scale of looser to tighter integration between two events (see Divjak 2007 for details). A binary approach (acceptable versus unacceptable) allows for eight logically possible combinations or degrees of integration, as shown in Table 1. Plusses indicate a positive test score for a test, minuses a negative one.

Tab. 1: Binding scale

1	2	3	4	5	6	7	8
+ thing	+ thing	+ thing	+ thing	- thing	- thing	- thing	- thing
+ that	- that	+ that	- that	+ that	- that	+ that	- that
+ time	+ time	- time	- time	+ time	+ time	- time	- time
main verbs						auxiliary verbs	

The eight different logically possible combinations of properties correlate with eight different degrees of integration between the two verbs in the [Vfin Vinf] construction. The categories were ordered according to the thing-test, followed by the time-test and, finally, by the *that*-test. The *that*-test was considered the linking diagnostic because it overlaps partially with the thing-test in that it tests for the object status of the infinitive structure and partly with the time-test in that it tests for separability.

[Vfin Vinf] combinations on the left-hand side of Table 1 score positively on all three diagnostic tests. They show the loosest type of bond and are considered multiple, independent events. [Vfin Vinf] combinations on the right-hand side of Table 1 score negatively on all three diagnostic tests. These exemplify the tightest type of bond and qualify as complex, integrated events. The finite verbs of the former combinations are considered standard main verbs while the finite verbs in the latter combinations are considered auxiliary verbs, in the most general sense of the word. Once the argument structures of each of the verbs is taken into account, several semantically coherent subgroups emerge within each category, as I demonstrated for Russian (Divjak 2007), which boasts about 300 verbs that combine with an infinitive.

In order to construct a Binding Scale for Polish, data needs to be collected on how each of the 95 Polish verbs that combines with an infinitive responds to each of the three diagnostic tests. This can be done by relying on one's intuitions or on the intuitions of a number of native speakers. In section 3.3, I will briefly discuss the way in which the acceptability of each of the 95 verbs in each of the three diagnostic tests was assessed by relying on a large sample of native speakers. In Section 4, I move on to finding semantically coherent groups in the data using cluster analysis.

3.3 Data

The vast majority of linguistic theories rest on a peculiar type of data: acceptability or grammaticality ratings. Ratings of usage events are proxies: if we accept that the system constrains the possibilities, the constructions that are licensed by the system should be judged more acceptable than the constructions that are not licensed. And more acceptable constructions should be used more frequently than constructions that are not licensed. Traditionally, these ratings were obtained through introspection by the analyst, an approach that is problematic in many (if not most) respects. Linguists have addressed (part of) the issue by eliciting ratings from larger numbers of native speakers.

Data on which to construct the Binding Scale for Polish were gathered in a large elicitation survey, following Cowart (1997), in which native speakers of Polish rated the acceptability of the 95 Polish verbs that combine with an infinitive in each of the three diagnostic tests that together reveal the degree of verb integration between the verbs in the [Vfin Vinf] structure (see Section 3.1).

Trigger sentences were constructed for each verb*test combination, i.e., all 95 verbs were used in the three test-constructions, resulting in 285 test sentences. To avoid lexical effects, three different examples were constructed per verb*construction combination. All sentences were adaptations of authentic sentences extracted from the Polish National Corpus (non-literary texts) that were comparable in complexity and length. 285 participants saw fifteen randomly selected verb*construction combinations in which fifteen different verbs were used and each of the three test-constructions was presented five times.

The trigger sentences were hidden among 30 filler sentences that are comparable in complexity and length and likewise exhibited grammaticality levels ranging from -2 to +2, as judged by native speakers. Both triggers and fillers were randomly assigned to blocks (to avoid order effects) that each contained one example of each construction type (three triggers) and one example of each mistake level (five fillers). These eight sentences were randomized within blocks, i.e., they were pseudo-randomized to ensure no questionnaire started with a trigger and triggers never followed each other. For an example, see Appendix 2.

Surveys of one page and a half were filled out in class by undergraduate students of English or German in Poland. Participants were asked to “tell me how Polish this sentence sounds” and their answers were recorded on a five-point Likert scale (-2 to +2 and ?). On this scale, they were told, -2 stands for unnatural Polish, i.e., a sentence that sounds strange and may even be difficult to understand. The middle value, 0, signaled “OK” Polish or sentences a native speaker could produce, although they are not perfect (this accommodates the strong pre-

scriptive tradition concerning the regulation and teaching of Polish to which participants would have been exposed). Finally, +2 was reserved for natural Polish sentences that are fully normal and understandable. Participants were ensured there were no right or wrong answers.

4 Finding groups in the data

Structure is an abstraction over usage data, yet very little is known about the amount of variation that is discarded in traditional linguistic analyses. In this section, I will use exploratory statistical techniques to detect natural groupings in the data and compare those to the eight degrees of integration that together make up the Binding Scale presented in Section 3.2.

The acceptability ratings were subjected to cluster analysis, an unsupervised learning technique that detects structure in data (see Baayen 2008; Johnson 2008; Gries 2009; Divjak and Fieller 2014; Levshina 2015). Cluster analysis is an exploratory data analysis technique, encompassing a number of different algorithms and methods for sorting different objects into groups. It requires the analyst to make choices about dissimilarity measures and grouping algorithms. Yet, in contrast to many other statistical methods, there seem to be fewer diagnostics informing of the weaknesses of any classification solution proposed. Therefore, “look[ing] for cluster groupings that agree with existing or expected structures” and “pick[ing] the one solution you like best” are not frivolous comments in the context of cluster analysis (Divjak and Fieller 2014: 430). Here, I will try a number of different dissimilarity measures and grouping algorithms to see whether any one combination can identify clusters that correspond to the eight degrees of integration from the Binding Scale discussed in Section 3.2.

The nature of the Likert scale used to collect grammaticality judgments poses a challenge in this respect. Whether the Likert scale is an ordinal or an interval scale is the subject of much debate. Although Likert himself assumed that the scale has interval qualities, as it was originally intended as a summated scale (after the questionnaire is completed, item responses are summed to create a score for a group of items), some consider a Likert scale to be ordinal in nature. Hence, treating the data as interval, or even ratio, is doubtful: summing ordinal data will not make it interval data, it will only make it summated ordinal data. The problem is compounded if only five levels of (dis)agreement are used, since respondents will not perceive all pairs of adjacent levels as equidistant. It has been objected, however, that, if the wording of response levels implies symmetry of response levels around a middle category, measurements would fall between ordinal and

interval level. To treat such data as ordinal could mean ignoring information it may contain. Furthermore, accompanying the item-to-be-rated with a visual analog scale where equal spacing of response levels is clearly indicated has been said to increase the likelihood that respondents construe the points as equidistant. Although both requirements were met in the questionnaires used, I remain doubtful as to whether the data could be considered anything but ordinal.

Since few clustering techniques deal with ordinal data, several work-arounds are explored, i.e., clustering summated responses (Section 4.1) and clustering summated proportions of responses (Section 4.2). Although the assumption that speakers have had less exposure to constructions they consider bad and are less likely to use such constructions themselves underlies both types of data summaries, there is a qualitative difference between these two approaches. Similarity in summated proportions of respondents assigning a particular score are slightly more precise in that they keep variation in the data, while similarities between summated responses may gloss over the very different combinations of judgments they are made up of. For example, a summed score of 10 might be the result of five respondents assigning the test construction a marginally unacceptable score or from two respondents considering the construction perfect and three others considering the construction unacceptable.

4.1 Cluster analysis on summated responses

For a first series of analyses, the fifteen ratings per verb*construction combination were summed up. Responses to several Likert questions can be summed, provided that all questions use the same Likert scale and that the scale is a defensible approximation to an interval scale, in which case they may be treated as interval data measuring a latent variable.

The data was then taken through hierarchical agglomerative cluster analysis, using `agnes()` from the package `cluster` in R, with Euclidean as the distance measure and Ward's as the amalgamation algorithm. Euclidean measures the distance between items "as the crow flies" and Ward's is known to yield small groups. The combination of both has proven to work well for linguistic data. The results are presented in the dendrogram in Figure 1. The dendrogram is read bottom up, with lower clusters representing items that are very similar and hence end up being clustered first. These lower-level clusters are then in turn grouped to form higher-level clusters and this process is repeated until all clusters are united in one overarching cluster.

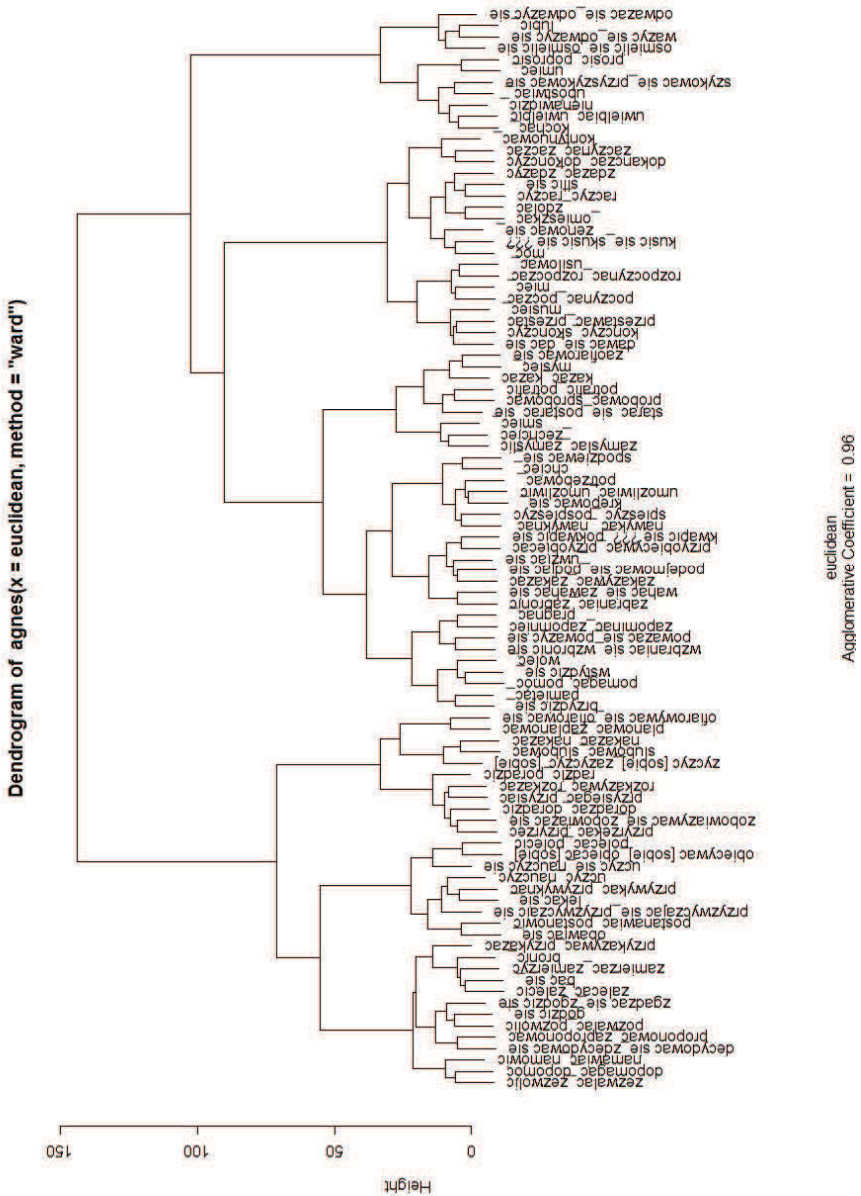


Fig. 1: Dendrogram of HAC cluster analysis on summated data with Euclidean as distance measure and Ward's as amalgamation algorithm

The agglomerative coefficient (AC), indicated at the bottom of the plot, is a measure of the clustering structure of the dataset that ranges from 0 to 1. An AC close to 1 indicates that a very clear structuring has been found whereas an AC close to 0 indicates that the algorithm has not found a natural structure. Do bear in mind that this measure is sensitive to sample size, i.e., the value goes up as the number of observations grows. In the present analysis, the AC for the dendrogram is very high (0.96) and this supports the presence of natural varieties (despite the indicator's sensitivity to the sample size).

Given the large number of clusters distinguished, a non-hierarchical cluster analysis was carried out to find the optimal clustering. This was done with `pam()` from the package `cluster` in R, using the same Euclidean distance measure. Silhouette plots were used to compare clustering solutions. These plots are read from left to right, and each silhouette represents one cluster.

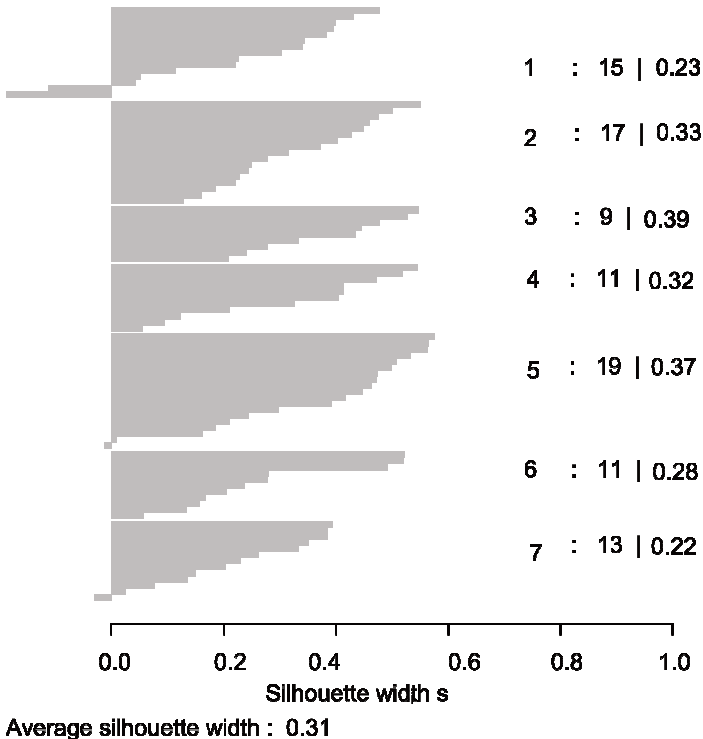


Fig. 2: Average silhouette width for seven-cluster solution

The more the silhouette shape resembles a rectangle, the higher the similarity of the elements in the cluster. The similarity is also expressed quantitatively by means of a silhouette value, which measures the degree of confidence in the clustering assignment of an observation. Well-clustered observations that are very distant from neighbouring clusters have values near 1, while poorly clustered observations that are probably assigned to the wrong cluster have values near -1. The average silhouette width is the average of the silhouette widths for all objects in the whole dataset and indicates the goodness of the overall clustering. Comparing average widths across clusterings reveals the best cluster solution. The optimal clustering solution for the data appeared to contain seven clusters, which is shown in the silhouette plot in Figure 2. Yet, each of the clusters has a relatively low silhouette width (ranging from 0.22 to 0.39) and the Average Silhouette Width for the optimal seven-cluster solution remains as low as 0.31, indicating that the proposed clustering may not be sensible.

This conclusion is confirmed by looking at the contents of each cluster. For each of the seven clusters a medoid is identified. A medoid is the most centrally located point in the given data set, representative of a data set in the sense that its average dissimilarity to all the objects in the cluster is minimal. The medoids are listed in Table 2. As mentioned in Section 3.2, the verbs in a cluster are expected to resemble each other semantically. The medoids do not show a strong semantic resemblance to the other verbs that are part of the same cluster, unfortunately. Table 3 contains details on one of the clusters listed in Table 2, i.e., the one for which the medoid is *bać się* ‘be afraid of, fear’ (the complete contents of each of the seven clusters is listed in Appendix 3). Apart from one verb, (*za*)*wahać się* ‘hesitate, waver’, all other verbs express rather the opposite of fear. There is some semantic cohesion between other verbs that are part of this cluster, however.

Tab. 2: Medoids for a non-hierarchical cluster analysis requesting 7 clusters

Cluster	Medoid	Translation
1	<i>bać się</i>	‘be afraid of, fear’
2	<i>śpieszyć</i> _ <i>pośpieszyć</i>	‘hurry, be in a hurry’
3	<i>zobowiązywać się</i> _ <i>zobowiązać się</i>	‘bind, pledge oneself’
4	<i>uwielbiać</i> _ <i>uwielbić</i>	‘adore, worship’
5	<i>kończyć</i> _ <i>skończyć</i>	‘end, finish, conclude, close’
6	<i>uczyć</i> _ <i>nauczyć</i>	‘teach, instruct’
7	<i>potrafić</i> _ <i>potrafić</i>	‘know how to, manage’

Tab. 3: Contents of one cluster resulting from non-hierarchical cluster analysis requesting seven clusters

Verb	Translation
<i>decydować się_zdecydować się</i>	‘determine, decide’
<i>pozwalać_pozwolić</i>	‘allow, permit, let’
<i>zgadzać się_zgodzić się</i>	‘agree, concur, consent’
<i>proponować_zaproponować</i>	‘offer, propose’
<i>bać się_</i>	‘be afraid of, fear’
<i>godzić się_</i>	‘agree, consent’
<i>zalecać_zalecić</i>	‘recommend, commend’
<i>przykazywać_przykazać</i>	‘order, command’
<i>bronić_</i>	‘defend, guard, vindicate, assert’
<i>namawiać_namówić</i>	‘induce, persuade’
<i>zamierzać_zamierzyć</i>	‘intend, mean, be going to’
<i>zezwałać_zezwolić</i>	‘allow, permit, let’
<i>dopomagać_dopomóc</i>	‘help, aid, assist’
<i>wahać się_zawahać się</i>	‘hesitate, waver’
<i>zakazywać_zakazać</i>	‘forbid, prohibit’

The shape of the clusters in Figure 2 and the low average silhouette width confirm that there is no clear structure. Instead, many verbs are close to verbs from other clusters. The fact that the structure found may be artificial would explain why the overarching semantics of individual clusters is difficult to capture.

4.2 Clustering summated proportions of responses

Instead of summing all judgments provided for one sentence, we could also summarize the data by proportions of respondents who assign a particular score. Summarizing by proportions of responses was done in two different ways, using the original five-point scale and a condensed three-point scale.¹

¹ Due to the instructions accompanying the rating scale, i.e., the fact that the middle point was conceived as 0 to capture the judgment “could be heard”, creating a binary solution would require second-guessing respondents’ intentions for assigning a 0 as it could mean “could be heard but I consider it unacceptable” or “could be heard and I consider it acceptable”.

4.2.1 Using a five-point rating scale

In a first analysis, proportions of responses were calculated using the original five-point ratings scale. Eight analyses were run, with both Euclidean and Manhattan distance in combination with complete, single, average linkage and Ward's amalgamation algorithms. Because both distance measures yielded virtually identical results, I will only present one set here.

The highest agglomerative coefficient was achieved by the Manhattan/Ward combination (0.87), followed by Manhattan/Complete (0.72), Manhattan/Average (0.54) and Manhattan/Single (0.29). To assess the replicability of the clustering, in the absence of an independent test-sample, p-values for all clusters contained in the clustering of the original data were calculated using the R package *pvclust*. For each cluster in hierarchical clustering, p-values are calculated via multiscale bootstrap resampling, a computer-based way of simulating similar datasets. *Pvclust* provides two types of p-values: the AU (Approximately Unbiased) p-value (on the left, normally in red) and BP (Bootstrap Probability) value (on the right, normally in green). The AU p-value, which is computed by multiscale bootstrap resampling, is a better approximation to unbiased p-value than the BP value computed by normal bootstrap resampling. Clusters that are highly supported by the data will have large p-values.

The two clusterings with the clearest structure as per the Agglomerative Coefficient do not yield any high-level replicable clusters. Based on 100 replications, the Manhattan/Ward combination yields nine clusters, each containing between two and six verbs, with AU values above 95. The likelihood that these clusters would not be found in another dataset is thus rejected at significance level 0.05. These clusters appear in (red) rectangles in Figure 3. All clusters are lower-level groupings; no higher-level clusters are likely to be found in other datasets, as the zeroes indicate. Of the lower-level groupings, only the six-verb cluster (second from the right) is semantically coherent, containing verbs like 'promise' or 'advise'. Manhattan/Complete yields a similar picture: eight replicable clusters with between two and four verbs each.

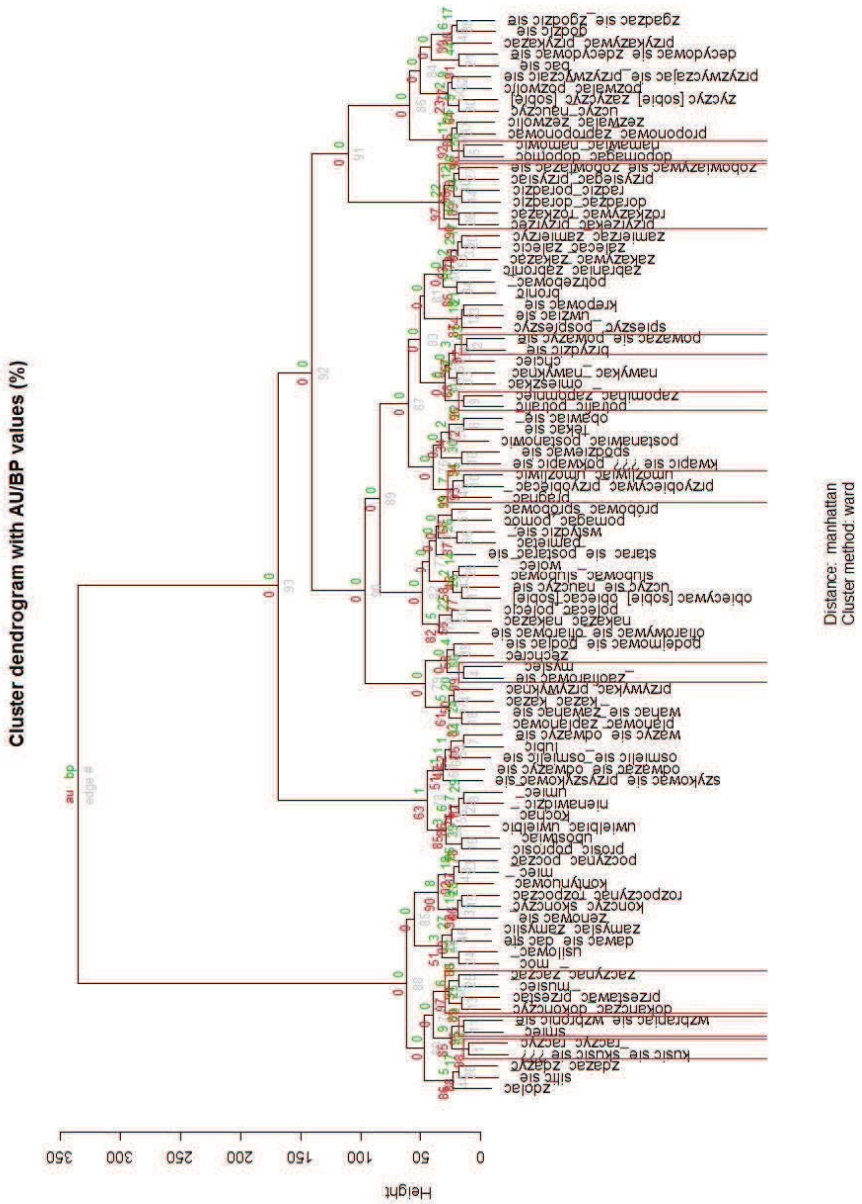


Fig. 3: Dendrogram of HAC with Manhattan/Ward and p-values on five-point scale

In other words, working with five levels of acceptability results in many low-level clusters. It is unclear from the data, however, what would motivate these clusters. If linguists would like to prefer low-level generalizations over high-level ones, some form of similarity between the verbs in one cluster would be expected. Dąbrowska (2008), for example, found that speakers prefer low-level generalizations over clusters of phonologically similar forms or clusters of words sharing the same derivational affix to more global generalizations. The clusters do, however, not contain verbs resembling each other from a semantic point of view and there is no phonological or morphological similarity either. It is rare to find a cluster containing infinitives ending in the same suffix, having a reflexive pronoun or exhibiting the same morphological aspectual alternation pattern.

4.2.2 Using a three-point rating scale

Clusters containing only two to four verbs contribute little to our understanding of the category of [Vfin Vinf] verbs as a whole. Therefore, in a next step, the five scoring options were reduced to three, by collapsing the scores -2 and -1 as well as 1 and 2. The same eight analyses as described in Section 4.2.1 were run, four with the Euclidean distance measure and four with Manhattan. For both sets, the agglomerative coefficients are the same depending on the amalgamation strategy used. Ward's does best, while Single linkage performs most poorly.

Of the clusterings run with the Euclidean distance measure, Ward-based clusterings achieve an agglomerative coefficient over 0.90 (both Euclidean/Ward and Manhattan/Ward get 0.93) while Complete-based clusterings receive an agglomerative coefficient over 0.80 (Manhattan/Complete gets 0.83 and Euclidean/Complete gets 0.82). Manhattan/Average gets 0.69 and Euclidean/Average 0.68 while Euclidean/Single gets 0.41 and Manhattan/Single 0.39.

These analyses were followed up with pvclust, to determine which clusters could be expected to replicate. Using pvclust with 1000 repetitions to assess the uncertainty in the Euclidean/Ward hierarchical cluster analysis, the two overarching groups that are amalgamated last both receive AU (approximately unbiased) p-values of 99. In other words, the hypothesis that these clusters do not exist is rejected at significance level 0.01.

The highlighted clusters in Figure 4, one on the left-hand side containing 22 verbs and the other one containing all remaining verbs, do not only seem to exist because of sampling error but may be stably observed if we increase the number of observations. The second-best clustering (running on Euclidean/Complete, not pictured here) suggests different clusters would replicate. The same high-level cluster of 22 verbs emerges but it is complemented by a medium-level seven-verb cluster expressing attitudes such as ‘like’ or ‘detest’, as well as by fifteen low-level clusters containing between two and four verbs each. These smaller clusters remain semantically unmotivated.

The two clusters in Figure 4 that are amalgamated last are of most interest from the point of view of the Binding Scale introduced in Section 3.2. It is also important that the leftmost cluster falls out of the second-best clustering as well. The two high-level clusters correspond to what I earlier called main verbs and auxiliary verbs respectively. The leftmost cluster contains the so-called auxiliary verbs whereas the rightmost cluster contains all the other verbs. In other words, auxiliary verbs behave differently enough from all other verbs to be rated in such a way by naïve speakers that they are picked up by a clustering program. The verbs listed in Table 4 qualify as auxiliary verbs. This diverse group of so-called auxiliary verbs is consistent with the results for English (Givón 2001: 54–58) and Russian (Divjak 2007), where semantic clusters of verbs expressing modality, intention, attempt, result and phase are attested within the category of auxiliary verbs. Comparable findings have been reported for non-Indo-European language systems, which may use verbal affixes, modifiers to a verb (including both adverbs and modal verbs) and non-inflecting particles within a clause to express similar concepts (Dixon 1996: 178).

Tab. 4: Replicating cluster of verbs with Manhattan/Ward on three-point scale

Verb	Translation	Classification
<i>_zdolać</i>	‘be able’	result
<i>_zechcieć</i>	‘become willing’	volition
<i>dawać się_dać się</i>	‘be possible, allow itself’	modality
<i>dokańczać_dokończyć</i>	‘finish up, conclude’	phase
<i>kończyć_skończyć</i>	‘end, finish, conclude, close’	phase
<i>kontynuować_</i>	‘continue’	phase
<i>kusić się_skusić się</i>	‘seek to obtain, attempt’	attempt
<i>mieć_</i>	‘have to’	modality
<i>móc_</i>	‘can, be able’	modality

Verb	Translation	Classification
<i>musieć_</i>	'be obliged to, have to'	modality
<i>począć_począć</i>	'begin, originate'	phase
<i>przestać_przestac</i>	'cease, stop, discontinue'	phase
<i>raczyć_raczyć</i>	'deign, condescend'	result
<i>rozpocząć_rozpocząć</i>	'begin, start, commence'	phase
<i>silić się_</i>	'make efforts, exert oneself'	attempt
<i>smieć_</i>	'dare, venture'	NA
<i>usiłować_</i>	'make efforts, endeavor, attempt'	attempt
<i>wzbraniać się_wzbronić się</i>	'forbid'	NA
<i>zaczynać_zacząć</i>	'begin, start, commence'	phase
<i>zamyslać_zamyslić</i>	'design'	volition
<i>zdać_zdać</i>	'manage to do on time'	result
<i>żenować się_</i>	'feel embarrassed'	NA

5 Is there a system in the variation?

It has been claimed that language is a social fact, an observable regularity in language use realized by a specific community. But it is also a cognitive fact because the members of the community have an internal representation of the existing regularities that allows them to realize the same system in their own use of the language (Geeraerts 2010: 237–238). In the case of the [Vfin Vinf] constructions discussed in this paper, would the proposed Binding Scale fall out of a social interpretation of acceptability ratings for the diagnostics that motivate the system? And how much of any Binding Scale would speakers need to have internalized to yield judgments that would seem to support the abstract system?

The one clear result that emerged from a series of cluster analyses supports a bifurcation of [Vfin Vinf] constructions into those built on a finite verb that is a main verb and those built on a finite verb that is an auxiliary verb. Small low-level classes exist but it is unlikely that there would be any widely shared local prototypes given that those lower-level classes did not exhibit any phonological, morphological or semantic coherence, which would be required to elevate the verb*construction combination from lexical idiosyncrasy to lower-level schema. Individual local prototypes may, however, have guided the ratings for individual

respondents and any divergence between these local prototypes may have further increased the variability in the data. The cline of eight different degrees of integration between the events expressed by means of a [Vfin Vinf] construction could not be reconstructed from acceptability ratings, when submitted to a (standard) statistical technique designed to find groups in data.

The observed two-way classification fell out from data summarized as the proportion of respondents who assigned a score on a three-point scale, i.e., it is a social construct and the result of summation and schematization. Summing the number of individuals who assigned a particular rating registered tendencies within the group of respondents. The scales had to tip for a (more) clear-cut judgment of ill-formedness to emerge. This process was facilitated by schematization: reducing the five-point scale to a three-point scale ensured that two experiences had a better chance of becoming equivalent, so that comparing them registered identity rather than disparity, thereby facilitating categorization.

The Binding Scale, like any other linguistic classification, abstracts away from variation to reveal the skeleton of a system that, if built on well-motivated diagnostic principles, should apply to a number of languages. For this study, usage data was used to populate the cells. A sufficient number of speakers of Polish recognized the syntactic limitations on auxiliary verbs for them to emerge as a category at the social level. The sample of speakers that I polled appears to have a strong aversion towards using auxiliary verbs in any other constructions than [Vfin Vinf]. At the same time, speakers diverged in their assessment of the extent to which the three diagnostic constructions are felicitous for main verbs. Because of the variation in their judgments, no crisply delineated categories of main verbs arise at the participants' group (i.e., social) level. This may mean that the finer details of the classification are not mentally real for any speakers, or maybe only for a small subgroup.

In this case, the Binding Scale could be partly reconstructed on the basis of acceptability data on the diagnostics but only if that data is summarized so as to reveal its social basis. The cluster analyses suggest that the Binding Scale captured conventionalization in society, not entrenchment in the mind. Language is very likely a complex adaptive system (Beckner et al. 2009) in which knowledge of the system's individual parts does not imply understanding of the system. The local agents or speakers know their task but the teleology of the system remains out of their grasp – if there is a goal to the overarching system at all. Knowledge is socially distributed: while each speaker individually knows part(s) of the system, no one speaker knows them all. By putting this distributed knowledge together, a picture of a socially supported system emerges, that in its entirety is unlikely mentally real for any one agent.

These findings limit what usage-based linguists, working within a cognitive framework, can expect from theoretical models that are not built on usage data from a large number of speakers but on binary acceptability judgments from an individual. Even if a proposed account is theoretically justified and each diagnostic has a plausible cognitive explanation, the overarching model may well lack psychological reality for other speakers of the language.

Acknowledgement: I am grateful to Neil Bermel, Petar Milin, James Street and two anonymous reviewers for commenting on an earlier version of this paper.

References

- Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bańko, Mirosław. 2000. *Inny Słownik języka polskiego* [Another dictionary of Polish]. Warsaw: PWN.
- 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 (1). 1–26.
- Boye, Kasper & Peter Harder. 2007. Complement-taking predicates: Usage and linguistic structure. *Studies in Language* 31 (3). 569–606.
- Boye, Kasper & Elisabeth Engberg-Pedersen. 2010. Introduction. In Kasper Boye & Elisabeth Engberg-Pedersen (eds.), *Language usage and language structure*, viii–xvi. Berlin: De Gruyter.
- Bybee, Joan. 2013. Usage-based theory and exemplar representation. In Thomas Hoffman & Graeme Trousdale (eds.) *The Oxford handbook of construction grammar*, 49–69. Oxford: Oxford University Press.
- Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage Publications.
- Dąbrowska, Ewa. 2000. From formula to schema: The acquisition of English questions. *Cognitive Linguistics* 11 (1/2). 83–102.
- Dąbrowska, Ewa. 2008. The later development of an early-emerging system: The curious case of the Polish genitive. *Linguistics* 46 (3). 629–650.
- Divjak, Dagmar. 2007. Degrees of verb integration: A binding scale for [VV] structures in Russian. In Dagmar Divjak & Agata Kochańska (eds.), *Cognitive paths into the Slavic domain*, 221–256. Berlin: De Gruyter.
- Divjak, Dagmar & Nick Fieller. 2014. Clustering linguistic data. In Dylan Glynn & Justyna Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 405–441. Amsterdam: John Benjamins.
- Dixon, R.M.W. 1996. Complement clauses and complementation strategies. In Frank R. Palmer (ed.), *Grammar and meaning: Essays in honour of Sir John Lyons*, 175–220. Cambridge: Cambridge University Press.

- Fuchs, Catherine & Bernard Victorri (eds). 1994. *Continuity in linguistic semantics*. Amsterdam: John Benjamins.
- Geeraerts, Dirk. 2010. Schmidt redux: How systematic is the linguistic system if variation is rampant? In Kasper Boye & Elisabeth Engberg-Pedersen (eds.), *Language usage and language structure*, 237–262. Berlin: De Gruyter.
- Givón, Talmy. 2001. *Syntax: An introduction*. Amsterdam: John Benjamins.
- Gomez, Rebecca L. & Louann Gerken. 1999. Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* 70 (2). 109–135.
- Gries, Stefan Th. 2009. *Quantitative corpus linguistics with R: A practical introduction*. London: Routledge.
- Janda, Laura A. 2009. What is the role of semantic maps in cognitive linguistics? In Piotr Stalmaszczyk & Wiesław Oleksy (eds.), *Cognitive approaches to language and linguistic data: Studies in honor of Barbara Lewandowska-Tomaszczyk*, 105–124. Frankfurt am Main: Peter Lang.
- Johnson, Keith. 2008. *Quantitative methods in linguistics*. Oxford: Wiley-Blackwell.
- Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald. 1991. *Foundations of cognitive grammar: Descriptive application*. Stanford: Stanford University Press.
- Langacker, Ronald W. 2006. On the continuous debate about discreteness. *Cognitive Linguistics* 17 (1). 107–151.
- Levshina, Natalia. 2015. *How to do linguistics with R*. Amsterdam: John Benjamins.
- Redington, Martin & Nick Chater. 1997. Probabilistic and distributional approaches to language acquisition. *Trends in Cognitive Sciences* 1 (7). 273–281.
- Redington, Martin & Nick Chater. 1998. Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes* 13 (2/3). 129–191.
- Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport. 1996. Statistical learning by 8-month old infants. *Science* 274 (5294). 1926–1928.
- Wierzbicka, Anna. 1975. Why ‘kill’ does not mean ‘cause to die’: The semantics of action sentences. *Foundations of Language* 13 (4). 491–528.
- Wierzbicka, Anna. 1988. *The semantics of grammar*. Amsterdam: John Benjamins.

Appendix 1: List of verbs that combine with an infinitive

	Verb (imperfective_perfective aspect)	Translation
1	<i>zezwałać_zezwoić</i>	‘allow, permit, let’
2	<i>brzydzić się_</i>	‘abhor, loathe, have an aversion’
3	<i>przyrzekać_przyrzec</i>	‘promise’

	Verb (imperfective_perfective aspect)	Translation
4	<i>kochać_</i>	'love'
5	<i>wzbraniać się_wzbronić się</i>	'forbid'
6	<i>ośmielić się_ośmielić się</i>	'venture, dare'
7	<i>zamyślać_zamyślić</i>	'design'
8	<i>obawiać się_</i>	'fear, be afraid, be anxious'
9	<i>umieć_</i>	'know how, be able'
10	<i>starać się_postarać się</i>	'endeavor, make efforts, take pain, try'
11	<i>decydować się_zdecydować się</i>	'determine, decide'
12	<i>dawać się_dać się</i>	'let, allow'
13	<i>pozwalać_pozwolić</i>	'allow, permit, let'
14	<i>przyzwyczajając się_przyzwyczaić się</i>	'become accustomed, get used'
15	<i>począć_począć</i>	'begin, originate'
16	<i>zabraniać_zabronić</i>	'forbid, prohibit, interdict'
17	<i>życzyć [sobie]_zażyczyć [sobie]</i>	'wish, desire'
18	<i>kazać_kazać</i>	'bid, order, let'
19	<i>proponować_zaproponować</i>	'offer, propose'
20	<i>zakazywać_zakazać</i>	'forbid, prohibit'
21	<i>móc_</i>	'can, be able'
22	<i>poważać się_poważyć się</i>	'dare'
23	<i>nawykać_nawyknąć</i>	'become accustomed'
24	<i>pomagać_pomóc</i>	'help, aid, assist'
25	<i>przysięgać_przysiąc</i>	'swear'
26	<i>próbować_spróbować</i>	'try, test, attempt'
27	<i>radzić_poradzić</i>	'advise'
28	<i>dokańczać_dokończyć</i>	'finish up, conclude'
29	<i>ślubować_ślubować</i>	'vow, make a vow'
30	<i>uczyć się_nauczyć się</i>	'learn'
31	<i>śpieszyć_pośpieszyć</i>	'hurry, be in a hurry'
32	<i>ubóstwiać_</i>	'idolize, adore'
33	<i>woleć_</i>	'prefer'
34	<i>kończyć_skończyć</i>	'end, finish, conclude, close'
35	<i>_zechcieć</i>	'become willing'
36	<i>godzić się_</i>	'agree, consent'
37	<i>nienawidzić_</i>	'hate, detest'
38	<i>pamiętać_</i>	'remember, keep in mind'
39	<i>obiecywać [sobie]_obiecać [sobie]</i>	'promise'

	Verb (imperfective_perfective aspect)	Translation
40	<i>_omieszkąć</i>	'fail'
41	<i>planować_zaplanować</i>	'plan'
42	<i>mieć_</i>	'have to'
43	<i>zobowiązywać się_zobowiązać się</i>	'bind, pledge oneself'
44	<i>_uwziąć się</i>	'set one's mind, become crazy'
45	<i>śmieć_</i>	'dare, venture'
46	<i>dopomagać_dopomóc</i>	'help, aid, assist'
47	<i>rozpocząć_rozpocząć</i>	'begin, start, commence'
48	<i>wstydzić się_</i>	'be ashamed'
49	<i>zgodzać się_zgodzić się</i>	'agree'
50	<i>kusić się_skusić się</i>	'seek to obtain, attempt'
51	<i>zalecać_zalecić</i>	'recommend, commend'
52	<i>zapominać_zapomnieć</i>	'forget'
53	<i>krępować się_</i>	'be embarrassed, feel uneasy'
54	<i>potrzebować_</i>	'need, want, be in need of'
55	<i>bronić_</i>	'defend, guard, vindicate, assert'
56	<i>raczyć_raczyć</i>	'deign, condescend'
57	<i>silić się_</i>	'make efforts, exert oneself'
58	<i>nakazać_nakazać</i>	'order, command'
59	<i>zaczynać_zacząć</i>	'begin, start, commence'
60	<i>bać się_</i>	'be afraid of, fear'
61	<i>postanawiać_postanowić</i>	'resolve, determine, make up one's mind'
62	<i>potrafić_potrafić</i>	'know how to do, manage'
63	<i>uwielbiać_uwielbić</i>	'adore, worship'
64	<i>musieć_</i>	'be obliged to, have to'
65	<i>odwagać się_odważyć się</i>	'dare, venture'
66	<i>usiłować_</i>	'make efforts, endeavor, attempt'
67	<i>ważyć się_odważyć się</i>	'dare, venture'
68	<i>doradzać_doradzić</i>	'advise'
69	<i>pragnąć_</i>	'desire'
70	<i>zdążyć_zdążyć</i>	'manage to do (on time)'
71	<i>prosić_poprosić</i>	'ask, beg, request'
72	<i>chcieć_</i>	'want, be willing, intend, desire, wish'
73	<i>przyobiecować_przyobiecąć</i>	'promise'
74	<i>polecać_polecić</i>	'recommend'
75	<i>_zdotać</i>	'be able'

	Verb (imperfective_perfective aspect)	Translation
76	<i>myśleć_</i>	'think, mean'
77	<i>zamierzać_zamierzyć</i>	'intend, mean, be going'
78	<i>wahać się_zawahać się</i>	'hesitate, weaver'
79	<i>umożliwiać_umożliwić</i>	'enable, make possible'
80	<i>lękać się_</i>	'fear, be anxious'
81	<i>kwapić się_pokwapić się</i>	'be eager'
82	<i>ofiarowywać się_ofiarować się</i>	'offer (oneself)'
83	<i>spodziewać się_</i>	'hope, expect'
84	<i>uczyć_nauczyć</i>	'teach, instruct'
85	<i>podejmować się_podjąć się</i>	'undertake'
86	<i>kontynuować_</i>	'continue'
87	<i>lubić_</i>	'like, love'
88	<i>przestawać_przestać</i>	'cease, stop, discontinue'
89	<i>szkować się_przyszkować się</i>	'prepare (oneself)'
90	<i>przykazywać przykazać</i>	'order, command'
91	<i>_zaofiarować się</i>	'offer (oneself)'
92	<i>namawiać_namówić</i>	'induce, persuade'
93	<i>rozkazywać_rozkazać</i>	'order, command'
94	<i>przywykać_przywyknąć</i>	'get accustomed to'
95	<i>żenować się_</i>	'feel embarrassed'

Appendix 2: Example questionnaire

Trigger sentences for each of the verbs in each of the three constructions were composed. To ensure naturalness as much as possible, the sentences were adapted from authentic sentences from the non-literary text sections from the PNC. The raw material for the sentences was extracted from the test version of the PNC (66 million words). Raw sentences were taken from written periodicals. If no examples were found, both dictionaries and (near-)native speakers were consulted. The sentences were then altered to contain the test constructions. To ensure comparability, every trigger item consisted of two sentences that formed a whole and could stand alone, i.e., were not context dependent. All sentences are declarative statements. Positive sentences were used unless there was a clear counter indication that the verb favored negative contexts. Sentence subjects are

male/female third person singular/plural. Finite verbs are past and perfective (if possible). Infinitives are proportional to ‘do something’.

The following is an example of one block. The capital letters A, B and C refer to the diagnostic tests (the thing-, that- and time-tests respectively). Small letters a, b and c refer to the lexical set, while numbers identify the verb. The capital letter F indicates filler sentences.

- Ac42 *Mieszkańcy Kołobrzegu mieli jeść, spać i oglądać telewizję w blokach poza centrum. Mieli to, aż nie naprawili przewodu gazowego w centrum.*
‘The inhabitants of K had to eat, sleep and watch TV in apartment buildings outside the center. They had this, until they fixed the gas pipes in the center.’
[example of an infelicitous thing- test]
- F8 *FBI prowadziło operację specjalną. Prowadzono operację w tak głębokiej tajemnicy, że w pewnym momencie nawet sam prezydent nie był do końca poinformowany.*
- F11 *Demokracja to dla wielu ludzi rzecz oczywista o której nie myślą. Nie wiedzą co to jest żyć w dyktaturze.*
- Ba23 *Jest w złym humorze, bo nawykł urlop spędzać w Kalifornii. Jak człowiek już nawykł, żeby spędzać urlop w słonecznym miejscu, to polskich deszczowych lat nie uwielbia.*
‘He is in a bad mood, because he is used to spending his holidays in California. Once you are used to spending your holidays in a sunny place, you no longer love Polish rainy years.’ [example of a felicitous that-test]
- F1 *Berlin był miastem podzielonym murem. W 1989 roku ludzie z obu stron zaczęli rozwalać mur.*
- F17 *Sztucer to broń myśliwska na grubego zwierza. Zawsze brał właśnie sztucer kiedy chodził na polowania.*
- Cc92 *Po wyborach był całkiem rozczarowany. On był jednym z tych, którzy podczas kampanii wyborczej namówili członków zespołu w dzień wyborów wesprzeć Kerry'ego do Białego Domu.*
‘The elections had left him completely disappointed. He was one of those who during the election campaign had talked members of the team into supporting Kerry into the White House on election day.’ [example of a felicitous time-test]
- F25 *Zgodnie z prawem księcia chronił królewski immunitet. Tylko królowa mogła go zdecydować o ukaraniu go jak normalnego obywatel*

Appendix 3: Contents of each of the seven clusters supported by K-means analysis on summated responses

Verb in cluster 1	Translation
<i>decydować się_zdecydować się</i>	'determine, decide'
<i>pozwalać_pozwolić</i>	'allow, permit, let'
<i>zgadzać się_zgodzić się²</i>	
<i>proponować_zaproponować</i>	'offer, propose'
<i>bać się_</i>	'be afraid of, fear'
<i>godzić się_</i>	'agree, consent'
<i>zalecać_zalecić</i>	'recommend, commend'
<i>przykazywać_przykazać</i>	'order, command'
<i>bronić_</i>	'defend, guard, vindicate, assert'
<i>namawiać_namówić</i>	'induce, persuade'
<i>zamierzać_zamierzyć</i>	'intend, mean, be going to'
<i>zezwałać_zezwolić</i>	'allow, permit, let'
<i>dopomagać_dopomóc</i>	'help, aid, assist'
<i>wahać się_zawahać się</i>	'hesitate, weaver'
<i>zakazywać_zakazać</i>	'forbid, prohibit'
Verb in cluster 2	Translation
<i>spieszyć_pospieszyć</i>	'hurry, be in a hurry'
<i>umożliwiać_umożliwić</i>	'enable, make possible'
<i>krępować się_</i>	'be embarrassed, feel uneasy'
<i>spodziewać się_</i>	'hope, expect'
<i>pragnąć_</i>	'desire'
<i>potrzebować_</i>	'need, want, be in need of'
<i>nawykąć_nawyknąć</i>	'become accustomed'
<i>_uwziąć się</i>	'set one's mind, become crazy'
<i>chcieć_</i>	'want, be willing, intend, desire, wish'
<i>kwapić się_pokwapić się</i>	'be eager'

2 Translations are missing if they were not included in Polish-English dictionaries.

Verb in cluster 2	Translation
<i>_omieszkąć</i>	'fail'
<i>przyobiecować _przyobiecać</i>	'promise'
<i>podejmować _podjąć się</i>	'undertake'
<i>wzbraniać _wzbronić się</i>	'forbid'
<i>zapominać _zapomnieć</i>	'forget'
<i>brzydzić się _</i>	'abhor, loathe, have an aversion'
<i>_zdotać</i>	'be able'
<i>zobowiązywać _zobowiązać się</i>	'bind, pledge oneself'
<i>radzić _poradzić</i>	'advise'
<i>przrzekać _przrzec</i>	'promise'

Verb in cluster 3	Translation
<i>doradzać _doradzić</i>	'advise'
<i>rozkazywać _rozkazać</i>	'order, command'
<i>planować _zaplanować</i>	'plan'
<i>życzyć[sobie] _zazyczyć[sobie]</i>	'wish, desire'
<i>ofiarowywać _ofiarować</i>	
<i>przysięgać _przysiąc</i>	'swear'

Verb in cluster 4	Translation
<i>uwielbiać _uwielbić</i>	'adore, worship'
<i>ubóstwiać _</i>	'idolize, adore'
<i>kochać _</i>	'love'
<i>ważyć _odważyć się</i>	'dare, venture'
<i>szkować _przyszykować się</i>	
<i>lubić _</i>	'like, love'
<i>nienawidzić _</i>	'hate, detest'
<i>ośmielić _ośmielić się</i>	'venture, dare'
<i>umieć _</i>	'know how, be able'
<i>odważać _odważyć się</i>	'dare, venture'
<i>prosić _poprosić</i>	'ask, beg, request'

Verb in cluster 5	Translation
<i>kończyć_skończyć</i>	'end, finish, conclude, close'
<i>przestawać_przestać</i>	'cease, stop, discontinue'
<i>począć_począć</i>	'begin, originate'
<i>dawać_dać się</i>	
<i>mieć_</i>	
<i>musieć_</i>	'be obliged to, have to'
<i>usiłować_</i>	'make efforts, endeavor, attempt'
<i>kusić_skusić się</i>	'seek to obtain, attempt'
<i>rozpocząć_rozpocząć</i>	'let, allow'
<i>kontynuować_</i>	'have to'
<i>móc_</i>	'can, be able'
<i>dokończyć_dokończyć</i>	'finish up, conclude'
<i>zdążyć_zdążyć</i>	'manage to do on time'
<i>zaczynać_zacząć</i>	'begin, start, commence'
<i>zamyślać_zamyślić</i>	'design'
<i>żenować się_</i>	'feel embarrassed'
<i>raczyć_raczyć</i>	'deign, condescend'
<i>silić się_</i>	'make efforts, exert oneself'
<i>smieć_</i>	'dare, venture'
Verb in cluster 6	Translation
<i>obiecywać[sobie]_obietcać[sobie]</i>	'promise'
<i>uczyć_nauczyć</i>	'teach, instruct'
<i>polecać_polecić</i>	'recommend'
<i>lękać się_</i>	'fear, be anxious'
<i>uczyć_nauczyć się</i>	'learn'
<i>przywykać_przywyknąć</i>	'get accustomed to'
<i>postanawiać_postanowić</i>	'resolve, determine, make up one's mind'
<i>ślubować_ślubować</i>	'vow, make a vow'
<i>nakazać_nakazać</i>	'order, command'
<i>obawiać się_</i>	'fear, be afraid, be anxious'
<i>przyzwyczajają_przyzwyczajają się</i>	'become accustomed, get used to'

Verb in cluster 7	Translation
<i>starać_postarać się</i>	'endeavor, make efforts, take pain, try'
<i>próbować_spróbować</i>	'try, test, attempt'
<i>kazać_kazać</i>	'bid, order, let'
<i>myśleć_</i>	'think, mean'
<i>potrafić_potrafić</i>	'know how to, manage'
<i>_zaofiarować się</i>	
<i>pomagać_pomóc</i>	'help, aid, assist'
<i>wstydzić się_</i>	'be ashamed'
<i>woleć_</i>	'prefer'
<i>pamiętać_</i>	'remember, keep in mind'
<i>_zechcieć</i>	'become willing'
<i>zabraniać_zabronić</i>	'forbid, prohibit, interdict'
<i>powążyć_powążyć się</i>	