Theme session proposal

Constructions at the mid-level of abstraction: linguistic diversity, variation and context

Convenors:
NATALIA LEVSHINA (Leipzig University) natalia.levshina@uni-leipzig.de
DORIS SCHÖNEFELD (Leipzig University) schoenefeld@uni-leipzig.de

Usage-based linguistics has placed strong emphasis on explaining schematic patterns as results of abstracting from repetitive and similar individual usage events. In between these layers, at the mid-levels of linguistic abstraction, speakers are assumed to have stored many mixed patterns consisting of schematic and lexical material. Material of this kind has been shown to play an important part in the organization of a speaker’s knowledge of his/her own language. Its effects could be traced in studies of language representation (e.g. Stefanowitsch & Gries 2003 and later works), acquisition (e.g. Tomasello 2005; Lieven & Ambridge 2011), processing (e.g. Gries et al. 2005) and change (e.g. Hilpert 2012). The phenomena in question, which have been made measurable by various types of association measures, come to the fore in the study of usage as patterns emerging through repetition; they are reminiscent of the Sinclairian differentiation between the open-choice and idiom principles (Sinclair 1991: 109–115). The claim derived from such patterns is that usage relies on lexico-grammatical patterns much more intensely than the traditional distinction between lexicon and grammar, or words and schemata, etc., implies (e.g. Boas 2003; Hampe & Schönefeld 2006).

Such patterns have enjoyed immense popularity in cognitive linguistic research since the seminal works on collostructional analysis (Stefanowitsch & Gries 2003 and later). However, most lexico-grammatical studies are based on one language (variety), most commonly the written one. At the same time, cognitive linguists nowadays take a growing interest in language-internal variation (e.g. Geeraerts 2005; Croft 2009; Schmid 2015) and linguistic diversity (cf. the special topic of the 14th International Cognitive Linguistics Conference). Although there have been a few variational collostructional studies (e.g. dialectal variation by Wulff et al. 2007 and register variation by Schönefeld 2013), as well as cross-linguistic collostructional analyses (e.g. Gilquin 2015), such studies are the exception rather than the rule. Moreover, most studies of lexico-grammatical patterns usually focus only on the co-occurrence of words and schemata, neglecting the gestural, interactional and other cues arising from the multimodal nature of language use.

This theme session aims at filling this gap at two complementary levels, which together provide a fully usage-based perspective, in accordance with the ongoing recontextualization of linguistics in general and Cognitive Linguistics in particular (Geeraerts 2010):

1) Macro-level: Investigation of the role of lexico-grammatical patterns in the processes of use, acquisition and change of more abstract constructions and linguistic categories in different languages and language varieties across time, space and discourse modalities;

2) Micro-level: Study of the rich conceptual, cultural, multimodal, interactional and other information associated with the use of lexico-grammatical patterns and representing the multidimensional character of the speaker’s knowledge.
List of participants

1. LUĐIVINE CRIBLE (Université de Louvain, Belgium). From co-occurrence to constructions: patterns of discourse markers and disfluencies across registers in English and French.

2. VOLODYMYR DEKALO (University of Erfurt, Germany). Salient verb collexemes of German modal constructions with the verbs vermögen, verstehen, wissen, bekommen.

3. ANTJE ENDEFELDER QUICK (Leipzig University and Max Planck Institute for Evolutionary Anthropology, Germany), AD BACKUS (Tilburg University, the Netherlands), ELENA LIEVEN (University of Manchester, the UK), and MICHAEL TOMASELLO (Max Planck Institute for Evolutionary Anthropology, Germany). Partially schematic constructions in Code-mixing of a German-English bilingual child.

4. SUSANNE FLACH (Freie Universität Berlin, Germany). Quantifying qualitative change: Collexeme paradigms and progressive constructionalization.

5. EMMELINE GYSSELINCK (University of Ghent, Belgium). Frequency and collocational constraints: the expansion of the Dutch intensifying fake reflexive resultative construction in the 19th-21st Century.

6. BEATE HAMPE (University of Erfurt, Germany) and IRENE MITTELBERG (RWTH Aachen University, Germany). Salient exemplars and syntactic constructions: Verbal and non-verbal usage evidence on prototypical realisations of the caused-motion construction in English.

7. STEFAN HARTMANN (University of Hamburg, Germany) and LAUREN FONTEYN (University of Manchester, the UK). Nominalization patterns in English and German: A contrastive study.


9. NATALIA LEVSHINA (Leipzig University, Germany). The principle of economy and language variation: help + (to) Infinitive in twenty varieties of web-based English.

10. NORIKO MATSUMOTO (Kobe University, Japan). Variations in the Frequency of V-V Sequences in English among 20 Different English-Speaking Countries.

11. DORIS SCHÖNEFELD (Leipzig University, Germany). Understanding novel denominal verbs.

Total length: 12 slots (11 talks plus introduction).
Abstracts

LUDIVINE CRIBLE (Université de Louvain, Belgium)

From co-occurrence to constructions: patterns of discourse markers and disfluencies across registers in English and French

Key-words: discourse markers; disfluencies; constructions; register variation; contrastive

Spoken language is characterized by online processes of production and comprehension happening over time. A natural consequence of this temporal nature is the presence of discourse markers (henceforth DMs, e.g. Schiffrin 1987) that can generally be defined as syntactically optional pragmatic expressions “fulfilling structuring functions with respect to local and global content and structure of discourse” (Fischer 2000: 20) such as well, but or I mean. These procedural devices express a wide range of discourse functions, from connective meanings such as cause or contrast, to interactional meanings such as turn-taking or monitoring. This heterogeneity within the DM category is reflected in preferences of syntactic category and position, as well as register and crosslinguistic variation (Aijmer 2013, Degand 2014).

DMs tend to co-occur with so-called disfluencies (e.g. Shriberg 1994) such as filled pauses, identical repetitions or false-starts, especially in impromptu speech though their presence is not excluded from more prepared and monologic situations. Following valid tertia comparisonis and operational definitions, DMs and disfluencies have been annotated in a comparable corpus of English and French encompassing eight different interaction settings from private conversation to news broadcast ([Author]). After combining information on the function and position of DMs and their co-occurring disfluencies, using statistical models such as conditional inference trees or multiple correspondence analysis, a number of interesting patterns emerged from the data: for instance, additive DMs (e.g. and) are typically utterance-initial and cluster with filled or silent pauses, while monitoring DMs (e.g. you know) are more strongly associated with the final position and disruptive disfluencies such as false-starts or truncated words. Such integration of formal and functional variables suggests tentative conclusions on the relative fluency or disfluency of these patterns.

In this paper, I will present the corpus-based method and major findings regarding the variation of these patterns across languages and registers. In doing so, the aim is to build a case for a constructional approach to patterns of DMs and disfluencies which would allow for their fine-grained investigation in varying degrees of abstraction, zooming in and out from broad categories to specific functions of DM lexemes (e.g. so uhm expressing a conclusive meaning). I argue that the high frequency and meaningful variation of co-occurrence patterns vouch for their treatment as constructions or schemas in usage-based terms, taking these notions slightly away from their lexico-grammatical core to a more discursive or interactional level, in line with recent works (e.g. Fischer & Alm 2013). This paper therefore paves the way for further study of the acquisitional and processing profile of these candidate constructions.

References

[Author]


Salient verb collexemes of German modal constructions with the verbs vermögen, verstehen, wissen, bekommen

Keywords: modal constructions, semantic network, semantic similarity, verb collexemes, salient collexemes

The proposed talk will discuss the results of a network analysis (Ellis, O’Donnell & Römer 2014) that has been carried out (in addition to a cluster analysis) in order to assess and further evaluate the results of a simple collexeme analysis of the lexical verbs occurring in four German modal constructions. These modal constructions contain the verbs vermögen, verstehen, wissen, bekommen and compete with the highly frequent können-construction in written German (ex. 1–4).

(1) Nur in seiner Quantität [...] vermag das Geld Widerstand zu leisten.¹
(2) Menschliche Schwäche unterstellte er und verstand er auszunutzen.
(3) Meine Schwester wußte meinen Einfall zu schätzen.
(4) Und wenn sie die Stöße und Erschütterungen der gesellschaftlichen Bewegungen zu spüren bekennen, [...] .

They express the modal meaning ‘possibility / capability’ and can be formalized by the following general schemas:

(5) [X VERMÖGENFIN (Y) (Z) zu VINF]²
(6) [X VERSTEHENFIN (Y) (Z) zu VINF]
(7) [X WISSENFIN (Y) (Z) zu VINF]
(8) [X BEKOMMENFIN (Y) (Z) zu VINF]

The verbal lexemes that are significantly associated with the VINF-slot in the schemas (5–8), i.e. its so-called collexemes, were defined by a simple collexeme analysis (Stefanowitsch & Gries 2003) and provide the input information for the semantic network analysis focussed on in this talk as well as a cluster analysis. It must be stressed that the collostruction-strength values (and the rank information derived from them) were only used in the identification of the attracted collexemes, but played no further role in either the network analysis or the cluster analysis.

Both of these further analyses were entirely based on the semantic relatedness of the collexemes (verb types), as presented in GermaNet (Hamp & Feldweg 1997), the German equivalent of the widely knownWordNet (Miller et al. 1990; Fellbaum 1998). More specifically, they are based on the results of the pairwise comparisons for semantic similarity of every lexeme with every other one (i.e. a measure of semantic similarity provided by GermaNet, see (Hamp & Feldweg 1997)).

While the cluster analysis was employed to identify verb groups on the basis of semantic similarity, the semantic network analysis makes it possible to arrange the verb collexemes into networks applying a graph-based algorithm from network science (Bastian, Heymann & Jacomy 2009). In these networks, the nodes represent verb collexemes and the edges semantic relations between them, as defined by the similarity measure chosen. The salience of the verb collexemes is quantified by betweenness centrality, which measures the extent to which a node lies on paths between other nodes (Newman 2010: 185), and also visually emphasizes this in the network graph.

The analysis thus determines which collexemes are central given their semantic relation to other collexemes. This information is very different from the output of the cluster analysis over the same semanticrelatedness values. In turn, both kinds of information are different from (and thus complement) the centrality information yielded by the collostruction-strength values themselves, which are based on usage, i.e. actual verb tokens.

We hold that, together, these diverse types of information about the constructions investigated enable the analyst to determine the prototypical or the most salient verb collexemes of the modal

¹ All examples are token from the DWDS-Kernkorpus des 20. Jahrhunderts
² The parentheses indicate optionality.
constructions explored, and thus make more informed claims about potential mid-level schemas that may play a specific role in the formation of the modal constructions investigated.

References
Code-mixing is one of the more salient phenomena that result from bilingualism. Bilingual acquisition data show that if children are mixing, this already occurs early on in the acquisition process. A rich literature exists that seeks to account for mixing patterns through the basic syntactic architecture of language (e.g. Cantone 2007). However, a focus on abstract syntax obscures our view on the acquisition of more concrete pieces of syntax. Bilingual child utterances often show evidence for the productivity of constructions, when open slots in a construction are filled in with lexemes from the other language. In many cases, these constructions come mostly from one of the languages, creating an asymmetry that is often found in adult code-mixing and which is known as the ‘matrix language effect’. We suggest a usage-based approach to this phenomenon, and argue that it gives us a better account of bilingual acquisition than accounts that focus on abstract syntax or UG.

Usage-based approaches (e.g. Tomasello 2003, Bybee 2010) assume that units can vary in their degree of schematicity, ranging from completely lexically fixed lexical items (e.g. How are you?) to wholly schematic constructions (e.g. NP VP NP). In between are partially schematic constructions (e.g. I want X), and these will be shown to play an important role in the code-mixing of a German-English bilingual child aged 2;03 – 3;11. Code-mixing often consists of the use of a partially schematic construction from one language with the open slot filled by material from the other language.

Code-mixed data were coded for schematicity (n=321). Identification of fixed slots was determined by previous occurrence of that specific unit e.g. partially schematic now I’m X was supported by earlier occurrences such as now I’m getting braun, and now I’am kaempfing you. We also analyzed whether any part of the code-mix was primed via prior discourse (Table 1).

Our first analyses of schema types revealed effects of age and of the language of the fixed slot ($\chi^2 (4, N=277)= 65.1439, p<.001$). The child’s stronger language, according to MLU, tends to provide the fixed parts of partially schematic constructions whereas the open slot may be filled by the weaker language, mainly with content lexemes (Figure 1). The strength of this factor shows in an interesting reversal of the language’s roles between the ages 3;00 and 3;10. During the period dominance changes from German to English which can be seen in the reversal of the language contributing the fixed slot. Concerning priming we found that in 72% of all mixed utterances at least one part (fixed or open) was primed; priming seems to equally affect both the fixed and the open parts of a schematic construction. This finding strengthens the need for a processing component in the explanation of code-mixing.

The results allow us to develop a more subtle account of what is generally referred to as a ‘matrix language effect’ in the code-switching literature, showing that this effect is mostly brought about by the selection of the most entrenched, partially schematic, constructions.
Tables

<table>
<thead>
<tr>
<th>Fion discourse, age 2;03</th>
<th>fixed slot</th>
<th>open slot</th>
<th>Discourse primed</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAT:</td>
<td>English</td>
<td>English</td>
<td>Input</td>
</tr>
<tr>
<td>Do you like sharks Fion ?</td>
<td>do you like x</td>
<td>sharks</td>
<td></td>
</tr>
<tr>
<td>CHI:</td>
<td>German</td>
<td>English</td>
<td>lexical</td>
</tr>
<tr>
<td>Da eine shark.</td>
<td>da eine x</td>
<td>shark</td>
<td>output</td>
</tr>
<tr>
<td>‘There, a shark.’</td>
<td>‘there, a X’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Discourse priming

Figures

Figure 1. Language of the fixed slot

References

Press.


Quantifying qualitative change: Collexeme paradigms and progressive constructionalization

The recent history of the English progressive is one of a rapid increase in quantitative usage (frequency). This has led to analyses of the construction as one of grammaticalization and/or ‘obligatorification’, especially in the course of the last two centuries (e.g., Kranich 2010). There is, however, strong collostructional evidence that the abstract schema of the progressive has been relatively stable for much of the Early and Late Modern English period (Flach & Johannsen 2015, work in progress). Among other indicators, there is a conspicuous rise in shared collexemes between the progressive and the simple constructions, which can be taken to represent a non-trivial construction-internal change. This qualitative change happens quicker than—and well before—the rapid rise in frequency (quantitative change), a development that can be shown by combining Distinctive Collostructional Analyses and VNC (e.g., Gries & Hilpert 2008). Linguistically, the results support a view of the progressive as a case of constructional change rather than one of grammaticalization (in the sense of the terminological distinction in Hilpert 2011a).

Methodologically, the value of Distinctive Collostructional Analysis for diachronic data has been shown, evaluated, and refined (Hilpert 2006, 2011b; Stefanowitsch 2006). On the other hand, what emerges generally in using Collostructional Methods, is that studies tend to put more emphasis on the most strongly attracted collexemes. This is warranted given the role of top-ranked items for the entrenchment of schemas. However, there is also a great potential in taking the entire ‘collexeme paradigm’ of (non-)shared types into account in approaching the interplay of constructional variation, alternation, and ultimately change.

Thus, this talk seeks to address two relevant issues, a linguistic and a methodological one: first, how can the nature of—and the change in—the ‘collexeme paradigm’ help us understand the development of the progressive prior to its major quantitative change? How does it support the assumption of constructionalization (rather than grammaticalization)? Second, methodologically, there are potentially serious statistical caveats especially in analysing diachronic data: could increased overlap between progressive and alternative constructions simply be a function of (greater) construction frequency and/or corpus size (e.g., COHA)? Should we treat top-, low-ranked items, and hapaxes alike? These are general challenges for (Diachronic) Collostructional Analyses, thus (re-)sampling methods are proposed and evaluated to minimize the risk of linguistic overinterpretation.

In sum, combining collexemic measurements and clustering methods helps us understand variation and/or alternation that can lead to change, by disentangling qualitative from quantitative change, thus refining our understanding of change mechanisms. Despite the focus on diachronic variation and change, both approach and results allow for generalizations to other areas of variational, acquisitional and/or lectal interests.

References


Recent work in Diachronic Construction Grammar has been concerned with the idea that the constructional network is a dynamic system which is constantly being reorganized as new subschemas emerge or fall out of use (see, e.g., Hilpert 2013, Colleman 2015). In Traugott & Trousdale (2013), it is further argued that the emergence of a new (sub)schema is often concomitant with the loss of collocational restrictions, resulting in expanded use. This paper aims to explore the relationship between frequency and collocational constraints by focusing on the diachronic development of one Dutch construction in particular.

The Dutch intensifying fake reflexive resultative construction is a formally transitive pattern in which a result is predicated of a reflexive object that is not selected by the verb. However, unlike ‘regular’ resultative constructions, the construction does not convey a literal resultative meaning. This is demonstrated in the following example.

(1) Ze lachten zich dood/rot/kapot/een bult/een breuk om die mop.
   ‘They laughed very hard (lit. laughed themselves dead/rotten/a hump/a fracture) at that joke.’

Instead of denoting a state resulting from the activity denoted by the verb, the bolded elements are used as intensifiers, boosting the verbal event of lachen ‘to laugh’ to a higher degree. Speakers of present-day Dutch are seemingly able to fill in just about anything in lieu of dood, rot or kapot, but there are indications that this has not always been the case. First of all, preliminary corpus research based on historical newspaper data shows that the raw frequency of occurrence of the construction has increased over fourfold between the late 19th century and the late 20th century. Secondly, this increase in frequency appears to have gone hand in hand with an overall relaxation of collocation constraints, presenting itself as an explosion of new intensifier types. Whereas language users in the late 19th Century only used a handful of intensifiers, mainly relying on dood ‘dead’ as a means of intensification, there are over 70 different intensifiers attested in present-day newspaper data. To give but one example, the verb zich ergeren ‘to be annoyed’ collocates with dood ‘dead’, kapot ‘broken’, rot ‘rotten’, wild ‘wezenloos’, as well as a broad spectrum of colour terms:

(2) Hij ergerde zich blauw/groen en geel/bont en blauw/rood en groen/groen, geel en blauw/blauw en groen/… aan hun gedrag
   ‘He was very annoyed (lit. annoyed himself blue/green and yellow/black and blue/red and green/green, yellow and blue/blue and green…) by their behaviour.

However, certain intensifiers or verb-intensifier combinations occur much more frequently than others, suggesting that there is also a great deal of conventionalization involved in the use of the construction. Returning to our example, the intensifiers blauw ‘blue’ and groen en geel ‘green and yellow’ account for over 50% of the examples with zich ergeren ‘to be annoyed’.

In order to elucidate this intriguing interplay between creativity and conventionalization, I will use the Delpher corpus of Dutch historical newspapers to investigate which intensifiers have gained (or lost) ground over the last two centuries and keep track of their collocations. In addition, I will discuss what these findings can tell us about the relationship between frequency and collocational restraints.

References:


This paper aims at investigating the influence of salient central exemplars on the usage tendencies of the syntactic construction they instantiate, by (i) following up corpus work on the complex-transitive constructions (Hampe & Schönefeld 2006; Hampe 2011a,b); and by (ii) adding multi-modal evidence from the RedHen database.

For the purpose of this study, we define "salient exemplars" as mid-level constructions, i.e. partially lexically specified constructions whose primary predicators are instantiated by those generic caused-motion verbs that a collostruction analysis determined as most closely associated with the construction (put, place, bring, get, set, take, turn, send, push, and possibly also the force-dynamically related verbs keep, leave, hold). In this work, selected linguistic properties of all of the tokens these collexemes of the caused-motion construction in the ICE-GB are further analysed, including such properties as: tendency towards particular lexical realisations of the obligatory adverbial/predicative, tendency towards motion meanings vs. metaphorical meanings at the first and second level, preference for particular realisations of aspectual and voice categories. These usage characteristics of the most typical instantiations will then be compared to the behaviour of the complete category, including less typical and highly atypical lexemes (i.e. repelled collexemes).

Appreciating the multi-modal nature of spontaneous oral communication, a representative sample of the same verbs are traced in a part of the multi-modal RedHen database providing American TV news coverage, and checked for the occurrence of co-speech gestures. Assuming that the most typical caused-motion verbs instantiate "basic action scenarios" (Mittelberg & Joue, to appear) when occurring literally, co-speech gestures of multi-modal tokens of the construction are expected to 'enact' basic aspects of the particular type of movement characteristic for these scenarios. Of particular interest here is the speakers' behaviour in the case of metaphorical uses of the construction. While first-level metaphor (i.e. non-spatial uses of the particles) has created (strongly conventionalized) aktionsart meanings, second-level metaphor takes the meaning of the construction out of the realm of motion into the domain of causation. In the latter case, we expect gestures to be "bleached", i.e. less specific and more uniform across verb-frames.

Overall, the paper aims at extending existing work on mid-level constructions as sources of analogical syntax motivating less typical/frequent or even completely novel/creative instances (see work on "frame blending", Fauconnier & Turner 1996; Hampe & Schönefeld 2006), of special interest are (i) the possibly category-wide effects of the behaviour of the prototypical instances/most salient exemplars, and (ii) the role of other semiotic channels than the verbal one in the formation of the symbolic category that defines the syntactic construction.

References


Nominalization patterns in English and German: A contrastive study

Keywords:  Word-formation change; collostructional analysis; comparative historical linguistics; diachronic construction grammar

It has often been noted that English nominalizations in *-ing*, such as (1), differ substantially from constructions with cognate suffixes in other Germanic languages, e.g. German nominalizations in *-ung*, exemplified in (2) (cf. e.g. Demske 1999; Iordachioaia & Werner 2014).

(1) Labour will ban the killing of foxes and encourage the killing of human foetuses. (BNC)
(2) Der Mensch in der Rüstung will fechten. (DWDS) ‘The person in the armor wants to fence’

However, to our knowledge, no systematic corpus-based study comparing these two patterns has been conducted so far. In this paper, we argue that collostructional methods provide a highly insightful method for studying these morphological patterns and comparing them with regard to their synchronic function and their diachronic development.

Drawing on several diachronic corpora – the Penn-Helsinki corpora for English and the GerManC corpus as well as the German Text Archive for German – we show that both patterns become increasingly “nominal”. While their word-formation products capture the event-like semantics of their base verbs quite faithfully in the initial stages of their development, *ung-* and *ing-*nouns denoting more concrete entities such as locations and objects become more and more prevalent. This can be shown using diachronic distinctive collexeme analysis (Hilpert 2006; Author2 & Author1 2016) as well as a more fine-grained semantic analysis of the individual types, tokens, and hapax legomena. In addition, morphological cross-tabulation analysis (Author1 2014), which assesses how strongly individual bases are attracted to or repelled by specific word-formation patterns, can be used to track the diachronic changes undergone by the word-formation constructions.

Importantly, the constructionist perspective adopted here sheds new light on the diachronic development of both patterns. In earlier studies, it has been claimed that English *ing*-nominals, unlike German *ung*-nominals, remain fairly verb-like in their semantics (cf. e.g. Demske 2002). However, we argue that this is only true if one takes a morpheme-based view on the data, which lumps together nominal and verbal gerunds. In taking a constructionist schema-based perspective, by contrast, it becomes clear that the development of English nominal gerunds bears striking similarities to the evolution of German *ung*-nominals as sketched by Demske (2000) and Author1 (2016). This shows that applying collostructional methods to comparative data can help reveal previously undetected cross-linguistic similarities in language change.

Corpora

*BNC: British National Corpus, version 3 (BNC XML Edition).* http://www.natcorp.ox.ac.uk/
DTA: Deutsches Textarchiv (German Text Archive). http://deutschestextarchiv.de/


GerManC: The GerManC corpus. http://ota.ox.ac.uk/desc/2544


References


Cross-linguistic aspectual variation and the mental predicate think: The case of English and Polish

Cross-linguistic aspectual variation is easy to observe. It is enough to compare Slavic languages, which make a distinction between grammatical types of aspect such as perfective and imperfective, and Romance languages, which distinguish between perfective and imperfective aspect for French, but progressive and perfect aspect for English, to consider just a few. Another aspectual distinction is made with regard to aspectual classes, also known as Aktionsarten or situation aspects, bearing “on inherent features of the verb” (de Swart 2012: 753). As de Swart (2012: 754) notes, grammatical and situation aspect are different types of aspectual classification; nevertheless, she observes that there are “clear interactions between them”. One should also consider languages in which grammatical aspect is not present, but its semantic value is manifested in one way or another. Thus, revealing “the internal temporal constituency of a situation” (Comrie 1976: 3), aspect can be treated as “a covert semantic category on the sentential (or propositional) level”(Croft 2012).

The aim of this study is to analyse aspectual behaviour of two distinct languages, belonging to two distinct language families and two distinct aspectual classifications, namely English and Polish, on the example of the English mental predicate think and its Polish equivalent myśleć. In order to investigate their internal temporal constituency, Croft’s (2012) two-dimensional model of aspectual structure is used. It combines a time dimension and a qualitative state dimension, which allow to see the verb’s aspectual behaviour depending on the grammatical and discourse context.

In order to investigate the aspectual potential of the mental predicates think and myśleć, a sample of corpus data has been gathered from the British National Corpus and from the PELCRA search engine (Pęzik 2012) of the National Corpus of the Polish Language (Przepiórkowski et al. 2012), respectively. The extracted data have been manually annotated with relevant ‘usage features’ (Glynn 2009). To see feature interaction and observe aspectual ‘behavioral profiles’ (Gries 2006) of the verbs, multivariate statistical methods such as Correspondence Analysis or Hierarchical Cluster Analysis have been applied. In sum, the corpus-driven statistical analysis is used in order to see how the mental process of thinking unfolds in time in the two languages by taking into consideration both the temporal and the qualitative dimension.

References:


The principle of economy and language variation: *help* + (*to*) Infinitive in twenty varieties of web-based English

**Aims and theoretical background**

The aim of this study is to demonstrate that the universal principle of economy (e.g. Haiman 1983) operating at the mid-level of abstraction, can explain constructional variation *help* + (*to*) $V_{inf}$, as exemplified by the sentences (1a) and (1b):

(1) a. Mary helped John write the letter.

b. Mary helped John to write the letter.

In many previous accounts (e.g. Dixon 1991), the choice between the variants was explained semantically, namely, the use of the bare or marked infinitive is assumed to be related to the degree of the subject’s involvement in the event represented by the infinitive. In addition, such factors as cognitive complexity, avoidance of identity (*horror aequi*) and the inflectional form of *help*, have been shown to constrain the choice between the bare or marked infinitive (Lohrmann 2011). In this study, I test the following hypothesis: The stronger the statistical association between *help* and the infinitive, the higher are the chances of the bare infinitive being used, according to the principle of economy (e.g. Haiman 1983). This hypothesis is tested in a series of multivariate analyses of the alternation in twenty national varieties of web-based English.

**Data and method**

I extract semi-automatically a sample of the constructions in question from Davies’ (2013) corpus of Global Web-based English, which represents twenty countries where English is spoken. The total size of the data set is approximately 120K observations, which are annotated semi-automatically for a number of variables. The strength of association between *help* and the infinitive is represented by a subset of popular unidirectional and bidirectional association measures, such as Collostructional Strength, Attraction, Reliance, log-odds ratio, $\Delta P$, etc. The other variables in the models are inspired by the previous studies of the alternation. They include the number of words between *help* and the infinitive (or cognitive complexity, as in Rohdenburg 1996), the presence of *to* before *help* (avoidance of identity, as in Rohdenburg 2003), the inflectional form of *help*, the valency of the infinitive and the degree of formality, which is represented by the average word length in the given online text.

The quantitative analyses involve twenty Bayesian mixed-effects binomial regression models, with the use of the bare or marked infinitive as the response variable, the above-mentioned contextual variables as fixed effects and the infinitives and text IDs as random intercepts. The analyses are performed with *RStan* (cf. Sorensen et al. In press). The preliminary results suggest that the strength of association is an important predictor of the variation in all twenty varieties. This finding is thus perfectly in line with the universal principle of economy.

**References**


Variations in the Frequency of $V$-$V$ Sequences in English among 20 Different English-Speaking Countries

Keywords: corpus, deixis, English dialects, grammar, verb phrase

This paper shows the existence of variations in the frequency of $V$-$V$ sequences in English among 20 different English-speaking countries, relying upon Corpus of Global Web-Based English (GloWbE). This paper deals with three types of $V$-$V$ sequences, the come-$V$, the go-$V$, and the help-$V$ sequences. The three types of $V$-$V$ sequences are discussed in comparison with the bare-come/go-and-$V$ and the help-to-$V$ sequences, which are in many ways semantically similar to the come/go/help-$V$ sequences, as shown in (1)-(3).

(1) a. Come join us.  
   b. Come and join us.
(2) a. You can go buy food somewhere else.  
   b. You can go and buy food somewhere else.
(3) a. He helped organize the party.  
   b. He helped to organize the party.

From a syntactic point of view, (1) and (2) are completely different from (3). Despite including two verbs, each sentence in (1) and (2) consists of one verb phrase. Each sentence in (3) involves two verb phrases. Based on Corpus of Historical American English (COHA), the come/go/help-$V$ sequences in American English have been recently gaining in currency in that they are replacing the come/go-$V$ and the help-to-$V$ sequences, respectively, as shown in Figure (1). This move is not related to grammaticalization or auxialization. This paper investigates whether this move applies to other varieties of English spoken around the world. With respect to the varieties of English, this paper also demonstrates that there is decidedly more to differences in grammar than well-known differences in pronunciation and vocabulary.

There are three main findings from our corpus data based on GloWbE, as shown in Table (1) and Table (2). First, the go-$V$ sequence that outnumbers the bare-go-and-$V$ sequence is observed in all 20 countries. Second, the help-$V$ sequence that outnumbers the help-to-$V$ sequence is observed in 19 countries except Jamaica. Third, the come-$V$ sequence that outnumbers the bare-come-and-$V$ sequence is observed in only 5 countries, America, Canada, Jamaica, Nigeria, and Singapore.

From these findings, as shown in Table (1), it can be concluded that USA and Great Britain represent two extremes of grammatical continuum with respect to $V$-$V$ sequences, with USA at the progressive pole and Great Britain at the conservative pole. America has a growing tendency to select all three types of $V$-$V$ sequences, but Great Britain does not have such a tendency. The national varieties of 7 countries except for USA and Great Britain in Table (1) are located roughly between the two extremes in relevant respects. The national varieties of 11 countries in Table (2) are, in some sense, located outside the grammatical continuum in relevant respects. These national varieties exhibit their own idiosyncratic grammatical divergences. For example, Jamaica and Singapore have a tendency to select come/go-$V$ sequences involving one verb phrase, whereas Australia and New Zealand have a tendency to select help-$V$ sequences involving two verb phrases. Our corpus-based study discussed here is to a large extent supported by using different corpora, such as British National Corpus (BNC), ICE (International Corpus of English) Corpora, and Collins Wordbanks Online.

Figure (1). Frequency of use in COHA per million words from 1810 to 2009.
<table>
<thead>
<tr>
<th>Ratio of go-V to bare-go-and-V is X:1</th>
<th>USA</th>
<th>Canada</th>
<th>Kenya</th>
<th>Malaysia</th>
<th>South Africa</th>
<th>Ireland</th>
<th>Tanzania</th>
<th>Great Britain</th>
<th>Sri Lanka</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.42 :1</td>
<td>6.44 :1</td>
<td>5.21 :1</td>
<td>2.77 :1</td>
<td>2.09 :1</td>
<td>2.77 :1</td>
<td>2.21 :1</td>
<td>1.85 :1</td>
<td>1.43 :1</td>
<td></td>
</tr>
<tr>
<td>Ratio of come-V to bare-come-and-V is X:1</td>
<td>1.51 :1</td>
<td>1.26 :1</td>
<td>0.75 :1</td>
<td>0.85 :1</td>
<td>0.76 :1</td>
<td>0.63 :1</td>
<td>0.51 :1</td>
<td>0.42 :1</td>
<td></td>
</tr>
<tr>
<td>2.88 :1</td>
<td>2.45 :1</td>
<td>1.99 :1</td>
<td>1.58 :1</td>
<td>1.45 :1</td>
<td>1.63 :1</td>
<td>1.51 :1</td>
<td>1.49 :1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table (1). Ratio of come/go/help-V sequences to bare-come/go-and-V and help-to-V sequences based on frequency of use in GloWbE per million words

<table>
<thead>
<tr>
<th>Ratio of help-V to help-to-V is X:1</th>
<th>Jamaica</th>
<th>New Zealand</th>
<th>Australia</th>
<th>Singapore</th>
<th>Hong Kong</th>
<th>Philippines</th>
<th>Nigeria</th>
<th>Ghana</th>
<th>Bangladesh</th>
<th>India</th>
<th>Pakistan</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.36 :1</td>
<td>2.21 :1</td>
<td>2.04 :1</td>
<td>3.37 :1</td>
<td>3.37 :1</td>
<td>2.09 :1</td>
<td>2.38 :1</td>
<td>2.29 :1</td>
<td>2.15 :1</td>
<td>1.67 :1</td>
<td>1.37 :1</td>
<td></td>
</tr>
<tr>
<td>Ratio of help-V to help-to-V is X:1</td>
<td>1.21 :1</td>
<td>0.45 :1</td>
<td>0.54 :1</td>
<td>1.52 :1</td>
<td>0.73 :1</td>
<td>0.72 :1</td>
<td>0.94 :1</td>
<td>0.97 :1</td>
<td>0.91 :1</td>
<td>1.12 :1</td>
<td></td>
</tr>
<tr>
<td>0.96 :1</td>
<td>2.00 :1</td>
<td>1.97 :1</td>
<td>1.57 :1</td>
<td>1.87 :1</td>
<td>2.83 :1</td>
<td>1.5 :1</td>
<td>2.07 :1</td>
<td>1.82 :1</td>
<td>1.8 :1</td>
<td>2.06 :1</td>
<td></td>
</tr>
</tbody>
</table>

Table (2). Ratio of come/go/help-V sequences to bare-come/go-and-V and help-to-V sequences based on frequency of use in GloWbE per million words
DORIS SCHÖNEFELD (Leipzig University, Germany)

Understanding novel denominal verbs

The talk was initiated by a layman’s discussion of novel verbs (https://www.1843magazine.com/content/ideas/anthony-gardner/youve-been-verbed). The question I am asking is how people can make sense of these new words when they encounter them in communication. The question will be pursued on the basis of usage data of English, as found in such corpora as COCA and the BNC.

The argument is developed within a (usage-based) construction-grammar framework and aims at an analysis of the cues speakers give their addressees when they use novel denominal verbs, such as card, pit or action. The problem of decoding these new words is mainly due to the fact that these novel verbs are (not yet) associated with the rich frame-semantic knowledge language users have accumulated for the verbs ‘established’ in a speech community. To understand such verbs, the decoder will probably rely on all the cues s/he can get from the situational and linguistic context. My exploratory investigation focuses on the latter, searching for triggers that help listeners to identify the event the novel verb is meant to encode. On the basis of the hits extracted from the corpora, it can be shown that cues to the meaning are available at different levels of schematicity.

Firstly, there is information in the form of the (fully schematic) argument structure construction (ASC) in which the novel verb is used. Some of these are, however, less helpful for the decoding process, as they have very general functions/meanings associated with them, such as the (mono-)transitive ASC (X ACT ON Y) or the intransitive ASC (where something is predicated of X – the subject). Such knowledge may suffice when the situation of the utterance is clear enough for the hearer to see what is meant.

Secondly, there is the cue given by the denominal verb itself. Its nominal base makes available the world knowledge surrounding its meaning, and the referent must be assumed to play a salient role as a participant in the event to be named (cf Clark & Clark, 1979: 787, Kiparsky 1997: 482). It is here that variational aspects come into play, namely, when nominal concepts have domain-specific readings and participate in domain-specific events, such as text, inbox, friend in electronic communication. Decoding the corresponding verbs presupposes fairly elaborate knowledge of the specifics of this field. And yet, the cue of the nominal base for an event meaning may not be strong enough, so that the novel verb’s meaning can be narrowed down rather than determined. Besides tracing the spread of such specific usages into ‘general English’, I will also look at the internal structure of the base nouns and how it affects the derived verbs’ decodability.

Thirdly, the novel verbs’ meanings can be further constrained by their contextual embedding at the fully lexical level. In particular, collocations have been found to trigger analogical verbs, as eg cars/smartphones/cameras/bookmaker system be (showcased) at a fair/exhibition -> shown, exhibited, introduced, and they (as well as individual words) may allude to specific events with which they are typically associated, as eg I was carded and denied admission (check the ID) vs the ref was asked why the player was carded (exclude from game).

As a last point, the study will ask what can be read off usage data with respect to the most effective cues for understanding the novel verbs investigated.

References

