

Milla Luodonpää-Manni, Esa Penttilä & Johanna Viimaranta. 2017. *Empirical approaches to cognitive linguistics: Analyzing real-life data*. Cambridge: Cambridge Scholars Publishing. Pp. 350.

Reviewed by Edyta Jurkiewicz-Rohrbacher

1 Introduction

The book is the outcome of two workshops on cognition, grammar, meaning, usage and corpora organized in 2014 and 2015 by the Finnish Cognitive Linguistics Association. It comprises nine papers covering diverse research areas (grammatical constructions and concepts, translation strategies, semantics, language acquisition) based on material from five European languages: English, Finnish, German, Russian and Spanish.

The papers are preceded by a short introduction written by the editors, which describes recent developments in empirical methods within cognitive linguistics. In a sense, it serves the purpose of a complementary view, which helps the reader to associate each of the following papers with an appropriate methodological frame. As it becomes apparent, usage-based approaches to language studies are quite heterogeneous both when it comes to the method of obtaining data and to further analysis. This is also reflected in the current volume, as should become clear from the following summary of the papers.

2 Summary of the chapters

Chapter 1 by Tuomas Huumo, Aki-Juhani Kyrölänen, Jenna Kanerva, M. Juhani Luotolahti, Tapio Salakoski, Filip Ginter & Veronika Lappala, *Distributional semantics of partitive A argument constructions*, examines Finnish clauses containing the partitive agentive argument in two-place predications as in (1):

- (1) *Huolto tunnusti, että useita uudehkoja*
 maintenance admit.PST that several.PL.PAR newish.PL.PAR
- autoja on reuttanut tämän testin.*
 car.PL.PAR AUX.PRES.3SG fail.PTCP this.GEN test.GEN
- ‘Maintenance admitted that several fairly new cars have failed the test.’

The first methodological problem to arise is obtaining the data itself, as the construction in question is a rare phenomenon. The authors thus turn to a web corpus, the Finnish Internet Parsebank, in order to retrieve observations. A manual revision of the concordance reveals 78 occurrences of the construction in the corpus with 41 distinct verbs. Basing on the obtained sample, the role of animacy and semantic roles of agentive arguments, quantifiers, epistemic usages, and the behaviour of patient-like arguments are examined.

The second part of the analysis focuses on the verbs themselves, and complex quantitative methods are employed. Namely, the semantic similarity between verbs appearing in the Partitive A construction is examined using a model based on the *word2vec neural network* trained to reconstruct linguistic contexts of words. A strong positive correlation between the frequency of use and semantic similarity between verbs is discovered. Thus, the authors conclude that these two measures form a gravitation centre fulfilling the function of attractor.

In Chapter 2, *Changes in figure-ground alignment in translation: Condensing information in subtitling*, Jukka Mäkisalo & Marjatta Lehtinen expand on the Cognitive Retention Hypothesis formulated in their former studies as their starting point. The authors assume that a study of a translation from source to target text should distinguish two levels: linguistic and cognitive, because while the latter is mostly preserved, the former is subject to syntactic and lexical change.

In the paper, Mäkisalo & Lehtinen examine this claim with regard to the cognitive Figure-Ground phenomenon – the unknown spatial or temporal property of an object called *Figure* is defined with reference to an entity with known properties (called *Ground*). They study English film dialogues and their Finnish audio-visual translations, and quantitatively evaluate the differences in Figure-Ground retention and lexical-semantic retention in target text. While cognitive structure is only slightly changed, differences in linguistic structure are significant and greater than in the case of the former pilot study on informative prose.

Chapter 3 *How light can a verb be? Predication patterns in V + NP constructions in English, Finnish, German and Russian* by Marja Nenonen, Juha Malli, Alexandre Nikolaev & Esa Penttilä deals with the problem of meaning distribution in V + NP predicates where the verbal part carries barely any descriptive content, as in *to make a face*. Such *light verb* constructions have previously been studied in Germanic languages. Here the authors also contribute to the question how to identify light verbs in Finnish and Russian.

As a starting point, they assume the existence of idiom-prone verbs like *to take*, *to make*, *to give* in all the studied languages, and they identify the most common ones in dictionaries. The authors determine that the distributions of verbal lexemes in V + NP constructions are similar across languages. Moreover, the sets of the most common lexemes include verbs with similar semantic explications. The authors examine more thoroughly the contributions of the meanings in V + NP constructions where the nouns refer to body parts. They come to the conclusion that while there are certain similarities between verbs with regard to their phonological, morphological and semantic properties, the semantic contribution to the meaning of the constructions varies. Therefore, in the last presented stage verbs reflecting the meaning of ‘take’ are examined in the corpora of fiction and newspapers. The meaning of the construction *take* + NP in each retrieved example is evaluated on a three-step scale with regard to the lightness of the verbal contribution. The distributions are fairly similar in all languages and allow the authors to create a *lightness continuum*. On the one end the verbal semantics strongly participates in the compositional meaning of the whole construction: the verb is “heavy”, as in *to take someone to the hospital*. Meanwhile on the other end the main semantic contribution belongs to the NP and the whole construction can be replaced by a separate verb of a similar meaning without loss of information. For example *to take a vote* is a light verb construction since it is interchangeable with *to vote*.

In Chapter 4, *A usage-based and contextual approach to clausal aspect in Finnish*, Salla Nurminen touches upon some rather understudied factors related to the Finnish clausal aspect – a semantic category modelled by the notions of duration and change (or lack of it). A clause is aspectually bounded when a limit indicating the completedness of situation is present:

- (2) *The children came to the park.*

The aspectually unbounded clause is related to the notion of continuity – proceeding without interruption or without reaching the limit:

(3) *The shy animal slowly came closer.*

Nurminen focuses on three frequent verbs: *olla* ‘to be’, *tehdä* ‘to do’ and *tulla* ‘to come’.

In the first part of the paper Nurminen discusses the role of temporal, durative and iterative adverbials such as *maanantaina* ‘on Monday’, *tunnin* ‘for an hour’, *kahteen kertaan* ‘twice’ as limit markers. Basing on an impressive number of nearly 2400 observations obtained from the digital version of Finnish daily *Helsingin Sanomat* she concludes that the role of expressions of repetition and duration does not appear to be as significant as formerly claimed by scholars since it is rather infrequent in the sample. Conversely, she considers adverbials expressing time to be potential triggers of bounds or bounded temporal windows for otherwise unbounded clauses. However, it must be underlined that the study is introspective, not quantitative.

In the second part of the paper, Nurminen discusses boundedness in the context of negation, modal constructions and generic context. She asserts that the previous layered models of the Finnish clausal aspect neglected those properties, and suggests that situations in clauses characterized by one of those three factors can be considered in terms of aspectuality to be unbounded through one of additional layers.

Finally, the author turns to the contextual perspective on clausal aspect, that is, on the interaction between the clause and its neighbour clauses in a text. In the paper mainly clause-external elements such as subordinated clauses containing the connector *kun* and *kunnes* ‘until’ are discussed, but the need for more detailed studies is expressed, as the study sample turned out to be insufficient for research on that topic.

Chapter 5, *The natural translation of idiomatic constructions*, by Esa Penttilä discusses concealed language contact between English and Finnish using the example of instant word-for-word translations of idioms. Obtaining the data is particularly challenging in this case, as the parameters of translated constructions are vague. The author turns to the Internet and applies Google search queries for English idiomatic phrases that do not have accurate equivalents in Finnish. Two primary areas give possible prospects for such studies: social media and journalism. In the case of fora and similar channels word-for-word translations are used and expected to

be understood by the narrow group of readers; similarly, in the case of the press the multinational, English-centred context increases the chance of code-switching. Penttilä manages to identify several idioms (e.g. *rampa ankka* ‘lit. lame duck, an authority in power whose successor has already been elected’, *apina selässä* ‘lit. monkey on one’s back, a burden’) that have infiltrated Finnish from English, but concludes that the phenomenon requires further study.

Chapter 6, *The choice between generic scientific terms in linguistic research written in Finnish*, by Milla Luodonpää-Manni describes the results of a qualitative corpus study on differences in usage of terms generally assumed to be self-explanatory that are associated with the scientific register, such as ‘theory’, ‘hypothesis’ and ‘model’. Analysis of a corpus comprising sixty scientific papers reveals various usages of the same term. Luodonpää-Manni examines possible contributing factors, namely: degree of certainty and stability as to the scientific constructs discussed by scholars, the scope of the scientific constructs, degree of abstraction of the constructs, and definition of scientific constructs in broader terms. An additional factor related to the onomasiological salience is the question of style and convention. The usage-based data is supplemented by questionnaire answers, where scholars were directly asked to characterize the differences between generic terms. This triangulation of methods enables the author to draw conclusions without relying on own intuition.

In Chapter 7, *Topic-marking prepositions in Spanish: contrasting corpus and questionnaire data in the analysis of prepositional synonymy*, Anton Granvik discusses four Spanish prepositions (*de*, *en torno a*, *sobre* and *acerca de*) used for topic marking, which are equivalent to the English group of prepositions: of, on, about, regarding and concerning. Granvik starts from the No-Synonymy Principle, and first characterizes the four prepositions as to their semantic structure and number of meanings they comprise besides the topic-marking function. In the next stage, Granvik extracts from the Corpus del Español 100 clause samples where the prepositions in question mark the topic. Granvik observes that the share of the 100 clauses in comparison to the sample from which they were obtained correlates directly with amount of all meanings each preposition covers. Basing on this distribution he tries to estimate the total number of topic-marking function occurrences in the corpus. The data set is then analysed quantitatively using statistical modelling (multinomial logistic

regression) and the final results are visualized with a dendrogram constructed by performing hierarchical cluster analysis.

The significant factors obtained from the quantitative corpus study allow the author to identify prototypical usage contexts for the prepositions in question and test them further in an open-ended and corpus-testing questionnaire. Granvik concludes that while *de* and *en torno a* form their own distinct behavioural profiles, the remaining *sobre* and *acerca de* are less constrained and more general. The results, however, do not provide a straightforward answer as to the constructional character of topic-marking prepositions.

In Chapter 8, *The role of morphological verb constructions in processing Russian reflexive verbs*, Aki-Juhani Kyröläinen, Vincent Porretta & Juhani Järvikivi examine real-time processing of Russian morphologically complex verbs with regard to the role of frequency and semantic similarity. Information on the frequency and dispersion of four-morphemic reflexive verbs is acquired from the Russian National Corpus in the form of nine variables describing verbal roots, bases, suffixes, reflexive roots and roots' verbiness. The possibly correlated variables are converted using a statistical method called *principle component analysis* into three linearly uncorrelated variables, which capture 96% of the variance. The authors describe an experiment in which twenty-six Russian native speakers were asked to perform a lexical decision task where they were expected to classify the stimulus appearing on the screen as a real Russian word or a pseudoword. The overall set of the potential stimuli consisted of 80 existing reflexive verbs, 80 existing prefixed verbs, 160 pseudowords and 10 words used as a warm-up before starting the experiment. During the experiment the response reaction time was measured. Additionally, having fulfilled the first task, participants were asked to evaluate reflexive – non-reflexive, and prefixed – non-prefixed verb pairs as to their semantic similarity on a five-level scale. In the analysis, the relations between reaction time and frequencies obtained from the corpus are tested.

The results of the psycholinguistic experiment and the principle component analysis are compared using a statistical model: *a generalized additive mixed model*. The results show that the reaction time is negatively correlated with the frequency of a particular morphological construction, as well as with the relatively high frequency of a verbal root. A relatively high frequency of the base from which a reflexive verb was derived, on the other hand, has an opposite effect on reaction time. These three factors correspond directly to the three components obtained from the corpus

analysis, so thanks to triangulation of methods the results seem robust. They are also in line with previous research on complex word processing, but in this case they contribute to the knowledge about processing of morphological structure.

Chapter 9, *Development of early directives in Finnish: A usage-based approach*, by Maija Surakka is a case study of a 2.5–3 year-old Finnish-speaking child's acquisition of directive expressions. The author uses two methods of obtaining data: recordings and notes, but her analysis is multimodal: that is, she also considers action and shared knowledge in the family to be important factors. The study provides a general numeric comparison of five types of grammaticalized meaning structures, but otherwise it is introspective. The author presents the studied utterance together with its paraphrase and then elaborates on them, taking into account both the verbal and nonverbal meaning resources. The author observes a considerable change in the child's development, in particular in an increasing usage of imperatives. Surakka also succeeds in identifying in the child's speech the presence of mental concepts of *wanting* and *others* that need to be established prior to the acquisition of such expressions.

3 Conclusions

One matter worth drawing attention to is the editors' choice to divide the book into *Part one: Corpus-based studies* and *Part two: Research based on information from language users*. These headings may be considered slightly misleading for potential readers, and it seems appropriate to clarify that matter here. First, as may be seen above in the chapter summaries, most studies in part two apply triangulation of data and methods, hence both corpus and elicited data are obtained. Secondly, the editors use the term *corpus* mainly in reference to collections of a defined size and content, often considered representative for some type of language variety, e.g. national corpora, reference corpora. It should not be forgotten that a text sample representative with regards to the research hypothesis may be called a corpus too. In that sense also the collection of child utterances from Surakka's paper can be considered a corpus documenting the development of directivity markers in the child's language.

On the whole, the papers in the volume provide extensive insight as to the range of empirical approaches currently applied in cognitive linguistics: next to corpus-illustrated, introspective studies, research supported by basic statistical testing and complex analytical models is presented. Also very

visible is the increasing orientation towards multimethodological approaches in contemporary cognitive linguistics.

Contact information:

Edyta Jurkiewicz-Rohrbacher

Department of Languages

P.O. Box 24

00014 University of Helsinki

Finland

e-mail: edyta(dot)jurkiewicz(at)helsinki(dot)fi

edyta(dot)jurkiewicz-rohrbacher(at)ur(dot)de