



# **8<sup>th</sup> Symposium on Corpus Approaches to Lexicogrammar (LxGr2023)**

6-8 July 2023

<http://ehu.ac.uk/lxgr>

## **ARCHIVE**

**FOCUS AND FORMAT**

**COMMITTEE**

**PROGRAMME & LINKS TO SLIDES**

**ABSTRACTS**

## SYMPOSIUM FOCUS AND FORMAT

The focus of LxGr is the interaction of lexis and grammar. It is influenced by Halliday's view of lexis and grammar as "complementary perspectives" (1991: 32), and his conception of the two as notional ends of a continuum (*lexicogrammar*), in that "if you interrogate the system grammatically you will get grammar-like answers and if you interrogate it lexically you get lexis-like answers" (1992: 64).

LxGr primarily welcomes papers reporting on corpus-based research on any aspect of the interaction of lexis and grammar — particularly studies that interrogate the system lexicogrammatically to get lexicogrammatical answers. However, position papers discussing theoretical or methodological issues are also welcome, as long as they are relevant to both lexicogrammar and corpus linguistics.

More specifically, presentations can:

- focus more on the lexis or grammar end of the continuum, or adopt an integrative approach.
- discuss different interpretations of the nature of lexicogrammar.
- operate within any theoretical approach that takes into account the interaction of lexis and grammar such as Construction Grammar, Lexical Grammar, Pattern Grammar, Systemic Functional Grammar, Valency Grammar.
- discuss empirical findings in need of theoretical interpretation.
- adopt a synchronic or diachronic approach.
- examine any language, or compare different languages.
- examine L1 and/or L2 use.
- discuss the implications of the findings of corpus-based lexicogrammatical research for applied linguistics (for example forensic linguistics, lexicography, language acquisition, language processing, language teaching, language testing and assessment, translation, sociolinguistics, discourse studies).
- report on the development of relevant research resources or applications (for example language teaching, translation, critical discourse studies).

Full papers are allocated a total of 35 minutes (including 10 minutes for discussion).

Work-in-progress reports are allocated a total of 20 minutes (including 5 minutes for discussion).

There are no parallel sessions.

Participation is free.

## ORGANISER

[Costas Gabrielatos](#) (Edge Hill University)

## LxGr2023 PROGRAMME COMMITTEE

[Federica Barbieri](#) (University of Swansea, UK)

[Belen Díez-Bedmar](#) (University of Jaén, Spain)

[Tine Breban](#) (University of Manchester, UK)

[Thomi Dalpanagioti](#) (Aristotle University of Thessaloniki, Greece)

[Eva Duran Eppler](#) (University of Roehampton, UK)

[Lise Fontaine](#) (Cardiff University, UK)

[Gaëtanelle Gilquin](#) (Université catholique de Louvain, Belgium)

[Nick Groom](#) (University of Birmingham, UK)

[Glenn Hadikin](#) (University of Portsmouth, UK)

[Andrew Hardie](#) (Lancaster University, UK)

[Sebastian Hoffmann](#) (University of Trier, Germany)

[Andrew Kehoe](#) (Birmingham City University, UK)

[Stefania Maci](#) (University of Bergamo, Italy)

[Marina Mattheoudakis](#) (Aristotle University of Thessaloniki, Greece)

[Gabriel Ozon](#) (De Montfort University, UK)

[Michael Pace-Sigge](#) (University of East Finland)

[Pascual Perez-Paredes](#) (University of Murcia, Spain)

[Paul Rayson](#) (Lancaster University, UK)

[Ute Römer-Barron](#) (Georgia State University, USA)

[Benet Vincent](#) (Coventry University, UK)

[John Williams](#) (University of Portsmouth, UK)

[Stefanie Wulff](#) (University of Florida, USA)

## **LxGr2023 PROGRAMME**

### **(AND LINKS TO SLIDES)**

## DAY 1: Thursday 6 July

09:00 – 09:30 WELCOME & ANNOUNCEMENTS

09:30 – 10:05

**Jonathan DUNN** (University of Canterbury Christchurch, New Zealand)

*Productivity as abstractness in computational construction grammar* [\[SLIDES\]](#)

10:15 – 10:50

~~Suzy PARK (Yonsei University, South Korea)~~

~~*A lexicogrammatical approach to analyzing the Korean MWE kes kathta as a pragmatic marker*~~  
~~[CANCELLED]~~

10:50 – 11:10 BREAK

11:10 – 11:45

**Detong XIA** (Southeast University, China)

*Exploring phrase frames across rhetorical functions in workplace request emails* [\[SLIDES\]](#)

11:55 – 12:30

**Mauro LE DONNE** (University for Foreigners of Perugia, Italy)

*Scraping the bottom of the corpus: Exploring lexical emergence in Italian time-stamped corpora*  
[\[SLIDES\]](#)

12:30 – 12:50 BREAK

12:50 – 13:25

**Frane MALENICA** (University of Zadar, Croatia)

*Schematicity and productivity of synthetic compounds in English and Croatian* [\[SLIDES\]](#)

13:25 – 14:30 BREAK

14:30 – 15:40 INVITED SPEAKER

**Thomas HERBST** (Friedrich-Alexander Universität, Germany)

*How corpus linguistics inevitably leads to Construction Grammar: On the interrelatedness of lexis and grammar* [\[SLIDES\]](#)

15:40 – 16:00 BREAK

16:00 – 16:35

**Heesun CHANG** (Texas A&M University, USA)

*Corpus-based analysis of lexico-grammatical features of international teaching assistants*

16:45 – 17:25

~~Xiaolong LU (University of Arizona, USA)~~

~~*Quantify the “big mess” construction in Chinese: A construction grammar approach*~~  
~~[CANCELLED]~~

## DAY 2: Friday 7 July

09:00 – 09:30 WELCOME & ANNOUNCEMENTS

09:30 – 10:05

**Dongchen YAO** (University of Queensland, Australia)

*Factors influencing the errors of number marking in English nouns: A study of Chinese learners of English* [\[SLIDES\]](#)

10:15 – 10:50

**Shuyi Amelia SUN & Kevin JIANG** (Jilin University, China)

*Acknowledging limitations in PhD theses across disciplines: A phrase-frame approach*

10:50 – 11:10 BREAK

11:10 – 11:45

**Susanne HANDL** (LMU München, Germany)

*Verbal binomials: A gradual phenomenon between coordination, construction and lexical unit* [\[SLIDES\]](#)

11:55 – 12:30

**Jiqiang LU & Caroline GENTENS** (Katholieke Universiteit Leuven, Belgium)

*Different speaker-related uses of “I think”: Disentangling pure subjective and epistemic modal uses* [\[SLIDES\]](#)

12:30 – 12:50 BREAK

12:50 – 13:25

**Seval ÖZEN** (Germany)

*Deviation patterns of multi-word verbs in the writings of Turkish speaking EFL learners* [\[SLIDES\]](#)

13:25 – 14:30 BREAK

14:30 – 15:40 INVITED SPEAKER

**Susan HUNSTON** (University of Birmingham, UK)

*Constructions and their Networks: Using system networks to derive constructions from grammar patterns* [\[SLIDES\]](#)

15:40 – 16:00 BREAK

16:00 – 16:35

**Ni LI** (Ocean University of China, Idaho State University, USA)

*Collocation processing in Chinese EFL learners with a wider range of proficiency-based on the L2-textbook frequency* [\[SLIDES\]](#)

16:45 – 17:25

**Victoria FENDEL** (University of Oxford, UK)

*Giving gifts and doing favours: support-verb constructions with χάρις in classical literary Attic*

## DAY 3: Saturday 8 July

09:00 – 09:30 WELCOME & ANNOUNCEMENTS

09:30 – 10:05

**Margo VAN POUCKE** (Macquarie University, Australia)

*Engagement in (pseudo-)medical menopause discourse on YouTube*

10:15 – 10:50

~~Chenghui WU (University of International Business and Economics, China)~~

~~*The semantics of credibility concepts and a lexicographic development of corporate online credibility dictionary (COCD) for corpus-based text analysis*~~

~~[CANCELLED]~~

10:50 – 11:10 BREAK

11:10 – 11:45

**Ling (Kathy) LIN & Yang ZHANG** (Shanghai Jiao Tong University, China)

*Part-of-speech patterns in research articles: A cross-sectional analysis*

11:55 – 12:30

**Marin KEŽIĆ** (University of Zagreb, Croatia)

*When was the last time your construction's meaning went undetermined? On the indeterminate aspectual behavior of GO un-participle constructions*

12:30 – 12:50 BREAK

12:50 – 13:25

**Quentin FELTGEN** (Ghent University, Belgium)

*The diachronic shaping of constructional meaning: A frequency trajectory cluster-based method to explore the emergence of polysemy* [[SLIDES](#)]

13:25 – 14:30 BREAK

14:30 – 15:40 INVITED SPEAKER

**Gaëtanelle GILQUIN** (Université catholique de Louvain, Belgium)

*Construction grammar and lexico-grammar, and why they matter to each other*

15:40 – 16:00 BREAK

16:00 – 16:35

**Lise FONTAINE** (Université de Québec à Trois-Rivières, Canada)

*The ragged middle: A functional approach to lexis* [[SLIDES](#)]

16:45 – 17:00 CONCLUDING REMARKS

## **ABSTRACTS: INVITED SPEAKERS**

## How corpus linguistics inevitably leads to Construction Grammar: On the interrelatedness of lexis and grammar

Thomas Herbst

Friedrich-Alexander Universität, Germany

[thomas.herbst@fau.de](mailto:thomas.herbst@fau.de)

There can be no doubt that computational corpus linguistics has led to a rather dramatic change in the way that many linguists perceive the nature of language as such. Apart from the fact that the widespread availability of data of language use at a previously unprecedented scale puts lexicographers, grammarians, language teachers and textbook authors in a position to arrive at much more accurate descriptions, corpus linguistics has also had an enormous impact on theoretical linguistics. This, I would like to argue, is not so much because of the "discovery" of new phenomena, but corpus analysis has enabled us to better understand the character of the elements that make up language, and in particular the interplay of lexical and grammatical units.

It is interesting to see that in the year 2008 one of the most renowned British corpus linguists, John Sinclair, and the founder of Cognitive Grammar, Ronald Langacker, highlighted the importance of prefabricated multi-word units and rejected the idea of a clear dividing line between lexis and grammar in almost identical terms. This talk will outline these developments and show how corpus linguistic findings play an instrumental role in a lot of usage-based research and in particular in the identification and description of different types of constructions (in the special sense in which the term is used in Construction Grammar, e.g. by Fillmore, Kay & O'Connor 1988; Goldberg 2006, 2019; Gilquin 2010; Boas, Lyngfelt & Torrent 2019; Herbst 2020; Hilpert 2020; Goldberg & Herbst 2021; Hoffmann 2022; Herbst & Hoffmann forthc.).

### References

- Boas, Hans C., Benjamin Lyngfelt, & Tiago Timponi Torrent. 2019. Framing *constructicography*. *Lexicographica*, 35, 15–59.
- Fillmore, Charles J., Paul Kay, & Mary Catherine O'Connor. 1988. Regularity and Idiomatcity in Grammatical Constructions: The Case of *let alone*. *Language* 64. 501–538.
- Gilquin, Gaëtanelle. 2010. *Corpus, Cognition and Causative Constructions*. Amsterdam: John Benjamins.
- Goldberg, Adele E. 2006. *Constructions at Work*. Oxford & New York: Oxford University Press.
- Goldberg, Adele E. 2019. *Explain me this*. Princeton: Princeton University Press.
- Goldberg, Adele E. & Thomas Herbst. 2021. The nice-of-you construction and its fragments. *Linguistics*. <https://doi.org/doi.org/10.1515/ling-2020-0274>.
- Herbst, Thomas. 2020. Constructions, generalizations, and the unpredictability of language Moving towards colloconstruction grammar. *Constructions and Frames*, 12(1), (SI *Construction Grammar across Borders*), 56–96. <https://doi.org/10.1075/cf.00035.her>.
- Herbst, Thomas & Thomas Hoffmann. forthcoming 2024. *A Construction Grammar of the English Language: CASA – a constructionist approach to syntactic analysis*. Amsterdam & Philadelphia: John Benjamins.

- Hilpert, Martin. 2020. Constructional Approaches. In Bas Aarts, Jill Bowie, & Gergana Popova (Eds.), *The Oxford Handbook of English Grammar*, pp. 106–123. Oxford & New York: Oxford University Press.
- Hoffmann, Thomas. 2022. *Construction Grammar: The Structure of English*. Cambridge: Cambridge University Press.
- Langacker, Ronald W. 2008. The relevance of Cognitive Grammar for language pedagogy. In Sabine de Knop & Teun Rycker (Eds.), *Cognitive Approaches to Pedagogical Grammar*, pp. 7–35. Berlin: de Gruyter.
- Sinclair, John McH. 2008. The phrase, the whole phrase and nothing but the phrase. In Sylviane Granger & Fanny Meunier (eds.), *Phraseology. An interdisciplinary perspective*, pp. 407–410. Amsterdam & Philadelphia: John Benjamins.

## **Constructions and their Networks: Using system networks to derive constructions from grammar patterns**

Susan Hunston

University of Birmingham, UK

[s.e.hunston@bham.ac.uk](mailto:s.e.hunston@bham.ac.uk)

This paper reports the first phase of a project to align Pattern Grammar with Construction Grammar and with Systemic Functional Grammar (Hunston 2022). The aim of this phase is to derive sets of constructions from each of 60 verb-based grammar patterns (Hunston & Francis 2000). Unlike other approaches to constructions, the starting point for this project is the grammar pattern. Analysis of patterns thus far suggests that up to 40 separate constructions can be identified for each pattern. To date, over 50 patterns have been analysed, resulting in the identification and description of approximately 1400 constructions.

In this paper, the process of identifying constructions from patterns is described and the information given about each construction is specified. Description of the construction has taken into consideration Perek and Patten's (2019) work in aligning patterns, constructions and FrameNet terminology and Hank's (2013) Corpus Pattern Analysis. The results of the identification and description of constructions is a searchable database that will be exemplified in the paper.

It is a major claim of the Construction Grammar literature that constructions form a hierarchy; Croft and Cruse (2004: 262), for example, refer to constructions as a 'structured inventory' which can be represented as 'a taxonomic network'. Initially in this project the aim was to identify verb-based constructions (phase 1) and subsequently to derive system networks (phase 2). It quickly became apparent, however, that the system network was a rational way to conceptualise the taxonomy of the constructions using any pattern. The paper provides illustrations of the system networks proposed by the project.

The system networks also make it possible to specify 'mid-point' constructions as well as 'end-point' constructions. For example, the 'V that' pattern is used in 20 end-point constructions, each with a fairly specific meaning ('say that', 'predict that', 'promise that', 'remember that', 'accept that', 'arrange that' etc). At a mid-point, more abstract constructions can be proposed: 'the V that communicate information construction'; 'the V that communicate future action construction'; 'the V that cognise information construction'; 'the V that cause event construction'.

The system networks also make it easier to specify the general semantic fields that constructions belong to. Work on these fields is ongoing, but the paper will give examples such as Communication and Cognition. Identifying such fields makes it possible to be consistent in the identification of the participant roles in each construction. This in turn will be crucial in the next phase of the project, where system networks with semantic fields as the entry-point, and constructions as the end-point, will be derived.

## References

- Croft, W. and Cruse, D.A. 2004. *Cognitive Linguistics*. Cambridge University Press.
- Hanks, P. 2013. *Lexical Analysis: Norms and Exploitations*. MIT Press.
- Hunston, S. 2022. The Other Grammarian's Dream: Constructions as the most delicate grammar. Paper given at LxGr2022.
- Hunston, S. and Francis, G. 2000. *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: Benjamins.
- Perek, F. and Patten, A. 2019. Towards an English constructicon using patterns and frames. *International Journal of Corpus Linguistics* 24(3): 354-384.

## **Construction grammar and lexico-grammar, and why they matter to each other**

Gaëtanelle Gilquin

Université catholique de Louvain, Belgium

[gaetanelle.gilquin@uclouvain.be](mailto:gaetanelle.gilquin@uclouvain.be)

In construction grammar (CxG), no distinction is made between linguistic phenomena traditionally described as lexical (e.g. idioms) and linguistic phenomena traditionally described as grammatical (e.g. phrasal patterns). They are all considered instances of constructions, in the sense of form-meaning pairings. This makes CxG an intrinsically suitable framework for the study of lexico-grammar. If we combine CxG with corpus linguistics, which is known for its close links with lexico-grammar, we are bound to be well-equipped to deal with lexico-grammatical features.

This presentation will show how the study of constructions such as causative and passive constructions in English can benefit from a CxG- and corpus-based approach which takes lexico-grammatical aspects into account. This starts with the extraction of the constructions, thanks to strategies which make it possible to retrieve more peripheral instances of a construction (e.g. pseudo-passives with an adjectival form like BE interested). This continues with the analysis, by means of techniques that seek to uncover the phraseological patterns of constructions, in particular collostructional analysis. This technique helps provide better descriptions of the constructions, including in an L1-L2 comparative perspective (showing, for example, that native writers prefer verbs such as seem or appear in the [X make Y Vinf] causative construction, whereas EFL learners prefer be and become). Finally, a CxG- and corpus-based approach centred around lexico-grammar can lead to pedagogical materials or methods that favour the production of more idiomatic constructions. Data-driven learning, for example, could be used in the classroom to help students notice the lexico-grammatical features of constructions. These different steps, from data extraction to pedagogical applications, will illustrate the main advantages of corpus approaches to lexico-grammar anchored in CxG.

## **ABSTRACTS: PRESENTATIONS**

## **Corpus-based analysis of lexicogrammatical features of international teaching assistants**

Heesun Chang

Texas A&M University, USA

[changh@tamu.edu](mailto:changh@tamu.edu)

Virtually all research universities in North America rely on international teaching assistants (ITAs) for a substantial portion of their undergraduate education. Unfortunately, ITAs have often received negative reactions from monolingual students and faculty members for insufficient English proficiency. Universities have made continuous efforts to properly train and assess ITAs' English proficiency. However, the emphasis of ITA training and assessment has been mainly on improving ITAs' pronunciation skills, while their lexico-grammatical features have been often overlooked. A few studies have elucidated the complex nature of ITAs' discourse and lexico-grammatical features, and researchers are still a long way from gaining a comprehensive understanding of ITAs' linguistic features. One reason is that many ITA discourse studies have relied on small sample sizes and have focused on analyzing selective lexico-grammar features, treating them as independent occurrences while overlooking their co-occurring patterns.

Based on a multidimensional framework (Biber & Conrad, 2019), this study aims to (1) increase our understanding of ITAs' instructional language use in ITA assessment by investigating co-occurring patterns of their lexico-grammatical features and (2) evaluate the extent to which those features correspond to the language and register features actually used in university settings. The study also examined the effects of ITAs' English proficiency levels, disciplines, and gender on their lexico-grammatical features. The data come from 186 prospective ITAs' mock-teaching presentations in ITA assessment at a large research university. Each presentation lasts approximately 10-13 mins, including a Q&A session where the audience asked questions regarding the topic of the presentation and the presenter responded to them. The presentations were transcribed, proofread, and built into an ITA corpus consisting of over 247,000 words, and over ninety linguistic features were annotated, using Biber Tagger and TagCount.

In the first phase, the study compared the lexico-grammatical features of ITA mock-teaching to those of ten spoken and written university registers, based on the four discourse functional dimensions of T2K-SWAL (Biber, 2006), a large-scale corpus designed to represent the range of spoken and written languages in various university contexts. The lexico-grammatical features of ITA mock-teaching were found to differ from those of university registers in manifold ways, and the features also varied across ITAs' proficiency levels, disciplines, and gender. The second phase of the study investigated how ITAs' lexico-grammar features vary internally by four discourse purposes in their presentations. Each presentation was manually segmented into the four discourse parts (i.e., introduction, lecture, concluding remarks, interaction with audience). Based on the six reference dimensions of general spoken and written English registers (Biber, 1988), distinctive patterns of internal lexico-grammatical variation were identified across the four discourse types. A k-means cluster analysis was performed to further explore and summarize the lexico-grammar patterns characterizing the four discourse types. The analysis identified three functional clusters: Impersonal discourse, Elaborative discourse, and Involved-Situated discourse. Overall, introduction, lecture, and interaction texts were mostly realized under Impersonal discourse, whereas concluding remarks were marked by Elaborative discourse features. The study discusses implications of the findings as well as limitations and directions for future studies.

## References

- Biber, D. (1988). *Variation across speech and writing*. Cambridge, UK: Cambridge University Press.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers* (Vol. 23). John Benjamins Publishing.
- Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2<sup>nd</sup> edition). Cambridge University Press.

## Productivity as abstractness in computational construction grammar

Jonathan Dunn

University of Canterbury Christchurch, New Zealand

[jonathan.dunn@canterbury.ac.nz](mailto:jonathan.dunn@canterbury.ac.nz)

This paper works within the Construction Grammar paradigm (CxG) in order to measure the productivity of constructions at different levels of abstraction. Drawing on a computational corpus-based approach to CxG (Dunn, 2017; Dunn, 2022), a *construction* is a representation learned from an unannotated corpus. For example, the constructions in (1) and (2) have been learned from a corpus of novels from Project Gutenberg and describe the set of utterances in (1a-1d) and (2a-2d). As shown in these examples, a construction consists of a series of slot-constraints; here the constraints are derived from lexical, syntactic, and semantic representations. The resulting tokens of the constructions capture lexico-grammatical units with a shared meaning but different forms. For example, (1) captures verb phrases with a meaning of causing changes to an abstract future state and (2) captures verb phrases which idiomatically occur with the argument "the importance".

(1) [ "to" — SYN: 113 <reassess\_demonstrate> — "the" — SYN: 230 <supposition\_contention> ]

- (1a) to improve the chance
- (1b) to affect the outcome
- (1c) to reduce the possibility
- (1d) to change the verdict

(2) [ SEM: 420 <underlines\_emphasizes> — "the" — "importance" ]

- (2a) underlining the importance
- (2b) highlight the importance
- (2c) emphasise the importance
- (2d) stresses the importance

From a lexico-grammatical perspective, these constructions differ in their level of abstraction: falling at different locations on the continuum between proto-typical lexical items and proto-typical grammatical items. The goal of this paper is to understand how constructions are distributed on this continuum by measuring the type/token ratio of constructions across corpora from different registers. The basic hypothesis is that more abstract constructions have a higher type/token ratio across corpora because there is a greater diversity of forms for that construction. Constructions with a low type/token ratio are expected to be less abstract.

Table. Type/Token Ratios by Construction Type and Corpus

	Blogs	Comments	Parliament	Gutenberg	Reviews	Wikipedia
<i>Adjectival</i>	0.63	0.62	0.61	0.75	0.56	0.59
<i>Adpositional</i>	0.68	0.68	0.50	0.68	0.66	0.55
<i>Adverbial</i>	0.74	0.75	0.68	0.84	0.69	0.80
<i>Idiom</i>	0.09	0.06	0.24	0.33	0.11	0.15
<i>Nominal</i>	0.86	0.85	0.70	0.86	0.82	0.77
<i>Sentential</i>	0.58	0.72	0.61	0.76	0.57	0.71
<i>Clausal</i>	0.63	0.71	0.64	0.73	0.64	0.83
<i>Transitional</i>	0.74	0.80	0.70	0.83	0.76	0.88
<i>Verbal</i>	0.76	0.81	0.76	0.88	0.75	0.74

We evaluate this hypothesis by annotating corpora from six registers with a grammar of constructions: Project Gutenberg (Rae et al., 2019), Wikipedia (Ortman, 2018), European Parliament proceedings (Tiedemann, 2012), news article comments (Kesarwani, 2018), product reviews (Zhang et al., 2015), and blogs (Schler et al., 2006). After calculating the type/token ratios for constructions in each corpus, we reach the average type/token ratios shown in the table above, here divided by type of construction (c.f., Dunn, 2023) and by corpus.

A linguistic analysis of constructions with different type/token ratios is undertaken to show that, with some interesting exceptions, this corpus-based measure can be used to sort constructions according to their location on the lexico-grammatical continuum.

## References

- Dunn, J. (2017). Computational Learning of Construction Grammars. *Language and Cognition*, 9(2): 254-292.
- Dunn, J. (2022). Exposure and Emergence in Usage-Based Grammar: Computational Experiments in 35 Languages. *Cognitive Linguistics*, 33(4): 659-699.
- Dunn, J. (2023). Exploring the Constructicon: Linguistic Analysis of a Computational CxG. In *Proceedings of the First International Workshop on Construction Grammars and NLP (CxGs+NLP, GURT/SyntaxFest 2023)*. Association for Computational Linguistics. 1-11.
- Kesarwani, A. (2018). *New York Times Comments*. Kaggle.  
<https://www.kaggle.com/datasets/aashita/nyt-comments>
- Ortman, M. (2018). *Wikipedia Sentences*. Kaggle.  
<https://www.kaggle.com/datasets/mikeortman/wikipedia-sentences>
- Rae, J.; Potapenko, A.; Jayakumar, S.; and Lillicrap, T. (2019). Compressive Transformers for Long-Range Sequence Modelling. *arXiv*: 1911.05507.
- Schler, J.; Koppel, M.; Argamon, S. and Pennebaker, J. (2006). Effects of Age and Gender on Blogging. In *Proceedings of AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*. Association for the Advancement for Artificial Intelligence.
- Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*. European Language Resources Association. 2214–2218.
- Zhang, Ziang, Junbo Zhao, and Yann LeCun. (2015). Character-level convolutional networks for text classification. *arXiv*: 1509.01626.

## The diachronic shaping of constructional meaning: A frequency trajectory cluster-based method to explore the emergence of polysemy

Quentin Feltgen

Ghent University, Belgium

[quentin.feltgen@gmail.com](mailto:quentin.feltgen@gmail.com)

Construction Grammar, insofar as it posits that all linguistic structures are signs, i.e., form-meaning pairs, highlights that meaning plays a foundational role in the linguistic knowledge of the language users. How this meaning comes to be construed in the case of schematic constructions (constructions with an open slot), however, is a paradox of sort: on the one hand, this slot's fillers participate to shape the construction's meaning (Stefanowitsch & Gries 2004); on the other hand, the coercion phenomenon (Lauwers & Willems 2011) allows constructions to imprint its own meaning onto unexpected fillers, such as the verb *sneeze* in Goldberg's famous example of the caused-motion construction, *Sue sneezed the napkin off the table* (Goldberg 1995). This paradox ties to the notion of prototypicality (Geeraerts 1997; core, frequent items define the construction's meaning, while marginal, rare items extend its use beyond this expected use), productivity (the ability of schema to be extended to new types), and analogy (the ability of the construction to recruit new fillers is fostered by the extent of the construction's coverage of this domain; Perek 2016).

Even though this view of the semantic structure is consistent in a synchronic perspective, the situation becomes more complex in diachrony. A construction may extend its use to new domains (Hilpert & Perek 2015, Desagulier 2022), weakening the identifiability of the prototypical meaning, and violating the isomorphic principle (De Smet et al. 2018). Furthermore, a construction with many types may be less productive than a construction with fewer types, but semantically more coherent (Barðdal & Gildea 2015). Without a clear core meaning, coercion is less efficient and the construction less productive; but a productive construction likely extends beyond its core meaning and loses semantic coherence.

We illustrate this complex dynamics over the semantic landscape with an example from French, the *venir de* 'immediate anteriority' construction, which grammaticalized in the sixteenth century (Bres & Labeau 2015), on the basis of over 60k occurrences from the Fantext corpus (ATILF 1998-2023). To study this construction, I use three methods in conjunction:

- (1) I model the token frequency with a succession of S-curves to detect three different phases of the evolution, based on the assumption that S-curves in frequency are the signature of semantic shifts (Feltgen et al. 2017);
- (2) I exploit the Herdan's law to describe productivity changes – showing remarkable consistency between (1) and (2)
- (3) I run a clustering algorithm on each fillers' frequency trajectory to detect clusters of fillers evolving in a cohesive way (Wiemer et al. 2021).

I select the best clustering model based on a custom two-fold cross-validation scheme – freely adapted from Wang (2010) –, which is 2 clusters using the relative frequency within the construction, instead of the collocate frequency itself.

The latter method offers an in-depth view into the diachronic dynamics that shape the construction's meaning through the behavior of the individual fillers. Different clusters correspond to separate semantic nuances that can be gained or left out by the construction (Jansegers & Gries 2020). The two clusters correspond respectively (based on a qualitative characterization) to a deictic/anteriority meaning, vs. a resultative/recency meaning. Interestingly, although these clusters compete over the construction, they mutually benefit from the success of the other: thanks to the later rise of the second cluster, the frequency of the first cluster takes off again, despite a loss in prototypicality.

## References

- ATILF (1998-2023). Base textuelle Frantext (online). ATILF-CNRS & Université de Lorraine. <https://www.frantext.fr/>
- Barðdal, J., & Gildea, S. (2015). Diachronic Construction Grammar: Epistemological context, basic assumptions and historical implications. In J. Barðdal, E. Smirnova, L. Sommerer, S. Gildea, *Diachronic construction grammar*. Amsterdam/New York : John Benjamins Publishing Company, 1-49.
- Bres, J., & Labeau, E. (2015). Venir de (+ infinitive) An immediate anteriority marker in French. *Diachronica. International Journal for Historical Linguistics*, 32(4), 530-570.
- De Smet, H., D'hoedt, F., Fonteyn, L., & Goethem, K. V. (2018). The changing functions of competing forms: Attraction and differentiation. *Cognitive Linguistics*, 29(2), 197-234.
- Desagulier, G. (2022). Changes in the midst of a construction network: a diachronic construction grammar approach to complex prepositions denoting internal location. *Cognitive Linguistics*, 33(2), 339-386.
- Feltgen, Q., Fagard, B., & Nadal, J. P. (2017). Frequency patterns of semantic change: corpus-based evidence of a near-critical dynamics in language change. *Royal Society open science*, 4(11), 170830.
- Geeraerts, D. (1997). *Diachronic prototype semantics: A contribution to historical lexicology*. Oxford University Press.
- Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. The University of Chicago Press.
- Hilpert, M., & Perek, F. (2015). Meaning change in a petri dish: constructions, semantic vector spaces, and motion charts. *Linguistics Vanguard*, 1(1), 339-350.
- Jansegers, M., & Gries, S. T. (2020). Towards a dynamic behavioral profile: A diachronic study of polysemous sentir in Spanish. *Corpus linguistics and linguistic theory*, 16(1), 145-187.
- Lauwers, P., & Willems, D. (2011). Coercion: Definition and challenges, current approaches, and new trends. *Linguistics*, 49(6), 1219–1235.
- Perek, F. (2016). Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics*, 54(1), 149-188.
- Wang, J. (2010). Consistent selection of the number of clusters via crossvalidation. *Biometrika*, 97(4), 893-904.
- Wiemer, B., Wrzesień-Kwiatkowska, J., & Wrzesień-Kwiatkowski, S. (2021). Diachronic dynamics of a stem-derivational aspect system: how cluster analysis helps discover patterns. *Korpus – Gramatika – Axiologie*, 23, 47-86.

## Giving gifts and doing favours: Support-verb constructions with *χάρις* in classical literary Attic

Victoria Fendel

University of Oxford, UK

[victoria.fendel@classics.ox.ac.uk](mailto:victoria.fendel@classics.ox.ac.uk)

Support-verb constructions, such as *to do a favour*, consist of a support verb which is (usually) form-identical with but shows a reduced argument grid vis-à-vis its full verb and auxiliary counterparts (Loporcaro 2022; Butt 1997) and a predicative noun which contributes the semantic weight. Support-verb constructions form a heterogeneous groups of constructions (Kamber 2008), such that they are discussed in the context of (morphological) periphrasis as well as (lexical) idiomaticity (Loporcaro 2022; Sheinfx et al. 2019). Structures that detransitivise, e.g. *to make a suggestion* vs *to suggest* (Marini 2010) could be considered under Haspelmath's (2000) categorial periphrasis; structures that have lost compositionality and/or analysability (Smith 2022) move further towards a lexical word (Taylor 2015).

The present paper explores the support-verb-construction family built with the predicative noun *χάρις* in classical literary Attic Greek (5<sup>th</sup> / 4<sup>th</sup> c. BC). *χάρις* combines with a range of support verbs, amongst them two recipient passive formations, one shared across the class of support-verb constructions, one specific to this family (cf. Keenan & Dryer 2007). *χάρις* also combines with a range of verbs of realisation (Mel'čuk 2004), including the preferred active which seems lexicalised to an extent as reflected by its discontinuity and constrained variability (cf. Savary et al. 2018).

The corpus of analysis is a selection of literary classical Attic prose, oratory and historiography (Lysias, Isaeus, Antiphon, Demosthenes; Plato, Aristotle; Thucydides, Xenophon). Support-verb constructions are extracted from this corpus semi-automatically using the Thesaurus Linguae Graecae and Sketch Engine (Fendel & Ireland forthcoming). The support-verb-construction family of interest is examined from the perspectives of prototypicality and canonicity (Smith 2022; Kamber 2008) with regard to their lexical idiomaticity and formal syntagma-character. The paper shows that even structures belonging to the same support-verb-construction family oscillate between more periphrastic and more idiomatic and do not form a homogenous group. Support-verb constructions lie at the interface of syntax and semantics as well as morphosyntax and lexicon.

### References

- Butt, Miriam. 1997. Complex predicates in Urdu. In Alex Alsina, Joan Bresnan & Peter Sells (eds.), *Complex predicates*, 107–149. Stanford: CSLI Publications.
- Fendel, Victoria & Matthew Ireland. forthcoming. Discourse cohesion in Xenophon's *On Horsemanship* through Sketch Engine. *Digital Humanities Quarterly*.
- Haspelmath, Martin. 2000. Periphrasis. In Geert Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphologie / Morphology. Ein internationales Handbuch zur Flexion und Wortbildung / An international handbook on inflection and word-formation*, vol. 1, 654–664. Berlin ; Boston: Mouton De Gruyter.
- Kamber, Alain. 2008. *Funktionsverbgefüge-- empirisch: eine korpusbasierte Untersuchung zu den nominalen Prädikaten des Deutschen* (Reihe Germanistische Linguistik). Vol. 281. Tübingen: Max Niemeyer.

- Keenan, Edward & Matthew Dryer. 2007. Passive in the world's languages. In Timothy Shopen (ed.), *Language Typology and Syntactic Description: Volume 1: Clause Structure*, vol. 1, 325–361. 2nd edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511619427.006>.
- Loporcaro, Michele. 2022. The morphological nature of person-driven auxiliation. In Adam Ledgeway, John Charles Smith & Nigel Vincent (eds.), *Periphrasis and inflexion in diachrony: a view from romance* (Oxford Studies in Diachronic and Historical Linguistics 48), 213–237. Oxford: University Press.
- Marini, Emanuela. 2010. L'antipassivo in greco antico: ποιείσθαι come verbo supporto in Aristotele. *Journal of Latin Linguistics* 11(1). 147–180. <https://doi.org/10.1515/joll.2010.11.1.147>.
- Mel'čuk, Igor. 2004. Verbes supports sans peine. *Lingvisticæ Investigationes* 27(2). 203–217. <https://doi.org/10.1075/li.27.2.05mel>.
- Savary, Agata, Marie Candito, Verginica Mititelu, Eduard Bejček, Fabienne Cap, Slavomír Čéplö, Silvio Cordeiro, et al. 2018. PARSEME multilingual corpus of verbal multiword expressions. In Stella Markantonatou, Carlos Ramisch, Agata Savary & Veronika Vincze (eds.), *Multiword expressions at length and in depth: Extended papers from the MWE 2017 workshop*, 87–147. Berlin: Language Science Press.
- Sheinfux, Livnat, Tali Greshler, Nurit Melnik & Shuly Winter. 2019. Verbal multiword expressions: Idiomaticity and flexibility. In Yannick Parmentier & Jakub Waszczuk (eds.), *Representing and parsing of multiword expressions* (Phraseology and Multiword Expressions), vol. 3, 35–68. Berlin: Language Science Press.
- Smith, John Charles. 2022. The boundaries of inflexion and periphrasis. In Adam Ledgeway, John Charles Smith & Nigel Vincent (eds.), *Periphrasis and inflexion in diachrony: a view from romance* (Oxford Studies in Diachronic and Historical Linguistics 48), 61–90. Oxford: University Press.
- Taylor, John. 2015. *The Oxford Handbook of The Word* (Oxford Handbooks in Linguistics). 1st edn. Oxford: Oxford University Press.

## The ragged middle: A functional approach to lexis.

Lise Fontaine

Université de Québec à Trois-Rivières, Canada

[Lise.Fontaine@uqtr.ca](mailto:Lise.Fontaine@uqtr.ca)

This paper offers a theoretical exploration of the reasons why textually-oriented approaches to lexis (e.g. systemic functional linguistics, Halliday 1994) and more lexically-driven approaches typically found within corpus linguistics (e.g. Hunston and Francis, 2000) struggle to find common ground. For Halliday (1961, p 69), '[t]he middle ranks of the grammar are often the most complex, presumably since they face both ways; so that a grammar which starts unidirectionally from the two ends will find it difficult to avoid leaving the middle ragged'. Halliday's 1961 proposal presented lexis as most delicate grammar, an idea developed further by Hasan (1996). As rather brutally pointed out by Daley, Jones & Sinclair (2004 p. 3), this approach seemed to give lexis 'the role of picking up the scraps from the tables of syntax'.

While I agree that such a position is problematic, it would be wrong to throw the baby out with the bath water. In this talk, I attempt to offer a reconciliatory position. A focus on lexical form provides important empirical evidence from which we can form abstractions, but we need a way of integrating this kind of meaning into the framework. Speakers 'necessarily associate each word with a network of paradigmatic and syntagmatic connections that conditions a specific sense' (Polguère 2015). This perspective seems compatible with Halliday's early views on lexis (1961, 1966). The SFL concept of 'meaning potential' can be developed, as proposed by Allwood (2003) and Fontaine (2017), to develop a meaning-driven account of lexis. Given that 'system', as meaning potential, is an abstraction from instances, lexical meaning can also be seen as an abstraction from instances.

Using a dataset of corpus instances of a small selection of verbs, I combine transitivity analysis with a corpus-based formal pattern analysis (Hanks 2004) and situation type analysis (e.g. drawing on Vendler's (1997) classes). This work in progress presents the initial steps towards a theoretical account of meaning potential in relation to lexis. In doing so, I challenge the existing metaphor of grammar as bigger than lexis. As Mel'cuk (1981, p 57) says, 'not only every language but every lexeme of a language is an entire world in itself'. The world of the word is vast and complex, and it is generally accepted that it is represented as part of a large complex network. For example, the 'knowing about' a context and the 'knowing about' a word is contained and maintained within an individual's networked cognitive system. In this sense, there might not be much difference between knowing about a context and knowing about a word.

This is an especially provocative thought if we consider how the concepts of meaning potential and instantiation can be applied to both context and lexis (Fontaine 2017). The ultimate aim of this paper is to examine the extent to which Hanks' (2013) theory of norms and exploitations can offer an important development in the problem of the ragged middle.

## References

- Allwood, J. 2003. Meaning potentials and context: Some consequences for the analysis of variation in meaning. In Cuyckens, H., Dirven, R. & Taylor, J. (eds.) *Cognitive approaches to lexical semantics*, Mouton De Gruyter, 29-66.
- Daley, R., Jones, S. & Sinclair, J. 2004. *English Collocation Studies*. London: Bloomsbury Publishing.
- Fontaine, L. 2017. Lexis as most local context: towards an SFL approach to lexicology. *Functional Linguistics*. 4: 17.
- Halliday, M.A.K. 1961. Categories of the theory of grammar. *WORD* 17(2). 241–92.
- Halliday, M.A.K. 1966. Lexis as a linguistic level. In Bazell, C.E. et al. (eds.) *In memory of J.R. Firth*. London: Longman.
- Halliday, M. A. K. 1994. *An Introduction to Functional Grammar*. London: Edward Arnold.
- Hanks, P. 2013. *Lexical analysis: Norms and exploitations*. Cambridge, Mass: MIT Press.
- Hasan, R. 1996. The grammarian's dream: Lexis as most delicate grammar. In Cloran, C., Butt, D. & Williams, G. (eds.) *Ways of saying, ways of meaning: Selected papers of Ruqaiya Hasan*, 73–103. London: Cassell.
- Hunston, S. & G. Francis. 2000. *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
- Polguère A. 2015. Lexical Contextualism: The Abélard Syndrome. In Gala, N., Rapp, R. & Bel-Enguix, G. (eds.) *Language Production, Cognition, and the Lexicon, Text, Speech and Language Technology* 48, Springer, Cham Heidelberg New York Dordrecht London, 53–73.
- Vendler, Z. 1967. *Linguistics in philosophy*. Ithaca N.Y.: Cornell University Press.

## Verbal binomials: A gradual phenomenon between coordination, construction and lexical unit

Susanne Handl

LMU Muenchen, Germany

[susanne.handl@anglistik.uni-muenchen.de](mailto:susanne.handl@anglistik.uni-muenchen.de)

In an early definition of the notion *binomial*, Malkiel (1959: 113) starts out with a purely descriptive approach, calling it "the sequence of two words pertaining to the same form-class, placed on an identical level of syntactic hierarchy, and ordinarily connected by some kind of lexical link". Since not all of those sequences exhibit the same level of formulaicity, he suggests "a continuum of subtly graded possibilities [ranging from] definitive coalescence (entailing irreversibility) [to] unimpaired freedom of variation" (Malkiel 1959: 116).

Still, research into binomials has generally focused on the coalescence extreme (e.g. Mollin 2014). So a more typical definition now would be – here for bi- and trinomials – "fixed sequences of two or three word forms that belong to the same part-of-speech category and are linked by the conjunction 'and' or 'or'" (Granger/Paquot 2008: 43).

Focusing on *and* can provide an insight into the more open section of the binomial fixedness continuum. The grammatical side of *and* is pure addition – from a phraseological perspective the conjunction might as well be responsible for creating collocational frames or constructions where different slot fillers lead to variations in fixedness and formulaicity.

To investigate the function of *and* in this context, I have carried out an analysis of verbal binomials in the British National Corpus. Analysing verbs seems most promising, as they are semantically much more versatile than the typical binomial adjectives. My hypothesis is that this field is strong on the formulaic side, but only for a very small range of verbs. The OED backs this, claiming that the first slot is mostly filled with *come*, *go*, *send*, and *try*, used in the infinitive or the imperative only. On the other end of the continuum there are V-and-V-structures that are less idiomatic but still not a simple addition of verb meaning.

Starting out from existing research (e.g. Bachmann 2013, Schönefeld 2012, Stefanowitsch 1999, Wulff 2006), I have collected data for a comprehensive overview of VERB *and* VERB sequences. Preliminary findings suggest that the word forms fulfill different functions in different registers, and that each slot has its own preferences. Infinitive and past tense forms, for instance, are mainly used with mental and communication verbs, whereas stative verbs in the first slot often suggest an intensifying function. Present participles are rare and the past participle is highly active in a more formal, sometimes legal register. Verbs in the base form are used for various semantic areas, including the typical concept of the form GO and DO.

The analysis so far suggests that verbal binomials like *go and buy*, *sit and eat* are not only sequences of the same part-of-speech but also of different valency. The less idiomatic V-and-V-sequences, however, combine verbs with the same valency. Therefore in a follow-up study I focus on the criterion of split vs. common valency of the two slots and analyse the semantic interrelation between the slots, the cognitive effect of the conjoint verbs, thus paving the way for a V-and-V-construction.

## References

- Bachmann, Ingo (2013) Has go -V ousted go-and -V? A study of the diachronic development of both constructions in American English. In: Hasselgård, Hilde, Ebeling, Jarle & Signe Oksefjell Ebeling, eds. (2013), *Corpus perspectives on patterns of lexis*, Amsterdam: John Benjamins, 91–112.
- Granger, Sylviane & Magali Paquot (2008), "Disentangling the phraseological web", in: Granger, Sylviane & Fanny Meunier, eds. (2008), *Phraseology: An interdisciplinary perspective*, Amsterdam: John Benjamins, 27-49.
- Malkiel, Yakov (1959), "Studies in irreversible binomials", *Lingua* 8, 113-160.
- Mittmann, Brigitta (2004), Mehrwort-Cluster in der englischen Alltagskonversation. Unterschiede zwischen britischem und amerikanischem gesprochenen Englisch als Indikatoren für den präfabrizierten Charakter der Sprache, Tübingen: Gunter Narr Verlag.
- Mollin, Sandra (2014), *On the (ir)reversibility of English binomials*, Amsterdam: John Benjamins.
- Schönefeld, Doris (2012), *Things going unnoticed - A usage-based analysis of go-constructions*. In Divjak, Dagmar & Stefan Gries, eds. (2012), *Frequency effects in language representation*, Berlin: Mouton de Gruyter, 11-49.
- Stefanowitsch, Anatol (1999) The go-and-Verb construction in a cross-linguistic perspective: image-schema blending and the construal of events. In Nordquist, Dawn and Catie Berkenfield (1999), *Proceedings of the Second Annual High Desert Linguistics Society Conference*, Albuquerque, NM: High Desert Linguistics Society, 123-134.
- Wulff, Stefanie (2006), Go-V vs. go-and-V in English: A case of constructional synonymy?, in: Gries, Stefan Th. & Anatol Stefanowitsch, eds. (2006), *Corpora in Cognitive Linguistics*, Berlin: Mouton de Gruyter.101–126.

## When was the last time your construction's meaning went undetermined? On the indeterminate aspectual behavior of *go un-participle* constructions

Marin Kežić

University of Zagreb, Croatia

[mke018@post.uit.no](mailto:mke018@post.uit.no)

Ever since Radden's landmark study (Radden 1996), *go un-participle* constructions – acting as more marked counterparts of *be un-participle* constructions (i.e. *unpassives*, see Siegel 1973) – have been a fascinating subject of enquiry within functionalism (Bourdin 2003, Schönefeld 2012, Brdar 2014). Existing functionalist literature agrees that the constructions in question code “abnormal and unexpected states of affairs” (Radden 1996: 449), thus conferring a sense of what might be termed *evaluative modality* (Bourdin 2003).

By employing a corpus-based approach, this paper attempts to broaden and partially reframe the debate on the underlying image-schematic structure (Radden 1996) and the aspectual behavior (Bourdin 2003) of the studied constructional type. In particular, the present account reports on distributional data from the British WaCky Web Corpus (UkWaC, see Ferraresi et al. 2008) to follow up on Bourdin's small-scale statistical survey of the BNC (Bourdin 2003: 108-109) and his cursory remark that “*go un-V-en* favours contexts involving temporal extension over those that focus on a moment in time.” (Bourdin 2003: 113). Usage examples involving aspectual cues such as duration and *scalar* (see Israel 2011) adverbs, as in (1), and (2) are contrasted with instances where the immediate context points to the construal of given scenarios as punctual events, as in (3) and (4):

- (1) The streets **went *uncleared*** *for weeks*. (Bourdin 2012: 112)
- (2) (...) Rohmer's surrealist moments are kept to a minimum and **go *mostly unexplained***. ([www.thezreview.co.uk](http://www.thezreview.co.uk))
- (3) Some of us get pretty depressed when the wine is finished, our letters **go *unanswered***, and the attractive stranger (...) just walks on by. ([humanism.org.uk](http://humanism.org.uk))
- (4) Not wishing to offend my new young cricketing friend at such a sensitive moment, the question **went *unasked*** (...). ([guardian.co.uk](http://guardian.co.uk))

The present analysis is concerned with contrasting instances of *go un-participle* constructions that seem to unequivocally call for continuative interpretations, and those that considerably favor inchoative interpretations (consider also the option to substitute the pseudo-passives under (3) and (4) with *get-passives* such as *don't get answered* and *didn't get asked*, respectively). Crucially, however, the study also points to instances where there the immediate context lacks any disambiguating *correlating elements* (see Ziem, Flick, Sandkühler 2019) and thus allows for ambiguity between *extent-durational* and *point-durational* (see Talmy 2000) interpretations, as is arguably the case in (5):

- (5) With a restless, inquisitive mind, he cannot let any problem **go *unsolved***. ([malefirst.co.uk](http://malefirst.co.uk))

To account for this constructional *indeterminacy* (Langacker 1998), the present approach re-evaluates the image-schematic structure of the verb *go* (see Radden 1996, Bourdin 2003). Specifically, it discusses its ability to profile both the SOURCE and PATH subpart of the SOURCE-PATH-GOAL schema and its association with both inchoative and continuative interpretations. As a means of elucidating the link between the underlying image-schematic motivation and aspectual behavior of these constructions, the study also considers their interconnectedness with kindred *go*-constructions and the impact of the verb *go*'s tense-aspect variation on its *(anti)resultative* (see Plungian 2001, Goldberg 1995, respectively) tinge.

## References

- Bourdin, P. (2003). On two distinct uses of *go* as a conjoined marker of evaluative modality. In R. Facchinetti, F. Palmer, & M. Krug (Eds.), *Modality in Contemporary English* (pp. 103–127). De Gruyter Mouton.
- Brdar, M. (2016). Going productive with metonymy. In I. Zovko Dinković & J. Mihaljević Djigunović (Eds.), *English Studies from Archives to Prospects: Volume 2; Linguistics and Applied Linguistics* (pp. 57–72). Cambridge Scholars Publishing.
- Ferraresi, A. et al (2008). Introducing and evaluating ukWaC, a very large web-derived corpus of English. In S. Evert, A. Kilgarriff, & S. Sharoff (Eds.), *Proceedings of the 4th Web as Corpus Workshop (WAC-4)* (pp. 47–54). ELRA
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*, Chicago University Press.
- Israel, M. (2011). *The grammar of polarity: Pragmatics, sensitivity, and the logic of scales*. Cambridge University Press.
- Langacker, R. W. (1998). Indeterminacy in semantics and grammar. In J. L. Cifuentes Honrubia (Ed.), *Estudios de Lingüística Cognitiva II* (pp. 649–672). Universidad de Alicante, Departamento de Filología Española, Lingüística General y Teoría de la Literatura.
- Plungian, V. (2001). Antirezultativ: Do I posle rezultata. In V. Plungian (Ed.), *Glagol'nye kategorii* (pp. 50–88). Russkie slovari.
- Radden, G. (1996). Motion metaphorized: The case of coming and going. In E. H. Casad (Ed.), *Cognitive linguistics in the redwoods* (pp. 423–458). Mouton de Gruyter.
- Schönefeld, D. (2012). Things going unnoticed – A usage-based analysis of *go*-constructions. In S. Th. Gries & D. Divjak (Eds.), *Frequency effects in language representations* (pp. 11–49). De Gruyter Mouton.
- Siegel, D. (1973). Nonsources of unpassives. In *Syntax and Semantics: Volume 2* (pp. 301–317). Brill.
- Talmy, L. (2000). *Toward a cognitive semantics: Volume I*. Concept structuring systems. MIT Press.
- Ziem, A., Flick, J., & Sandkühler, P. (2019). The German Constructicon Project: Framework, methodology, resources. *Lexicographica*, 35(1), 15–40.

## **Scraping the bottom of the corpus: Exploring lexical emergence in Italian time-stamped corpora**

Mauro LE DONNE

University for Foreigners of Perugia, Italy

[mauro.ledonne@unistrapg.it](mailto:mauro.ledonne@unistrapg.it)

In recent years there has been growing interest on lexical emergence, defined as ‘the process through which new word forms spread across a population of speakers’ (Grieve, Nini, Guo 2017: 102). Whereas past studies considered productive certain formative patterns only inasmuch they were massively represented in dictionaries (Aronoff, 1976), in the past two decades, quantitative approaches based on language corpora have been considered more reliable to investigate the productivity of word formation processes. The percentages of hapax legomena in corpora have been considered a valid test for the viability of a given formative and its status in terms of profitability in a given language (Baayen, Renouf 1996: 72-73; Bauer 2001). Nonetheless, it has been noted that also such measure can provide dubious results if the corpus is not large enough; indeed, what could be labeled as hapax inside a small corpus could be instead an old-fashioned term or a rare word in a larger corpus (Gaeta, Ricca 2006: 70).

In order to overcome this and other limitations, diverse tools that exploit the web as a potentially infinite source of linguistic data have been developed (Kilgariff 2001; Baroni, Ueyama 2006), for instance, queries may be performed on a web corpus or on a web-based interface (Lýdeling, Evert, Baroni 2007). More recently, data mining tools, like Neocrawler (Kerremans, Stegmayr, Schmid 2012) and Neoveille (Cartier 2016), allow to identify neologisms from the web, to collect them and to constantly enlarge their database. However, such instruments are still not available for many languages. As this study is focused on lexical emergence in Italian, a different corpus-driven approach must be adopted. To trace back the spreading of Italian hapaxes and occasionalisms, a time-stamped corpus is consulted through the Sketch Engine interface (Kilgariff et al. 2004), namely, the Italian web corpus Timestamped JSI 2014-2021 (Buřta et al. 2017), made up of news articles gained from RSS feeds annotated diachronically. In order to set up a manageable sample of neologisms of Contemporary Italian, the query is restricted using text-specific filters that contain information on year of occurrence, source country and web domain. As the aim is to obtain a sample of representative hapaxes and occasionalisms, we set the absolute frequency to a value ranging from 10 to 50. An additional filter is added to retrieve all the entries marked by metasignals (Svanlund 2018) on the left and right context of the KWIC.

This action returns a valuable sample of neologisms and fewer instances of other words, which can be excluded at a later stage. Following this methodology, a list of different neologisms will be compiled for each year covered by the corpus. A diachronic analysis of the range of years considered for the most recurrent neologism will follow, together with considerations relating the neologism origin (word formation process) and its meaning. This will consent us, at least to some extent, to draw a comparison on recent word formation studies focused on Contemporary Italian (Rainer 2015; Iacobini, Thornton 2016; Micheli 2020).

## References

- Aronoff, M. (1976). *Word Formation in Generative Grammar*. Cambridge, London. MIT Press.
- Baayen, H., Renouf, A. (1996). Chronicling the Times: Productive Lexical Innovations in an English Newspaper. *Language*, 76 (1): 69-96.
- Baroni, M., Ueyama, M. (2006). Building General- and Special-Purpose Corpora by Web Crawling. In *Proceedings of 13th NIJL International Symposium: Language Corpora. Their Compilation and Application*. Tokyo: 31-40.
- Bauer, L. (2001). *Morphological Productivity*. Cambridge. Cambridge University Press.
- Bušta, J., Herman, O., Jakubíček, M., Krek, S., Novak, B. (2017). JSI Newsfeed Corpus. In *9<sup>th</sup> International Corpus Linguistics Conference*, University of Birmingham, 25-28 July.
- Cartier, E. (2016). Neoveille, système de repérage et de suivi des néologismes en sept langues. *Neologica*, 10: 101-131.
- Gaeta, L., Ricca, D. (2006). Productivity in Italian word formation: a variable-corpus approach. *Linguistics*, 44 (1): 57-89.
- Grieve, J., Nini, A., Guo, D. (2017). Analyzing lexical emergence in Modern American English online. *English Language and Linguistics*, 21(1): 99-27.
- Kerremans, D., Stegmayr, S., Schmid, H.-J. (2012). The NeoCrawler: Identifying and Retrieving Neologisms from the Internet and Monitoring Ongoing Change. In Allan, K., Robinson, J.A. (eds.), *Current Methods in Historical Semantics*. De Gruyter Mouton.
- Kilgariff, A. (2001). Web as corpus. *Proceedings of Corpus Linguistics 2001*, Lancaster University, 30 March. Lancaster: 342-344.
- Kilgariff, A., Rychlý, P., Smrž, P., Tugwell, D. (2004). The Sketch Engine. In *Proceedings of the 11th EURALEX International Congress. Lorient*: 105-116.
- Lüdeling, A., Evert, S., Baroni, M. (2007). Using Web Data for Linguistic Purposes. *Language and Computers*, 59: 7-24.
- Micheli, M. S. (2020). *La formazione delle parole in italiano*. Roma. Carocci editore.
- Rainer, F. (2016). Italian. In Mýller,ÉP., O., Ohnheiser,ÉI., Olsen, S., Rainer, F. (eds.). *Word-Formation. An International Handbook of the Languages of Europe 4*. De Gruyter Mouton.
- Svanlund, J. (2018). Metalinguistic comments and signals: What can they tell us about the conventionalization of neologies? *Pragmatic & Cognition*, 25(1): 121-140.
- Thornton, A. M., Iacobini, C. (2016). Morfologia e formazione delle parole. In Lubello, S. (ed.) *Manuale di linguistica italiana*. De Gruyter: 190-221.

## **L2-textbook collocation frequency, congruency and syntactic type effects on Chinese adolescent and adult EFL learners**

Ni Li

Ocean University of China; Idaho State University, USA

[nili@isu.edu](mailto:nili@isu.edu)

The current research aims to unravel the collocation processing pattern by investigating the performance of Chinese EFL learners with a wider range of proficiency levels. As for the L2 vocabulary development, Kroll and Stewart (1994) and Jiang (2000) respectively proposed the Revised Hierarchy Model and a lexical development model with three stages. However, these two models are applied to single words rather than collocations that contain more than one lexical item. So far, there is little evidence of a development model for collocation processing and representation. Hence, we tentatively assume that an L2 collocation will go through the same developmental stages as the single words do in the above two models. Based on previous research, we make an assumption regarding the L2 learners' collocation processing: with the gain of L2 proficiency, the L1's influence on L2 learners will gradually attenuate and be replaced by L2-specific knowledge. The topic is important to language acquisition as studies have found that it is imperative for language learners to gain the knowledge of collocations in order to become more natural and competent.

In order to represent the authentic language input as closely as possible, the experimental items were culled from a L2 textbook corpus with a finer-grained classification of collocation congruency. An acceptability judgment task was conducted to evaluate language performance with congruent, L1-only, L2-only and partially-congruent collocations, as a more nuanced congruency category. Mixed-effects models were used to analyze the data, revealing that L2 learners were sensitive to the congruency condition, L2 collocation frequency and collocation syntactic types. Specifically, the congruency effect is more pronounced among lower proficiency learners, while the frequency effect is stronger for higher proficiency groups.

Verb-noun (VN) collocations were easier to be processed than adjective-noun (AN) collocations. These findings are partially in line with the three-stage model proposed by Jiang (2000) where the connection between concept and L2 word is strength with the increasing proficiency levels. On the other hand, the persistent difficulty in L1-only collocations inhibits the model's applicability to the domain of collocations. Methodologically, our study highlights the need to consider the syntactic structure when investigating the collocation processing.

The present study revealed significant differences between VN and AN collocations in both response times (RTs) and accuracy rates (ACCs) measures, and these differences were found to vary across different levels of language proficiency. In addition, the analyses of the current study have reaffirmed that RTs and ACCs are measures representing different latent factors and thus revealing different information. Furthermore, the findings may have implications for the design of collocation teaching materials and curricula, by highlighting the importance of considering the congruency and linguistic features of collocations in order to optimize learners' processing and retention of these lexical units.

## Part-of-Speech patterns in research articles: A cross-sectional analysis

(Kathy) Ling Lin & Yang Zhang

Shanghai Jiao Tong University, China

[kathyll@sjtu.edu.cn](mailto:kathyll@sjtu.edu.cn), [zhangyanga9@sjtu.edu.cn](mailto:zhangyanga9@sjtu.edu.cn)

This study combines the corpus-based Part-of-Speech gram (PoS-gram) approach with ESP genre analysis to examine key PoS-grams and their associated language patterning and variability across the four major sections of 300 Introduction-Methods-Results-Discussion (IMRD)-structured empirical research articles from medical science. A PoS-gram, as defined by Stubbs (2007, p. 91), is “a string of part-of-speech categories”, “the tokens of which are strings of words that have been annotated with these PoS tags” (Pinna & Brett, 2018, p. 107). Stubbs (2007) considered it as a type of “routine phraseology”, in addition to n-grams and phrase-frames. Yet, as phraseology is generally defined in corpus linguistics research as “the recurrent co-occurrence of words” (Clear, 1993, p. 277) and the compositional unit of a PoS-gram is a PoS category (grammatical category) rather than a word form, PoS-grams in our understanding may arguably not be a type of phraseology (Lin and Liu, 2021). Accordingly, we only treat it as a phraseology-related concept, since the exponents of each PoS-gram may be potential phraseology and the identification of it can be an effective way to extract recurrent phraseologies and patterns (Pinna & Brett, 2018).

Compared to abundant literature on academic phraseology study, the corpus-based PoS-gram analytical procedure has hardly been applied to the analysis of academic language use in research genres, except our recent study of PoS-grams in academic introductions across two contrasting disciplines (Lin and Liu, 2021). Lin and Liu (2021) have demonstrated the applicability of the corpus-based PoS-gram procedure to the analysis of language patterning in article introductions as a particular part-genre and has effectively yielded the associated cross-disciplinary similarities and differences in PoS-gram use and the links between PoS-gram use and the rhetorical functions of introductions. Given the huge potential of the corpus-based PoS-gram procedure to academic discourse study, the present work, to the best of our knowledge, for the first time extends it to the whole research articles structured in the conventional Introduction-Method-Results-Discussion (IMRD) format in medical science (Lin and Evans, 2012).

Using a self-compiled 1.3-million-word corpus, comprising 300 IMRD-structured internal medicine articles, as the reference corpus, we have identified key PoS-grams and their associated lexico-grammatical frames in each of the four major sections (viz., the Introduction, Method, Results, and Discussion) and made a cross-generic comparison. In the identification and concordance search of key PoS-grams, Sketch Engine with their modified English TreeTagger PoS tagset was adopted (Kilgariff et al., 2014).

The study reveals the close connection between key PoS-grams and their representative language patterns and the communicative functions of each part-genre. For instance, the top six 6-PoS-grams ranked by key scores in the Introduction have all expressed the function of purpose-statement and the lexicogrammatical frame generated based on them is “the {aim/goal/purpose/objective} of {this/the} {study/analysis} was to {maintain/ assess/ compare/ determine/ test/ examine/ investigate}”. In the Method section, the key PoS-gram with the highest key value is “DT NNS VVD JJ JJ NN” (“All patients provided written informed consent”). In the Result, the key PoS-grams identified are associated with the functions of Pointer, Substantiation, and Evaluation of findings (Brett, 1994; Williams, 1999), while in the Discussion,

the key PoS-grams such as NN IN DT NN VBZ IN/that (“limitation of this study is that”) and DT NNS IN DT NN VVP (“The strengths of this study include”) are linked to the functions of summarizing and reflecting on the study’s value and weaknesses. The results provide robust evidence for the feasibility of the corpus-based PoS-gram procedure for academic genres. The key PoS-grams and their associated language patterns identified could characterize salient generic and disciplinary features, which can be directly used in EAP teaching and material writing. The findings have important implications, particularly for medical science research writing and pedagogical practice.

### Acknowledgements

This work was supported by the National Social Science Fund of China (Project no.: 18CYY054).

### References

- Kilgarrieff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7-36.
- Lin, L. & Evans, S. (2012). Structural patterns in empirical research articles: A cross-disciplinary study. *English for Specific Purposes*, 31(3), 150-160.
- Lin, L. and Liu, M. (2021). Towards a Part-of-Speech (PoS) gram approach to academic writing: A case study of research introductions in different disciplines. *Lingua*, 254(2), 1-18.
- Stubbs, M. (2007). An example of frequent English phraseology: Distributions, structures and functions. In R. Facchinetti (Ed.), *Corpus linguistics 25 years on* (pp. 89-105). Amsterdam: Brill Rodopi.

## Different speaker-related uses of “I think”: Disentangling pure subjective and epistemic modal uses

Jiqiang Lu & Caroline Gentens

KU Leuven, Belgium

[jiqiang.lu@kuleuven.be](mailto:jiqiang.lu@kuleuven.be), [caroline.gentens@kuleuven.be](mailto:caroline.gentens@kuleuven.be)

It is well acknowledged that except for canonical modal auxiliaries like *might* and *must*, adjectives like *possible* and *certain*, adverbials like *possibly* and *certainly*, and discourse secondary complement-taking predicates (henceforth, CTPs) such as *I think*, *I am sure*, *there is a chance* are increasingly establishing themselves as an integral part of the English modal system (Halliday, 1970; Davidse et al., 2022). Among these CTPs, *I think* has received the most extensive attention from the perspective of grammaticalisation and subjectification (Boye & Harder, 2007; Thompson & Mulac, 1991ab; Van Bogaert, 2011; Shank et al., 2016).

However, Janssens and Nuyts (2021) proposed that Dutch expressions similar to English *I think*, can also be used to express the speaker’s personal attitude towards the state of affairs at issue, which they termed as “pure subjectivity”, i.e., the assessor’s “strictly personal opinion.” For example, in (1), *I think* clearly signals an epistemic stance, which can semantically be substituted by the modal auxiliary *might* as *He might be in the library*. In contrast, *I think* in (2) is used to express the speaker’s personal attitude or opinion about the state of affairs, meaning “According to me/In my opinion, she is amazing.” *I think* as a whole as a grammatical construction, in this sense, is used in a way contrasting with the original lexical meaning of the mental state verb *think*, as in (3).

(1) A: Where is John? B: **I think** he’s in the library, but you need to check. (Janssens & Nuyts, 2021: 228)

(2) I know Jocelyn and **I think** she is amazing. (WB)

(3) The present is hateful but when I **think** of the future it is so nice. (WB)

Janssens and Nuyts (2021) developed this semantic distinction in their diachronic analysis of Dutch *denken* ‘think’, *dunken* (impersonal variant of) ‘think’, *geloven* ‘believe’ and *vinden* ‘find’. We would like to extend their analysis to English examples as in (1) and (2), whilst at the same time developing new criteria to distinguish the subjective use as in (1) from the epistemic use as in (2), by focusing on both the grammatical behaviour and the possible discursal motivations.

To do this, we take a corpus-based approach to examine the English counterpart *I think*. There are 67,871 potential hits of *I think* in the British Spoken Corpus of the Collins Wordbanks Online (WB), but lexical uses and unintelligible examples will be filtered out. Based on a random sample of 400 valid hits, we then investigate the proportion of *I think* used as subjective markers and epistemic modal markers respectively as well as the criteria to distinguish their different use. Some crucial parameters that underlie the distinction may be the positions *I think* takes within the clause, the presence of the complementiser *that*, and interactional motivation, for example, whether *I think* is associated with answering preceding questions as in (1) or initiating a dialogue as in (2).

## References

- Boye, K., & Harder, P. (2007). Complement-taking predicates: Usage and linguistic structure. *Studies in Language*, 31(3), 569-606.
- Davidse, K., Van linden, A., & Brems, L. (2022). A semiotic approach to grammaticalization: modelling representational and interpersonal modality expressed by verbonominal patterns. *Language Sciences*, 91, 101473.
- Halliday, M. A. K. (1970). Functional diversity in language as seen from a consideration of modality and mood in English. *Foundations of Language*, 322-361.
- Janssens, K., & Nuyts, J. (2021). On the origins of the epistemic, evidential, and subjectivity meanings in the mental state predicates: The case of Dutch. *Jezikoslovlje*, 22(2), 227-250.
- Shank, C., Van Bogaert, J., & Plevoets, K. (2016). The diachronic development of zero complementation: A multifactorial analysis of the *that*/zero alternation with *think*, *suppose*, and *believe*. *Corpus Linguistics and Linguistic Theory*, 12(1), 31-72.
- Thompson, S. A., & Mulac, A. (1991a). A quantitative perspective on the grammaticization of epistemic parentheticals in English. In E. Traugott & B. Heine (eds) *Approaches to Grammaticalization* Vol.2, Amsterdam: John Benjamins Publishing, 313-329.
- Thompson, S. A., & Mulac, A. (1991b). The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of pragmatics*, 15(3), 237-251.
- Van Bogaert, J. (2011). *I think* and other complement-taking mental predicates: a case of and for constructional grammaticalization. *Linguistics*, 49(2), 295–332.

## Schematicity and productivity of synthetic compounds in English and Croatian

Frane Malenica

University of Zadar, Croatia

[fmalenica@unizd.hr](mailto:fmalenica@unizd.hr)

Synthetic compounds, such as *wine maker* and *taxi driver*, are generally accepted as an area of interaction between morphology and syntax. Their formation has been analysed in numerous studies, both from formal (*inter alia*, Roeper and Siegel 1978, Selkirk 1982, Lieber 1983, Grimshaw 1990, Oshita 1995, Ackema and Neeleman 2004) and functionalist perspectives (Booij 2010, Gaeta 2010, Gaeta & Zeldes 2017, Mattiello & Dressler 2022). One of the main assumptions of the Construction Morphology (CxM) paradigm is that the creation of new lexemes is licensed by word-level constructions known as morphological schemas (Booij 2010, 2015, 2018; Gaeta 2010, Gaeta and Zeldes 2017, Jackendoff and Audring 2016, 2018). Like their syntactic counterparts (Goldberg 1995 & 2006), these morphological constructions encompass the full range from fully lexically specified idiosyncratic constructions to fully schematic constructions (Booij 2010).

The aim of this paper is to analyse the interaction between productivity, measured as ratio of hapaxes and tokens (Baayen 1991, Baayen and Lieber 1991), and schematicity, measured as ratio of nominal complements to verbal head elements, in synthetic compounds in English and Croatian. These two languages provide an interesting point of comparison as nominal compounds are regarded as highly productive in Germanic languages and relatively less productive in Slavic languages (Kastovsky 2009) and this difference in degree of productivity should be reflected in the different degrees of schematicity of this word-formation pattern in the two languages.

The data analysed in this paper were collected within the scope of the doctoral dissertation (Malenica 2021) which utilized the data from two corpora based on newspaper articles. The data for English were collected from the Daily Mail subcorpus of the SiBol corpus of English broadsheets (cca. 23 million tokens) and it included synthetic compounds created with the *-er* suffix, such as *dog breeder* and *deer hunter*. The data for Croatian were collected from the Večernji list subcorpus of the HrWac corpus (cca. 49 million tokens), and it included compounds created with the 4 suffixes with the highest number of tokens: *-ac* (*najmodavac* 'lease giver, landlord'), *-nik* (*zemljoposjednik* 'land-owner'), *-telj* (*brodograditelj* 'ship builder') and *-ø* (*kišobran* 'rain protector, umbrella').

The results of the analysis show that English synthetic compounds are markedly more productive than the analysed synthetic compounds in Croatian and this difference in productivity is reflected in higher degree of schematicity. Based on the collected data, we can argue that English compounds are distributed along the whole continuum from fully lexically specified to fully schematic constructions, while the Croatian compounds include only fully and partially specified patterns, but seem to lack the fully schematic part of the continuum.

## References

- Ackema, P., & Neeleman, A. (2004). *Beyond morphology: Interface conditions on word formation*. Oxford University Press.
- Baayen, H. (1991). Quantitative aspects of morphological productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1991* (pp. 109–149). Springer Netherlands. [https://doi.org/10.1007/978-94-011-2516-1\\_8](https://doi.org/10.1007/978-94-011-2516-1_8)
- Baayen, H., & Lieber, R. (1991). Productivity and English derivation: A corpus-based study. *Linguistics*, 29(5). <https://doi.org/10.1515/ling.1991.29.5.801>
- Booij, G. (2010). *Construction Morphology* (3rd ed.). Oxford University Press.
- Booij, G. (2015). 12. Word-formation in construction grammar. In P. O. Müller, I. Ohnheiser, S. Olsen, & F. Rainer (Eds.), *Word-Formation* (pp. 188–202). DE GRUYTER. <https://doi.org/10.1515/9783110246254-014>
- Booij, G. (2018). The Construction of Words: Introduction and Overview. In G. Booij (Ed.), *The Construction of Words* (Vol. 4, pp. 3–16). Springer International Publishing. [https://doi.org/10.1007/978-3-319-74394-3\\_1](https://doi.org/10.1007/978-3-319-74394-3_1)
- Gaeta, L. (2010). Synthetic compounds. *Cross-Disciplinary Issues in Compounding, Amsterdam: John Benjamins*, 219–236.
- Gaeta, L., & Zeldes, A. (2017). Between VP and NN: On the constructional types of German -er compounds. *Constructions and Frames*, 9(1), 1–40. <https://doi.org/10.1075/cf.9.1.01gae>
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Grimshaw, J. B. (1990). *Argument structure*. MIT Press.
- Jackendoff, R., & Audring, J. (2016). Morphological schemas: Theoretical and psycholinguistic issues. *The Mental Lexicon*, 11(3), 467–493. <https://doi.org/10.1075/ml.11.3.06jac>
- Jackendoff, R., & Audring, J. (2018). Relational Morphology in the Parallel Architecture. In J. Audring & F. Masini (Eds.), *The Oxford Handbook of Morphological Theory* (pp. 389–408). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199668984.013.33>
- Kastovsky, D. (2011). *Diachronic Perspectives*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199695720.013.0016>
- Lieber, R. (1983). Argument Linking and Compounds in English. *Linguistic Inquiry*, 14(2), 251–285.
- Malenica, F. (2021). *Sintetske složenice i tematske uloge u engleskom i hrvatskom jeziku (Synthetic Compounds and Thematic Roles in English and Croatian)* [Unpublished doctoral dissertation]. University of Zagreb.
- Mattiello, E., & Dressler, W. U. (2022). Dualism and superposition in the analysis of English synthetic compounds ending in -er. *Linguistics*, 60(2), 395–461. <https://doi.org/10.1515/ling-2021-0235>
- Oshita, H. (1995). Compounds: A view from Suffixation and A-Structure Alteration. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1994*. Springer Netherlands. <https://doi.org/10.1007/978-94-017-3714-2>
- Roeper, T., & Siegel, M. E. (1978). A lexical transformation for verbal compounds. *Linguistic Inquiry*, 199–260.

## Deviation patterns of multi-word verbs in the writings of Turkish-speaking EFL learners

Seval Özen

Germany

[oezen.seval@gmail.com](mailto:oezen.seval@gmail.com)

Multi-word verbs (MWVs), which remain a difficult aspect of English vocabulary for L2 learners, regardless of their proficiency and L1 background, are the focus of this doctoral study. It provides a comprehensive insight into Turkish learners' usage of four different categories in their writings (Özen 2021); namely, phrasal, prepositional, and phrasal-prepositional verbs, as well as verb-noun collocations. The Turkish sub-corpus (TICLE) of the *International Corpus of Learner English* (ICLE) was the basis for the investigation. The aim was to explore how the typological distance resulting in the structural differences in the verb system of Turkish and English affects the use of different MWV categories by Turkish-speaking learners. The lack of structurally equivalent MWVs was believed to pose this learner group difficulties.

In order to assure that all occurring MWVs in TICLE were accounted for, no pre-established list for the process of extraction was used. The indivisible unit of thought, i.e. lexical unity, was taken as the most essential criterion in the definition of MWVs. The only criterion in the extraction of verb-noun collocations was that noun was abstract in nature. The aspect of idiomaticity or transparency was not considered as a defining criterion. The extraction of prepositional verbs and verb-noun collocations was done fully manually whereas a semi-automatic approach was opted for in the extraction of phrasal and phrasal-prepositional verbs. The data extraction yielded 6,129 tokens. The acceptability of 'unrecorded' and 'inappropriate' MWVs has been evaluated by four NSs of different English varieties.

In this talk, the focus will be on common error types attested in the data and their possible explanations. On the basis of the findings it can be concluded that all four categories of MWVs investigated are a major hurdle for Turkish learners. Their MWV use displays deviation from the NS norms to a great extent. The findings converge with previous studies that showed that EFL learners lack sufficient collocational knowledge, displaying a high reliance on the 'open choice principle' (Sinclair 1991) in the use of MWVs, and that they show a tendency to rely on high-frequency, 'easy', verbs that can be employed in a large number of contexts rather than more specific ones in collocations (Lennon 1996).

With regard to the deviation types, the broad patterns of the results align well with earlier studies conducted on MWV use by ESL and EFL learners with other L1 backgrounds. The deviations attested fall into four major categories:

- (a) omission of a necessary constituent of a MWV (e.g., omission of prepositions, particles, determiners),
- (b) use of additional constituents (e.g. redundant uses of prepositions, particles),
- (c) substitution errors, and
- (d) unnatural (i.e. unidiomatic) uses.

Although L1 interference (evaluated by the existence of incongruity between the learners' L1 and L2) revealed itself noticeably in nearly all deviation types, the learners also displayed some degree of 'creativity' in the formation of 'new' aspectual phrasal verbs and prepositional verbs,

following similar patterns observed in other EFL and ESL varieties (e.g., Mukherjee 2009; Nesselhauf 2009; Edwards & Laporte 2015; Gilquin & Granger 2011; Schneider & Zipp 2013; Schneider & Gilquin 2018; Gilquin 2011; Mondor 2008). In addition, other, more subtle factors were found to underlie the inappropriate and unidiomatic use of MWVs in the data. These and a few tentative conclusions with regard to the factors contributing to the difficulty of using MWVs will be discussed.

## References

- Edwards, Alison and Samantha Laporte. 2015. Outer and expanding circle Englishes: The competing roles of norm orientation and proficiency levels. *English World-Wide* 36(2): 135-169.
- Gilquin, Gaëtanelle. 2011. 'Corpus linguistics to bridge the gap between World Englishes and Learner Englishes'. In Leonel Ruiz Miyares and María Rosa Álvarez Silva (eds.), *Comunicación social en el siglo XXI, Vol. II* (pp. 638–642). Santiago de Cuba: Centro de Lingüística Aplicada.
- Gilquin, Gaëtanelle and Sylviane Granger. 2011. 'From EFL to ESL: Evidence from the *International Corpus of Learner English*'. In Joybrato Mukherjee and Marianne Hundt (eds.), *Exploring second-language varieties of learner Englishes. Bridging a paradigm gap* (pp. 55-78). Amsterdam/Philadelphia: John Benjamins.
- Lennon, Paul. 1996. Getting 'easy' verbs wrong at the advanced level. *International Review of Applied Linguistics in Language Teaching* 34(1): 23-36.
- Mondor, Monika. 2008. *Figuring it out. A corpus-based comparison of the verb-particle construction in argumentative writing by Swedish advanced learners and native speakers of English*. Doctoral dissertation. Gothenburg: University of Gothenburg.
- Nesselhauf, Nadja. 2009. Co-selection phenomena across New Englishes: Parallels (and differences) to foreign learner varieties. *English World-Wide* 30(1): 1-25.
- Özen, Seval. 2021. *Multi-Word Verb Usage by Turkish Learners of English: A Corpus-Based Study*. Book Series "New Ideas in Human Interaction (NIHIN)". Freiburg: Rombach. Full text: <https://doi.org/10.6094/UNIFR/223433>
- Mukherjee, Joybrato. 2009. 'The lexicogrammar of present-day Indian English: Corpus-based perspectives on structural nativisation'. In Ute Römer and Rainer Schulze (eds.), *Exploring the lexis-grammar interface* (pp. 117-135). Amsterdam & Philadelphia: John Benjamins.
- Schneider, Gerold and Gaëtanelle Gilquin. 2018. 'Detecting innovations in a parsed corpus of learner English'. In Sandra C. Deshors, Sandra Götz and Samanantha Laporte (eds.), *Rethinking linguistic creativity in non-native Englishes* (pp. 47-74). Amsterdam: John Benjamins.
- Schneider, Gerold and Lena Zipp. 2013. Discovering new verb-preposition combinations in New Englishes. *Studies in Variation, Contacts and Change in English*, 13:online. <http://doi.org/10.5167/uzh-82328>
- Sinclair, John McHardy. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

## Acknowledging limitations in PhD theses across disciplines: A phraseological approach

Shuyi Amelia Sun & Kevin Jiang

School of Foreign Language Education, Jilin University

[shuyi.amelia.sun@uq.net.au](mailto:shuyi.amelia.sun@uq.net.au); [kevinjiang@jlu.edu.cn](mailto:kevinjiang@jlu.edu.cn)

With the burgeoning research interest in English for Specific Academic Purposes (ESAP), PhD theses alongside their disciplinary rhetorical structures and linguistic resources have received an upsurge of scholarly attention (Paltridge & Starfield, 2020; Sun & Crosthwaite, 2022a; Xiao & Sun, 2020). As an underexplored “high-stakes part-genre” (p. 2), the ‘limitations’ section is considered crucial yet challenging, especially for novice thesis writers (Sun & Crosthwaite, 2022b). Composing a quality ‘limitations’ section not only manifests the writer’s capacity to critically reflect on the research but also contributes to a claim of expertise in a certain discipline (Zhou & Jiang, 2023). Nevertheless, due to insufficient pedagogical instructions, novice academic writers are usually unclear about the appropriate rhetorical ways to present limitations while minimizing the risk of losing validity (Glasman-Deal, 2010).

To date, research on ‘limitations’ sections has examined their interpersonal strategies (Sun & Crosthwaite, 2022b; Zhou & Jiang, 2023), or in most cases, researchers tended to draw on a structural view to explore how the ‘limitations’ section functions as a constituent of macro part-genres (Cotos et al., 2015; Paltridge & Starfield, 2020; Sheldon, 2019). By comparison, little is known about the articulation of ‘limitations’ sections from a phraseological view, and whether any variations with respect to the use of phraseology in acknowledging limitations lie across disciplines. This is despite the fact that phraseology as a key lexicogrammatical phenomenon has been found to play a fundamental role in language acquisition, processing, and pedagogy (Lu et al., 2021).

Accordingly, this study seeks to remedy the oversight by investigating the use of four-word lexical bundles in the ‘limitations’ sections of PhD theses across hard-applied, hard-pure, soft-applied, and soft-pure disciplines, aiming to address the following research questions.

- (1) What are the frequencies, structures, and functions of four-word lexical bundles in the limitations discussed by thesis writers from different disciplinary fields?
- (2) To what extent does the above featured use of bundles vary across disciplines?

In the pilot study, the authors constructed a corpus of 120 ‘limitations’ sections, incorporating 30 texts from each disciplinary field, while ensuring the comparability, representativeness and sampling of the specialized corpus designed to answer our research questions. We used *AntConc* (Anthony, 2022) to search for four-word bundles following Hyland’s (2008) criterion of strings occurring in at least 10% of texts and then manually coded each instance for its grammatical structure and function (Hyland & Jiang, 2018). Cross-disciplinary comparisons were finally conducted using Log-Likelihood tests (Rayson, 2016).

According to our results, at the structural level, clause-related and noun/preposition-related bundles were more commonly witnessed compared with verb phrase-related bundles, while there was a balanced distribution of research-oriented, text-oriented and participant-oriented bundles at the functional level. Our cross-disciplinary comparisons further showed that soft

disciplines adopted notably more four-word bundles at both structural and functional levels while acknowledging limitations. We close by discussing how the findings can inform ESAP pedagogy and thereby develop novice academic writers' lexicogrammatical awareness by helping them understand the typical ways to acknowledge research limitations by their disciplinary community members.

## Acknowledgements

This work was supported by the National Social Science Foundation of China (21AYY013) and the Graduate Innovation Fund of Jilin University.

## References

- Anthony, L. (2022). AntConc (Version 4.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Cotos, E., Huffman, S., & Link, S. (2015). Furthering and applying move/step constructs: Technology-driven marshalling of Swalesian genre theory for EAP pedagogy. *Journal of English for Academic Purposes*, 19, 52–72. <https://doi.org/10.1016/j.jeap.2015.05.004>
- Glasman-Deal, H. (2010). *Science research writing for non-native speakers of English*. Imperial College Press.
- Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41–62. <https://doi.org/10.1111/j.1473-4192.2008.00178.x>
- Hyland, K., & Jiang, F. (2018). Academic lexical bundles. *International Journal of Corpus Linguistics*, 23(4), 383–407. <https://doi.org/10.1075/ijcl.17080.hyl>
- Lu, X., Yoon, J., & Kisselev, O. (2021). Matching phrase-frames to rhetorical moves in social science research article introductions. *English for Specific Purposes*, 61, 63–83. <https://doi.org/10.1016/j.esp.2020.10.001>
- Paltridge, B., & Starfield, S. (2020). *Thesis and dissertation writing in a second language* (2nd ed.). Routledge.
- Rayson, P. (2016). Log-likelihood and effect size calculator [Computer Software]. Available from: <http://ucrel.lancs.ac.uk/llwizard.html>
- Sheldon, D. E. (2019). Knowledge construction of discussion/conclusion sections of research articles written by English L1 and L2 and Castilian Spanish L1 writers. *Journal of English for Academic Purposes*, 37, 1–10. <https://doi.org/10.1016/j.jeap.2018.11.002>
- Sun, S., & Crosthwaite, P. (2022a). “Establish a niche” via negation: A corpus-based study of negation within the move 2 sections of PhD thesis introductions. *Open Linguistics*, 8(1), 189–208. <https://doi.org/10.1515/opli-2022-0190>
- Sun, S., & Crosthwaite, P. (2022b). “The findings might not be generalizable”: Investigating negation in the limitations sections of PhD theses across disciplines. *Journal of English for Academic Purposes*, 59, 101155. <https://doi.org/10.1016/j.jeap.2022.101155>
- Xiao, W., & Sun, S. (2020). Dynamic lexical features of PhD theses across disciplines: A text mining approach. *Journal of Quantitative Linguistics*, 27(2), 114–133. <https://doi.org/10.1080/09296174.2018.1531618>
- Zhou, H., & Jiang, F. K. (2023). ‘The study has clear limitations’: Presentation of limitations in conclusion sections of PhD dissertations and research articles in applied linguistics. *English for Specific Purposes*, 71, 34–47. <https://doi.org/10.1016/j.esp.2023.02.001>

## Exploring phrase frames across rhetorical functions in workplace request emails

Detong Xia

Southeast University, China

[xiadg@mail.uc.edu](mailto:xiadg@mail.uc.edu)

As an important medium for organizational communication, email is extensively used in professional communities. Research has shown that it may be challenging for non-native speakers to write effective request emails in the workplace, which requires not only proficient language skills but also the ability to use linguistic functions appropriately in a given situational context (Evans, 2014; Xia et al., 2022). To address learners' difficulties of incorporating linguistic knowledge into their writing in functionally appropriate ways, a growing body of research has integrated both genre analysis (of rhetorical functions) and corpus analysis (of linguistic features) in L2 writing research (e.g., Lu et al., 2021; Yoon & Casal, 2020). This study seeks to add to this growing body of literature by investigating phrase frames (or p-frames), i.e., a form of recurrent multi-word sequence with one variable slot (e.g., *as \* as the [far, soon, well]*), in terms of the rhetorical goals that working professionals employ them to realize in request emails.

Two research questions were addressed in this paper:

- (1) How are p-frames distributed across rhetorical functions in a corpus of workplace request emails?
- (2) What are the linguistic characteristics of the identified p-frames in the request emails?

A workplace corpus of authentic request email messages derived from the Avocado Research Email Collection (Oard et al., 2015) was used in this study (total number of emails = 1148; total number of words = 125,471, *M* length = 109.49, *SD* = 54.81). Two coders independently annotated all emails based on the revised version of Park et al.'s (2021) rhetorical move framework of request emails. A rhetorical move is characterized as a rhetorical unit of a text that performs a specific communicative purpose (Swales, 1990). P-frames of five-word length were extracted based on the frequency threshold of 6 (roughly corresponding to 43 per million words) and the range threshold of 5 email texts using *kfNgram* (Fletcher, 2012). I wrote a script in Python 2.7.16 to automatically map a phrase frame to one of the rhetorical moves in which they occurred.

For the first research question, I calculated the number of emails containing each rhetorical move as well as the type and token of p-frames occurring in each rhetorical move. The association strength between each p-frame and the move was also calculated by dividing the number of times a p-frame occurred in its primary rhetorical move by the total occurrences of the p-frame. For the second research question, I used the Cross-Cultural Speech Act Realization Project (CCSARP; Blum-Kulka et al., 1989) framework to conduct both quantitative and qualitative analysis of linguistic characteristics embedded in p-frames. CCSARP framework provides guidelines to investigate linguistic devices for making requests at both syntactic and lexical levels.

Results showed that the rhetorical moves in request emails relied on p-frames, yet the degree of such reliance differed across the moves. The moves of *making the inquiry* and *closing* had the largest number of p-frames solely associated with them, which indicated that working professionals used highly formulaic language to pose their inquiries or to signal the ending of an

email. In addition, I found a wide range of syntactic and lexical mitigation devices embedded in p-frames, especially in the move of *making the inquiry*, including politeness makers (e.g., *could you please \* me*), embedded *if*-clauses (e.g., *let \* know if this*), progressive aspect (e.g., *I was wondering if \**), downtoners (e.g., *is it possible to \**), and hedges (e.g., *do you have any \**). Such findings suggested that highly frequent multi-word sequences were useful linguistic devices to display genre-specific characteristics of request emails. The discussion will include some pedagogical points for English for Business Purposes curricula.

## References

- Blum-Kulka, S., House, J., & Kasper, G. (1989). *Cross-cultural pragmatics: Request and apologies*. Norwood, NJ: Ablex.
- Casal, J. E., & Kessler, M. (2020). Form and rhetorical function of phrase-frames in promotional writing: A corpus- and genre-based analysis. *System*, 95(December).
- Evans, S. (2014). Teaching business correspondence: lessons from the globalised workplace. *The Asian Journal of Applied Linguistics*, 1(2), 102–120.
- Fletcher, W. (2012). *KfNgram*. Annapolis, MD: USNA.
- Lu, X., Casal, J. E., & Liu, Y. (2021). Towards the synergy of genre- and corpus-based approaches to academic writing research and pedagogy. *International Journal of Computer-Assisted Language Learning and Teaching*, 11(1), 59–71.
- Oard, D., Webber, W., Kirsch, D., Golitsynskiy, S. (2015). *Avocado research email collection*. Philadelphia: Linguistic Data Consortium.
- Park, S., Jeon, J., & Shim, E. (2021). Exploring request emails in English for business purposes: A move analysis. *English for Specific Purposes*, 63, 137–150.
- Swales, John M. 2004. *Research genres: Explorations and applications*. New York: Cambridge University Press.
- Xia, D., Ai, H., & Pae, H. K. (2022). “Please let me know”: Lexical bundles in business emails by business English learners and working professionals. *International Journal of Learner Corpus Research*, 8(1), 1–30.
- Yoon, J., & Casal, J. E. (2020). Rhetorical structure, sequence, and variation: A step-driven move analysis of applied linguistics conference abstracts. *International Journal of Applied Linguistics (United Kingdom)*, 30(3), 462–478.
- Zhu, W. (2012). Polite requestive strategies in emails: An investigation of pragmatic competence of Chinese EFL learners. *RELC Journal*, 43(2), 217–238.

## Factors Influencing the Errors of Number Marking in English Nouns: A Study of Chinese Learners of English

Dongchen Yao

The University of Queensland, australia

[dongchenyao628@gmail.com](mailto:dongchenyao628@gmail.com)

The study explores the suppliance of English number marking by intermediate Chinese learners of English, using the *International Corpus Network of Asian English* (ICNALE). Previous studies have acknowledged the challenge of number marking for learners from classifier languages. Several factors have been identified as potential contributors to number marking errors in English nouns, including the count/mass distinction, noun classes, atomicity, determiners, and L1 transfer (Hiki, 1990; Choi & Ionin, 2021; Yoon, 1993). However, these studies have been experimental in nature. Research investigating L2 learners' natural or near-natural production of English number marking has been limited. The current study aimed to fill in this gap by focusing on some of the factors established by previous research namely the count/mass distinction, atomicity, and determiners. The research question addressed in the study is: to what extent do these factors contribute to L1-Chinese learners' erroneous production of English number marking?

Employing mixed-effects binomial logistic regression analysis, it explored the factors that contribute to errors in English number marking by Chinese learners. 2,000 concordance lines containing erroneous number marking were extracted from the Chinese section of the Spoken Monologue (SM) component of the *International Corpus Network of Asian Learners of English* (ICNALE) as compiled by Ishikawa (2014). The data processing and analysis were run in the programming environment R (R CRAN team, 2022). The data were first part-of-speech tagged using the *udpipe* R package (Wijffels, 2021). Then, 2,000 concordance lines of nouns that were based on the pos-tags NN and NNS were extracted using the *kwic* function from the *quanteda* package. The variables were annotated, and the regression model was fitted using the *glmuti* function in R.

The results revealed a correlation between the count/mass distinction and the use of determiners with errors in number marking. Specifically, the findings indicate that Chinese learners made more number marking errors with count nouns than with mass nouns, particularly with singular count nouns. Additionally, the findings show that the omission of determiners is a significant predictor of errors in number marking, indicating a higher likelihood of such errors in the absence of determiners than in their presence, regardless of whether the nouns are countable or uncountable. These errors encompassed the erroneous usage of plural markers (e.g., *\*childs*, *\*peoples*) as well as the absence of determiners preceding the bare form of a noun (e.g., *\*in restaurant*).

Applying the Complex Adaptive System Principles (CASP) model (Filipović & Hawkins, 2013), the current study attempts to account for these findings. This usage-based model views language as a dynamic system shaped by both internal and external factors, and specifically emphasizes the interplay of multiple linguistic factors, such as complexity, frequency and L1 transfer. For instance, errors on mass nouns were minimal due to their structural simplicity, whereas there were more errors on count nouns because these nouns have to be morpho-syntactically marked for number. Implications for language pedagogy are discussed.

## References

- Choi, S. H., & Ionin, T. (2021). Plural marking in the second language: Atomicity, definiteness, and transfer. *Applied Psycholinguistics*, 42(3), 549-578.  
<https://doi.org/10.1017/S0142716420000569>
- Filipović, L. & Hawkins, J. (2013). Multiple factors in second language acquisition: The CASP model. *Linguistics*, 51(1), 145-176.
- Ishikawa, S. (2014). Design of the ICNALE Spoken: A new database for multi-modal contrastive interlanguage analysis. *Learner Corpus Studies in Asia and the World*, 2, 63-76.
- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Wijffels J (2022). \_udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit\_. R package version 0.8.9,  
<https://cran.rproject.org/web/packages/udpipe/index.html>
- Yoon, K. Y. (1993). Challenging prototype descriptions: Perception of noun countability and indefinite vs. zero article use. *International Review of Applied Linguistics*, 31, 269–289.