



# **7<sup>th</sup> Symposium on Corpus Approaches to Lexicogrammar (LxGr2022)**

8-9 July 2022

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

## **ARCHIVE**

**FOCUS AND FORMAT**

**COMMITTEE**

**PROGRAMME**

**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)

## LxGr2022 PROGRAMME COMMITTEE

[Bogdan Babych](#) (Heidelberg University)

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

[Belen Diaz-Bedmar](#) (University of Jaén)

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

[Lise Fontaine](#) (Cardiff University)

[Gaëtanelle Gilquin](#) (Université Catholique de Louvain)

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

[Andrew Hardie](#) (Lancaster University)

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

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

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

[Geraldine Mark](#) (University of Cambridge)

[Gabriel Ozon](#) (University of Sheffield)

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

[Magali Paquot](#) (Université Catholique de Louvain)

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

[Paul Rayson](#) (Lancaster University)

[Ute Römer](#) (Georgia State University)

[Serge Sharoff](#) (University of Leeds)

[Benet Vincent](#) (Coventry University)

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

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

## **LxGr2022 PROGRAMME**

## DAY 1: Friday 8 July 2022

9:00 am – 9:30 am WELCOME

9:30 am – 10:30 am GUEST SPEAKER

**Magali PAQUOT** (Université catholique de Louvain, Belgium)

*Investigating phraseological complexity in learner language: From collocations to collocations and other lexicogrammatical structures*

10:30 am – 11:00 am BREAK / CHAT

11:00 am – 11:35 am

**Paulina VALENZUELA-LAGOS & Sy-Ying LEE** (National Taiwan University of Science and Technology, Taiwan)

*The applicability of reading illustrated books with learners of Spanish as a foreign language: A corpus analysis*

11:40 am – 12:15 pm

**Tan GEDIK** (Erlangen-Nürnberg University, Germany)

*Creating lexicogrammatical language materials: A practical guide for foreign language teachers*

12:20 pm – 12:40 pm

**Tomasz MICHTA** (University of Białystok, Poland)

*A dictionary of legal collocations: Looking back and looking forward [WIP]*

12:45 pm – 1:15 pm BREAK / CHAT

1:15 pm – 1:50 pm [CANCELLED]

**Leila GHOLAMI** (Arizona State University, USA)

*Formulaic vs. non-formulaic constructions in second language writing: A text-analytic learner corpus research*

1:55 pm – 2:15 pm

**Sangeun KIM** (University of Exeter, UK)

*Do stance constructions become more formulaic in L2 academic writing development?*

2:20 pm – 2:40 pm

**Mirjam WEDER** (University of Basel, Switzerland)

*How formulaic sequences link to larger rhetorical patterns: The case of the German formulae “Fakt ist” and “Tatsache ist” [WIP]*

2:45 pm – 3:15 pm BREAK / CHAT

3:15 pm – 3:35 pm

**Yating TAO** (Université catholique de Louvain, Belgium)

*Valency patterns of 'take' in Hong Kong English [WIP]*

3:40 pm – 4:15 pm

**David CALLAGHAN** (University of Birmingham, UK)

*"I'm beginning to wonder what it's for": A corpus-based investigation into lexicogrammatical incongruency and its relation to irony.*

4:20 pm – 4:55pm

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

*A corpus study of headless compound chi-zao in Chinese: A constructionalization approach*

5:00pm – 5:30 pm DAY 1 WRAP-UP & DISCUSSION

## DAY 2: Saturday 9 July 2022

9:30 am – 10:00 am WELCOME

10:00 am – 10.35 am [CANCELLED]

**Sugene KIM** (Nagoya University of Commerce & Business, Japan)

*A corpus study of the linking adverbial 'beside'*

10:40 am – 11:15 am

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

*The other grammarian's dream: Constructions as the most delicate grammar*

11:15 am – 11.45 am BREAK / CHAT

11:45 am – 12:20 pm

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

*The grammatical organization of synonymy: Evidence of nestedness in a paradigm of schematic constructions*

12:25 pm – 1:00 pm

**Laura ESTEBAN-SEGURA & Vanessa DÍAZ-GONZÁLEZ** (Universidad de Málaga, Spain)

*A study of the understood reflexive object alternation in early modern English*

1:00 pm – 1:30 pm BREAK / CHAT

1:30 pm – 2:05 pm

**Aysel Şahin KIZIL** (İzmir Bakırçay University, Turkey), **Lee MCCALLUM & Benet VINCENT** (Coventry University, UK)

*Adopting a phrase-frame approach to investigate a corpus of research article introductions: Methodological issues*

2:10 pm – 2:45 pm

**Flavio PISCIOTTA** (University of Bologna, Italy)

*Not all subjects are “raised” equally: The alternation between the constructions with sembrare (‘seem’) in written and spoken Italian*

2:45 pm – 3:15 pm    BREAK / CHAT

3:15 pm – 3:50 pm

**Aleksandra MILOSAVLJEVIĆ** (Institute for the Serbian Language of SASA, Serbia)

*Concessive conjunctions in Serbian: Subordinators and/or coordinators?*

3:55 pm – 4:30 pm

**Marta PETRAK** (University of Zagreb, Croatia)

*Word-formation as a complex product of lexico-grammatical relations: A study of Croatian nouns prefixed with nad- and pod-*

4:30 pm – 5:00 pm    DISCUSSION & FEEDBACK

## **ABSTRACTS**



***I'm beginning to wonder what it's for: A corpus-based investigation into lexicogrammatical incongruency and its relation to irony***

David Callaghan  
University of Birmingham, UK  
[DJC055@student.bham.ac.uk](mailto:DJC055@student.bham.ac.uk)

My research is a corpus-based investigation into the lexicogrammatical features of irony. Irony is a well-established area of linguistic study, heavily influenced by the cooperative principal theory of Grice (1975). Subsequently, much linguistic theory of irony is concerned with its pragmatic nature or focuses on its paralinguistic features (Wilson and Sperber 1992, Attardo 2000). A common understanding of irony is of a trope in which the dictum and the implicatum are seen as incongruous (Kotthoff 2003). I argue that patterns of lexicogrammatical incongruency can reflect this incongruency at the pragmatic level. Additionally, a bottom-up examination of genuine examples of ironic utterances can reveal common lexicogrammatical patterns. In this presentation, I will present two notable patterns: collocational patterns of hedging items; and collostructions (Stefanowitsch and Gries 2003) involving the progressive aspect. This study attempts to readdress the paucity of linguistic studies into irony which focus on real-world examples of irony as a source of data.

The patterns presented in this study are therefore supported by examples of irony taken from two irony-rich discourse environments: Prime Minister Questions sessions from British Parliament and White House Press Briefings from the US presidential office. These environments were the source for the creation of two political corpora (PMQ and WHPB). Following this, real-world examples of irony were extracted from the two corpora using an independent framework of irony influenced by both Hutcheon (1994) and Burgers et al (2011). Finally, using the Sketch Engine software, commonalities of patterning were first identified, and then interrogated across the two DIY corpora, as well as the BNC and enTenTen15 general corpora, in order to measure both frequency (raw/t-score) and fixedness. Finally, a deeper examination of the concordance line revealed whether such patterns carry an ironic force.

Two significant findings will be presented. Firstly, the study explores lexicogrammatical patterns of collocation concerning *multiple hedging*: that is, two or more lexical items which ostensibly have a hedging function, yet often frame strong evaluative or rhetorical statements. Secondly, influenced by both construction grammar theory (Goldberg, 1995; Jackendoff, 1997) and corpus-based investigations into the progressive aspect (Mindt, 2000; Römer, 2005), I present patterns of collostruction in which the progressive aspect colligates with cognition verbs. Again, such phrases incongruently frame strong rhetoric. It is the lexicogrammatical incongruency within these patterns that is often a source of irony.

Usage of these phrases does not, however, guarantee that the statement will always be ironic. Yet, when compared within larger general corpora, these patterns do demonstrate high tendencies of pragmatic characteristics related to irony. Therefore, I adopt Hoey's (2005) terms by arguing that such patterns can be considered as having ironic *priming*.

Identification and awareness of such patterns may help audiences in accurately reaching ironic interpretations. More practically, these patterns may also help NLP methodology by building upon previous attempts of automated irony detection (Hao and Veale, 2010; Fariás et al, 2016) to create more robust algorithms. Furthermore, there are wider implications to what corpus

linguistic methodology can explore in regard to connections between pragmatics and lexicogrammar.

## References

- Attardo, S., 2000. Irony as relevant inappropriateness. *Journal of pragmatics*, 32(6), pp.793-826.
- Burgers, C., van Mulken, M., & Schellens, P.J., 2011. Finding irony: An introduction of the Verbal Irony Procedure (VIP). *Metaphor and Symbol*, 26(3), 186-205.
- Farías, D.I.H., Patti, V. and Rosso, P., 2016. Irony detection in twitter: The role of affective content. *ACM Transactions on Internet Technology (TOIT)*, 16(3), pp.1-24.
- Grice, P., 1975. Logic and conservation. Reprinted in *Studies in the Way of Words*. Harvard University Press (1989). Retrieved from <http://grammar.about.com/od/c/g/coopeartiveprincipleterm.htm>
- Hao, Y. and Veale, T., 2010. An ironic fist in a velvet glove: Creative mis-representation in the construction of ironic similes. *Minds and Machines*, 20(4), pp.635-650.
- Hoey, M., 2005. *Lexical priming: A new theory of words and language*. London: Routledge.
- Hutcheon, L., 1994. *Irony's edge: The theory and politics of irony*. London: Routledge
- Kotthoff, H., 2003. Responding to irony in different contexts: On cognition in conversation. *Journal of pragmatics*, 35(9), pp.1387-1411.
- Mindt, D., 2000. *An empirical grammar of the English verb system*. Berlin: Cornelsen
- Römer, U., 2005. *Progressives, patterns, pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics*. John Benjamins Publishing.
- Wilson, D. and Sperber, D., 1992. On verbal irony. *Lingua* 87:53-76.

## A study of the understood reflexive object alternation in Early Modern English

Laura Esteban-Segura & Vanessa Díaz-González

Universidad de Málaga, Spain

[lesteban@uma.es](mailto:lesteban@uma.es)

The Understood Reflexive Object Alternation is a particular type of verb alternation in which the reflexive object may appear either marked (*xself*) or unmarked ( $\emptyset$ ) in cases of “self-directed situations” (König & Gast, 2002, p. 6), as in the following examples: (i) *John dressed himself* vs (ii) *John dressed*. While (ii) entails that “the action is understood to be directed towards the subject of the verb” (Levin, 1993, p. 228), this coreference is also projected syntactically in (i). From a historical viewpoint, the proliferation of both constructions was particularly active in the seventeenth century, even though their behaviour differs across verbs because of their different polysyllabic dimension (Baugh & Cable, 2002, p. 244).

The set of verbs that allow this alternation is restricted to some semantic fields, e.g. verbs of caring for the whole body (*dress, shave*) and load verbs, among others (Levin, 1993, p. 36), being the marked variant the predominant one in terms of frequency in Present-Day English. Likewise, locational verbs (*keep, move*) also allow for both constructions, although they bear slightly syntactic and semantic changes (Fawcett, 1987). Thus, the marked variant is in principle preferred whenever there is more emphasis on the resulting event, while the unmarked one suggests a routine action seen as a unitary event (Wierzbicka, 1996, pp. 422-423).

Taking all this into consideration, the present study investigates the role of syntax and semantics in cases of self-directed situations, focusing on both the *xself* (marked) and  $\emptyset$  (unmarked) constructions in Early Modern English medical writing with the following objectives: (i) to analyse their distribution in the period 1470s-1690s; (ii) to investigate the phenomenon from the perspective of verb typology; and (iii) to evaluate the contribution of the following factors in the choice of one particular reflexive form: (a) the typology of the intervening subject; (b) the morphology of the verb; and (c) the linguistic complexity of the reflexive construction.

The data, which have been analysed from a quantitative and qualitative perspective, come from the corpus of *Early English Books Online* (Davies, 2017). A set of ten verbs has been selected for the study: *cut, dress, keep, load, move, pride, pull, shave, shower* and *soap* (Quirk et al., 1985; Fawcett, 1987; Levin, 1993). Wildcards were employed to retrieve all the reflexive structures, specifically 13 expressions for each verb, which amounted to a total of 13,332 instances, divided into 7,865 occurrences of the marked reflexive construction and 5,467 of the unmarked one.

As for the main findings, the unmarked form appears to decline sharply towards the mid-sixteenth century, which coincides with a significant diffusion of *xself* towards the late sixteenth century, particularly in combination with three verbs, i.e., *cut, dress* and *keep*, but following different time scales. The progressive rise of *xself* was to counterbalance the reduction of verbal inflections in English (Baugh & Cable, 2002).

Secondly, the data examined confirm that the phenomenon behaves differently across different types of verbs. Their polysyllabic dimension is also found to be a triggering factor in the choice of *xself* and  $\emptyset$ , i.e., monosyllabic verbs such as *cut, dress* or *keep* favour the use of *xself*, while  $\emptyset$  is mostly opted for in disyllabic verbs such as *load, shower* or *soap*. The concomitant use of both variants in these environments, in turn, is determined by different conceptualisations. Thus, while the marked form is in principle preferred whenever there is more emphasis on the

resulting event, the unmarked one suggests a routine action seen as a unitary event (Wierzbicka, 1996).

Thirdly, the typology of the subject, whether nominal or pronominal, seems to directly elicit the use of *xself* or  $\emptyset$ . Even though a balanced distribution of both subjects has been obtained, the marked form of the verbs *keep*, *load* and *move* is noted to diffuse more widely when there is a pronominal subject. On the contrary, the unmarked form of these three verbs is more likely to occur with a nominal subject. With the verb *pull*, the nominal subject noticeably prevails over the pronominal subject in the *xself*-construction.

Concerning the role of the morphology of the verb, in terms of person, the differences are not relevant as the third person predominates with both variants. With respect to the verbs *dress*, *pride*, *pull* and *shave*, the second person is more widespread when these verbs appear with the  $\emptyset$ -construction, thus distinguishing *you* for direct addressing and *thou*, for special registers, i.e., poetry, prayer, factual present.

Morphological differences as to number, on the other hand, present a different state of affairs, with a more widespread diffusion of the singular in the  $\emptyset$ -construction, while the plural is preferred in the *xself*-construction. This preference for the plural may be justified from an inflectional point of view as a result of the loss of the dual number in English.

Finally, with regard to linguistic complexity, both variants appear more frequently in main clauses. In the particular case of *keep-xself*, however, the subordinate clause is favoured, and this then complies with the *complexity principle* since “the more explicit variants are preferred in more complex environments” (Rohdenburg, 1996). *Cut- $\emptyset$*  and *pride- $\emptyset$* , on the other hand, also show a preference for the subordinate clause, which may be due to the widespread usage of the relative pronouns *who*, *which* and *when* in these environments.

## References

- Baugh, A. C., & Cable, T. (2002). *A History of the English Language* (5<sup>th</sup> ed.). London: Routledge.
- Davies, M. (2017). *Early English Books Online (EEBO)*. Available at: <https://corpus.byu.edu/eebo/>.
- Fawcett, R. P. (1987). The Semantics of Clause and Verb for Relational Processes in English. In Halliday, M. A. K., & Fawcett, R. P. (Eds.). *New Developments in Systemic Linguistics, Vol. 1: Theory and Description*. London: Pinter, pp. 131-183.
- König, E., & Gast, V. (2002). Reflexive Pronouns and Other Uses of *Self*-Forms in English. *Zeitschrift für Anglistik und Amerikanistik*, 50(3), 1-14.
- Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: The University of Chicago Press.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- Rohdenburg, G. (1996). Cognitive Complexity and Increased Grammatical Explicitness in English. *Cognitive Linguistics*, 7(2), 149-182.
- Wierzbicka, A. (1996). *Semantics: Primes and Universals*. Oxford: Oxford University Press.

**The grammatical organization of synonymy:  
Evidence of nestedness in a paradigm of schematic constructions**

Quentin Feltgen  
Ghent University, Belgium  
[quentin.feltgen@gmail.com](mailto:quentin.feltgen@gmail.com)

Language shows the striking property that it usually provides a wide variety of lexical means to express more or less the same idea – that is, synonymy is pervasive in the lexicon. Lexical series of words with closely related use and meaning, like the one led by *brave* and *courageous* (Cruse 2000, Moon 2013), are overly abundant. This property has stirred considerable efforts to disentangle the fine-grained semantic nuances between these synonyms (Rubenstein & Goodenough 1965, Edmonds & Hirst 2002).

Implicit behind these attempts is the idea that synonyms cover slightly different semantic niches – see (Rainer 2018) for an extensive overview of the concept. In this contribution, I would like to challenge this very idea by making use of another concept from Ecology, nestedness (Atmar & Patterson 1993, Mariani et al. 2019). Nestedness is a concept that applies to bipartite networks, that is, networks of inter-relations between two kind of species, such as flowers and pollinators. These networks could be of three major types: random, compartmentalized, that is, pollinators focus on sub-sets of flowers, the whole system being a collection of niches slightly overlapping with each other, or nested: generalists each pollinate a large range of flowers, while specialists pollinate small sub-sets of flowers, already pollinated by generalists – hence the nestedness. This structure happens to reinforce the biodiversity of the system (Bastolla et al. 2009).

That nestedness could readily be applied to language, and especially schematic constructions (that is, constructions featuring a free slot that can be filled with other words), has already been suggested (Petré 2014), but, to my knowledge, never empirically verified. To undertake this task, one must focus on a given type of constructions, that are both schematic (they offer a free slot) and organized into a paradigm – in a sense, these are constructions that can be described as two slots, each corresponding to a different level in the construction network. In this study, I rely on ‘obstruction’ constructions of the form [BE {V-ed} from] + {V-ing}, (e.g. *BE prevented from entering*). Based on the COCA corpus (Davies 2008), I have listed 33 such constructions, each with very similar meanings, and covering several hundreds of fillers in the {V-ing} slot. The constructions (that is, the choice of verb in the {V-ed} slot) play the role of pollinators, and the fillers that of the flowers. The nestedness of this bipartite network is then measured according to a freely available Python algorithm (Straka 2017) based on (Almeida-Neto et al. 2008). I finally show that the nestedness score, which measures the degree of nestedness of the system, significantly differ from a random distribution of the interactions between the two slots.

This result shows that synonymy at the constructional level does not result in constructions that specialize into fine-grained meanings (i.e. that allocate the fillers between themselves in a way to avoid overlap), but in constructions that are contrasted in their degree of specialization. This is consistent with previous observations on lexical synonyms, and with the major frequency differences between synonyms as well. It also leads to the intriguing observation that language is organized in such a way to optimize diversity.

## References

- Almeida-Neto, Mário, Paulo Guimaraes, Paulo R. Guimaraes Jr, Rafael D. Loyola, and Werner Ulrich. "A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement." *Oikos* 117, no. 8 (2008): 1227-1239.
- Atmar, Wirt, and Bruce D. Patterson. "The measure of order and disorder in the distribution of species in fragmented habitat." *Oecologia* 96, no. 3 (1993): 373-382.
- Bastolla, Ugo, Miguel A. Fortuna, Alberto Pascual-García, Antonio Ferrera, Bartolo Luque, and Jordi Bascompte. "The architecture of mutualistic networks minimizes competition and increases biodiversity." *Nature* 458, no. 7241 (2009): 1018-1020.
- Cruse, Alan. *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford University Press, 2000.
- Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>.
- Edmonds, Philip, and Graeme Hirst. "Near-synonymy and lexical choice." *Computational linguistics* 28, no. 2 (2002): 105-144.
- Hüning, Matthias. "Semantic niches and analogy in word formation: Evidence from contrastive linguistics." *Languages in contrast* 9, no. 2 (2009): 183-201.
- Mariani, Manuel Sebastian, Zhuo-Ming Ren, Jordi Bascompte, and Claudio Juan Tessone. "Nestedness in complex networks: observation, emergence, and implications." *Physics Reports* 813 (2019): 1-90.
- Moon, Rosamund. "Braving synonymy: from data to dictionary." *International Journal of Lexicography* 26, no. 3 (2013): 260-278.
- Petré, Peter. *Constructions and environments: Copular, passive, and related constructions in Old and Middle English*. Oxford Studies in the History, 2014.
- Rainer, Franz. "Patterns and niches in diachronic word formation: the fate of the suffix-men from Latin to Romance." *Morphology* 28, no. 4 (2018): 397-465.
- Rubenstein, Herbert, and John B. Goodenough. "Contextual correlates of synonymy." *Communications of the ACM* 8, no. 10 (1965): 627-633.
- Straka, Mika. <https://github.com/tsakim/nestedness.2017>.

## Creating lexicogrammatical language materials: A practical guide for foreign language teachers

Tan Gedik

Erlangen-Nürnberg, Germany

[tan.gedik@fau.de](mailto:tan.gedik@fau.de)

Construction grammar (CXG) has been taken up by various subfields of linguistics. Its application to language teaching, however, is relatively new and few (see Boas, 2022; Gilquin & DeKnop, 2016; DeKnop & Mollica, 2022; Hudson, 2022). Its application to producing teaching materials, especially for L2 English, has been scarce. While there are seemingly lexicogrammatically-informed textbooks (see for instance Berlitz, 2018), or corpus-based teaching materials (e.g., Le Foll, 2021; Reppen, 2021), pre- and in-service teachers generally do not learn these new theoretical developments and usage-based constructionist materials do not seem to exist. Therefore, this paper addresses the gap by subscribing to a usage-based constructionist approach to propose a practical guide for teachers to produce lexicogrammatically-informed and corpus-based English as a foreign language teaching materials. The input, i.e., the sentences used in the activities, is adopted from the corpus. The guide assumes teachers have used corpora at least once in their training.

The talk focuses on the teaching of the way-construction, based on the BNC, to illustrate the proposed steps (copies of the materials can be accessed [here](#)). The preliminary materials were developed for a flipped class based on Bloom's taxonomy, which is predominantly used in flipped teaching, and is a set of hierarchical models used for classification of educational learning objectives into levels of complexity and specificity. Based on the principles of pedagogical CXG (Herbst, 2016) and corpus-based material development (e.g., Bardovi-Harlig et al., 2015; Chang & Kuo, 2011), the steps to create lexicogrammatically-based English as a foreign language teaching materials are (1) select a construction, (2) identify form and meaning using the BNC using Sketch Engine - or use constructions if the construction selected is available, such as [constructicon.de](http://constructicon.de), (3) pick 200 randomized sentences and analyze the meaning in further detail, (4) pick a theme congruent with the meaning -if possible, (5) create objectives to teach the form-meaning pairing at once, (6) follow Bloom's taxonomy -or other lesson plan structures, e.g., present, practice, produce, (7) use the randomized sentences to create tasks that would showcase the usages of the construction, (8) create a collo-profile for the construction to give speakers an idea of the most prototypical examples of it, and (9) modify the ambient language input in tasks in line with the proficiency level. Thus, the concordance lines act as the basis for the activities and the input in the materials.

To validate the proposed steps here, 10 pre-service teachers and 10 in-service teachers were given a survey on their knowledge of, thoughts of a lexicogrammatical approach to language teaching and how the integration of the approach in the materials can prove beneficial for their English classes. The findings suggest that pre-service teachers are more likely to adopt new approaches to material development than in-service teachers. The survey also reflects that teachers' beliefs and classroom experiences about language acquisition reflect a lexicogrammatical approach, however, their training lacks such approaches.

As such, this paper establishes a practical guide for the creation of lexicogrammatically-informed, corpus-based teaching materials for pre- and in- service teachers. Academics in teacher training programs may also find this guide useful for their micro-teaching courses (e.g., [METU FLE304](#)). The practical guide with the information teachers might need on the central tenets of CXG is available [here](#).



## References

- Bardovi-Harlig, K., Mossman, S., & Vellenga, H. E. (2015). Developing Corpus-Based Materials to Teach Pragmatic Routines. *TESOL Journal*, 6(3), 499–526. <https://doi.org/10.1002/tesj.177>
- Berlitz. (2018). *Berlitz English 3*.
- Boas, H. C. (Ed.). (2022). *Directions for Pedagogical Construction Grammar: Learning and Teaching (with) Constructions*. De Gruyter. <https://doi.org/10.1515/9783110746723>
- Chang, C.-F., & Kuo, C.-H. (2011). A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes*, 30(3), 222–234. <https://doi.org/10.1016/j.esp.2011.04.001>
- De Knop, S., & Mollica, F. (2022). Construction-based teaching of German verbless directives to Italian- speaking learners. In H. C. Boas (Ed.), *Directions for Pedagogical Construction Grammar* (pp. 123–160). De Gruyter. <https://doi.org/10.1515/9783110746723-005>
- Gemmell Hudson, M. (2022). Teaching second year German using frames and constructions. In H. C. Boas(Ed.), *Directions for Pedagogical Construction Grammar* (pp. 265–304). De Gruyter. <https://doi.org/10.1515/9783110746723-009>
- Gilquin, G., & De Knop, S. (2016). *Applied construction grammar*. de Gruyter Mouton.
- Herbst, T. (2016). Foreign language learning is construction learning – what Else? Moving towards Pedagogical Construction Grammar. In S. De Knop & G. Gilquin (Eds.), *Applied Construction Grammar*. De Gruyter. <https://doi.org/10.1515/9783110458268-003>
- Le Foll, E. (2021). *Creating Corpus-Informed Materials for the English as a Foreign Language Classroom: A step-by-step guide for (trainee) teachers using online resources*. Zenodo. <https://doi.org/10.5281/ZENODO.4992504>
- Reppen, R. (2021). *Grammar and beyond. 1 1*. Cambridge: Cambridge University Press.



## The other grammarian's dream: Constructions as the most delicate grammar

Susan Hunston  
University of Birmingham, UK  
[S.E.HUNSTON@bham.ac.uk](mailto:S.E.HUNSTON@bham.ac.uk)

This paper makes the argument that a comprehensive account of mid-level constructions can be derived from corpus-based work on verb complementation patterns, and that such constructions can be used to populate sections of the transitivity network. The end points of the network are represented by such constructions rather than by individual verbs as in Hasan's (1987) seminal paper: *The grammarian's dream*. The paper gives examples of what sections of the proposed network look like, and discusses the potential of further work in this area.

In the SFG tradition, the transitivity network expresses the meaning distinctions observable in different process types. At the most general (left-wards) end of the network categories such as 'material' and 'relational' are identified. At a more delicate level, distinctions are drawn between, for example, transformative and creative material processes, or between perceptive and emotive mental processes (Halliday and Matthiessen 2014). At the right-most, or most delicate, end of the network, Hasan has attempted to map the meaning choices that lie behind use of one of a set of closely related verbs, such as *gather* and *accumulate* (Hasan 1987). While this work is both ambitious and inspiring, further development is hindered by the difficulty in tying together both ends of the network, and by the emphasis on the single lexical item – the verb – as the object being explained as the outcome of meaning choices. It has been suggested that concept of verb pattern or construction might be brought into service to populate the mid-way points in the network (Halliday & Matthiessen 2014: 66; Davidse 2017) or to supplement the network by including meaning made by structures longer than the transitive or intransitive verb structure that typifies discussion in the SFG literature (Lin and Peng 2006).

This paper argues that the right-most points of the transitivity network should be conceived of as populated by constructions, of a specific type, rather than by individual words. For example, instead of accounting for the verbs *ask* and *enquire*, the network would account for '*ask/enquire* wh-clause' and '*ask* someone wh-clause'. The constructions so used are form-meaning pairings derived from the extensive corpus research that lies behind Pattern Grammar (Francis 1993; Hunston and Francis 2000). It has been demonstrated in a number of papers (e.g. Hunston 2019; Hunston and Su 2019) that exploiting and adapting the 'meaning groups' identified for each complementation pattern associated with verbs, nouns and adjectives in English (Francis et al 1996; 1998) can lead to the identification of specifiable mid-level constructions and the range of verbs used with them. This approach suggests a potential alignment between patterns and constructions, to the benefit of both.

This paper demonstrates the feasibility of this idea with four detailed examples of pattern-derived networks with constructions as their most delicate exponent. It also assesses the potential for extending this work in the future, raising the question of its applicability to SFG, to Construction Grammar, and to Applied Linguistics.

## References

- Davidse K. 2017. Systemic functional linguistics and the clause: the experiential metafunction. In T. Bartlett and G. O'Grady (eds.) *The Routledge Handbook of Systemic Functional Linguistics*. London: Routledge.
- Francis G. 1993. A corpus-driven approach to grammar: principles, methods and examples. In M. Baker, G. Francis and E. Tognini-Bonelli (eds.) *Text and Technology: In Honour of John Sinclair*. Amsterdam: Benjamins.
- Francis G., Hunston S. and Manning E. 1996/1998. *Collins Cobuild Grammar Patterns 1 and 2*. London: HarperCollins. Also available at [grammar.collinsdictionary.com/grammar-pattern](http://grammar.collinsdictionary.com/grammar-pattern)
- Halliday M. and Matthiessen C. 2014. *Introduction to Functional Grammar*. 4<sup>th</sup> edition. London: Routledge. Pp 137-156.
- Hasan R. 1996 [1987]. The grammarian's dream: lexis as most delicate grammar. In C. Cloran, D. Butt and G. Williams (eds.) *Ways of Saying: Ways of Meaning. Papers by Ruqaiya Hasan*. London: Cassell. Pp 73-103.
- Hunston, S. 2019. Patterns, constructions, and applied linguistics. *International Journal of Corpus Linguistics* 24(3): 324-353.
- Hunston S. and Francis G. 2000. *Pattern Grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam: Benjamins.
- Hunston S. and Su H. 2019. Patterns, constructions, and local grammar: a case study of 'evaluation'. *Applied Linguistics* 40(4): 567-593.
- Lin F. and Peng A. 2006. Systemic Functional Grammar and Construction Grammar. *Proceedings of the 33<sup>rd</sup> International Systemic Functional Congress*.

## Do stance constructions become more formulaic in L2 academic writing development?

Sangeun Kim  
University of Exeter, UK  
[sk688@exeter.ac.uk](mailto:sk688@exeter.ac.uk)

Stance constructions have been of interest in writing development research for their potential to illustrate register-specific characteristics of academic prose (e.g., Hyland & Tse, 2005; Biber & Gray, 2016). This study adopts usage-based approaches to language learning, which posit that learning occurs through language use and formulaic structure is useful chunks memorised through exposure (Durrant, 2019; Ellis, 2002). Formulaicity is more salient at the early stages of writing development but less so at later stages, as verb–argument constructions (VACs) get diverse, and more irregular or specific head verbs get more frequent (e.g., Romer & Berger, 2019; Kyle & Crossley, 2018; Kyle, Crossley & Verspoor, 2021). This study aims to answer the following question: How does the formulaicity in VACs associated with stance function in academic texts written by university students of English as an additional language (L2) speakers of eight different first languages systematically vary across three academic semesters?

The primary data is the University of Pittsburgh I Language Institute corpus (PELIC; Juffs et al. 2020), an L2 longitudinal learner corpus collected in an intensive I programme. The selected 326 texts for this study consist of written texts produced by learners of 8 linguistic backgrounds who contributed written essays throughout three or more academic semesters with one or two revised versions. All the first and revised drafts are included in the analysis. In addition, the Michigan Corpus of Upper-Level Student Papers (MICUSP; O'Donnell & Römer, 2012; Römer & O'Donnell, 2011) was used as a reference corpus representative of the target student register. The entire MICUSP corpus is used to compute the collostruction score, which will be used as a reference to decide on the verbs highly associated with the target student register.

This study uses longitudinal, quantitative corpus-based approaches to explore the frequency and association strength variation of VACs associated with stance function, verb–that clauses, and *verb-to infinitives* in L2 writing. The basic procedures are as follows: First, the PELIC corpus is pre-processed so that each text can be identified with the writer and coded by 1) the personal course of study, reflecting the chronological order of different semesters in which the writer submitted written texts, and 2) the version of drafts, as to whether the draft is a first, second, or third version. Then, the corpus is tagged to all the occurrences of to-infinitives and that-clauses using the Multi-dimensional Analysis Tagger (Nini, 2019). Next, the constructions headed by stance verbs were identified based on the verb class taxonomy in Biber et al. (2021).

Approximate association strength scores of the stance verb constructions in the L2 texts and the MICUSP were computed using the collostructional analysis method, which is calculated by multiplying the delta P-value by the frequency of the verb (Gries et al., 2005; Kyle, 2016). Subsequently, two nonlinear growth models (Preacher & Hancock, 2015) were fitted to identify the individual and overall trajectories of 1) the overall association strength scores of the stance constructions in the PELIC texts and 2) the association strength scores of individual verbs selected due to their highest association strength. The trend is significantly inconsistent along individual trajectories due to the nature of the longitudinal data with relatively short time points mixed with revised versions. Nonetheless, the increase in the scores between some time points indicates development in producing more target-like constructions. In other words, the stance constructions with high formulaicity scores tend to be more saliently observable in later stages of the writing process. The results indicate that writers expand their VAC repertoire and productivity over time, and verb-VAC associations move closer to the target-like usage.

## References

- Biber, D., & Gray, B. (2016). Phrasal versus clausal discourse styles: A synchronic grammatical description of academic writing contrasted with other registers. In *Grammatical Complexity in Academic English: Linguistic Change in Writing* (pp. 67-124). Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (2021). *Grammar of Spoken and Written English*. John Benjamins. <https://doi.org/https://doi.org/10.1075/z.232>
- Durrant, P. (2019). *Formulaic language in English for Academic Purposes*. In A. Siyanova-Chanturia and A. Pelicer-Sanchez (Eds). *Understanding formulaic language: A second language acquisition perspective*. London: Routledge.
- Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24, 143–188.
- Gries, S. T., Hampe, B., & Schönefeld, D. (2005). Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. 16(4), 635-676. <https://doi.org/doi:10.1515/cogl.2005.16.4.635>
- Hyland, K., & Tse, P. (2005). Hooking the reader: a corpus study of evaluative that in abstracts. *English for Specific Purposes*, 24(2), 123-139. <https://doi.org/https://doi.org/10.1016/j.esp.2004.02.002>
- Juffs, A., Han, N-R., & Naismith, B. (2020). The University of Pittsburgh English Language Corpus (PELIC) [Data set]. <http://doi.org/10.5281/zenodo.3991977>. Retrieved from: <https://eli-data-mining-group.github.io/Pitt-ELI-Corpus>. Accessed 24 June 2021.
- Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine-grained indices of syntactic complexity and usage-based indices of syntactic sophistication* (Doctoral dissertation). Retrieved from: [https://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1035&context=alesl\\_diss](https://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1035&context=alesl_diss)
- Kyle, K., & Crossley, S. A. (2018). Measuring Syntactic Complexity in L2 Writing Using Fine-Grained Clausal and Phrasal Indices. *The Modern Language Journal*, 102(2), 333-349. <https://doi.org/10.1111/modl.12468>
- Kyle, K., Crossley, S., & Verspoor, M. (2021). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. *Studies in Second Language Acquisition*, 43(4), 781-812. <https://doi.org/10.1017/s0272263120000546>
- Nini, A. (2019). The Multi-Dimensional Analysis Tagger. In Berber Sardinha, T. & Veirano Pinto M. (eds), *Multi-Dimensional Analysis: Research Methods and Current Issues*, 67-94, London; New York: Bloomsbury Academic. Retrieved from: <https://sites.google.com/site/multidimensionaltagger/>. Accessed 21 July 2021
- O'Donnell, M. B., & Römer, U. (2012). From student hard drive to web corpus (part 2): The annotation and online distribution of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 7(1), 1-18.
- Preacher, K., & Hancock, G. (2015). Meaningful Aspects of Change as Novel Random Coefficients: A General Method for Reparameterizing Longitudinal Models. *Psychological Methods*, 20, 84-101. <https://doi.org/10.1037/met0000028>
- Römer, U., & Berger, C. M. (2019). Observing the emergence of constructional knowledge. *Studies in Second Language Acquisition*, 41(5), 1089-1110. <https://doi.org/10.1017/s0272263119000202>
- Römer, U., & O'Donnell, M. B. (2011). From student hard drive to web corpus (part 1): The design, compilation and genre classification of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 6(2), 159-177.

## Adopting a phrase-frame approach to investigate a corpus of research article introductions: Methodological issues

Aysel Şahin Kızıl<sup>1</sup>, Lee McCallum<sup>2</sup> & Benet Vincent<sup>2</sup>

1. İzmir Bakırçay University, Turkey; 2. Coventry University, UK  
[ad8848@coventry.ac.uk](mailto:ad8848@coventry.ac.uk); [ad7602@coventry.ac.uk](mailto:ad7602@coventry.ac.uk); [ab6667@coventry.ac.uk](mailto:ab6667@coventry.ac.uk)

The importance of gaining a better understanding of phraseology as a key lexicogrammatical phenomenon has been recognised for some time now in the area of English for Academic Purposes (EAP). A well-established approach is to use corpus software to extract lists of potentially useful phraseological phenomena from a corpus; see e.g., Ackermann and Chen, (2013) academic collocation list, Gilmore and Millar's (2018) list of lexical bundles in Civil Engineering research articles and Le and Harrington's (2015) 'word clusters' in discussion sections in Applied Linguistics research articles. Studies like this focus on continuous sequences and thus greatly limit the variation allowed in the phrases identified. An approach that attempts to address this limitation uses the 'phrase frame' (p-frame), a fixed sequence of items occurring frequently in a corpus but with one or two empty slots (Lu, Yoon, & Kisselev, 2021). This approach allows researchers to retrieve the most frequent p-frames in a particular corpus, then identify which specific variants are typically found, where they are found and what function they may have. This approach to identifying phrases can inform EAP pedagogy and thereby develop learners' lexicogrammatical awareness by helping them better understand how members of their discourse community typically express themselves (Lu, Yoon, & Kisselev, 2018).

The aim of the project for which this paper represents work in progress is to use a p-frame approach to create a list of pedagogically useful phrases and match these to their rhetorical functions in RA introductions in Health Sciences. The idea will be for the list to inform the writing teaching of novice researchers in this area. A number of studies have used a p-frame approach with similar aims though for different discipline areas, including Fuster-Márquez and Pennock-Speck (2015), Cunningham (2017) and Lu et al., (2018, 2021). However, these studies disagree on a number of issues central to p-frame methodology, presenting a challenge for new work in this area.

Our aim in this work in progress, therefore, is to evaluate some key variables of the approach that affect the composition of the list that is produced. To this end, we will discuss the following and explain how we came to decisions regarding the p-frame parameters used:

- corpus compilation: how did we decide on the size and compilation, who was consulted?
- p-frame length: previous studies have opted for 4, 5 or 6-item p-frames but how can one decide in a principled way?
- p-frame slot position: is it better to include the full range of options or focus only on frame-internal slots?
- frequency thresholds: what is the effect of choosing different thresholds? What might be the optimum level?
- completeness / meaningfulness of resulting phrases: what does it mean to say that a phrase is (in)complete or meaningful/meaningless? How does one decide?

This discussion will include examples of phrase types that are included or excluded based on the decisions arrived at and other lexicogrammatical implications.

## References

- Ackermann, K., & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL) - A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235–247. <https://doi.org/10.1016/j.jeap.2013.08.002>
- Cunningham, K. J. (2017). A phraseological exploration of recent mathematics research articles through key phrase frames. *Journal of English for Academic Purposes*, 25, 71. <https://doi.org/10.1016/j.jeap.2016.11.005>
- Fuster-Márquez, M., & Pennock-Speck, B. (2015). Target frames in British hotel websites. *International Journal of English Studies*, 15(1), 51–69. <https://doi.org/10.6018/ijes/2015/1/213231>
- Gilmore, A., & Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes*, 51, 1–17. <https://doi.org/10.1016/j.esp.2018.02.002>
- Le, T. N. P., & Harrington, M. (2015). Phraseology used to comment on results in the discussion section of applied linguistics quantitative research articles. *English for Specific Purposes*, 39, 45–61. <https://doi.org/10.1016/j.esp.2015.03.003>
- Lu, X., Yoon, J., & Kisselev, O. (2018). A phrase-frame list for social science research article introductions. *Journal of English for Academic Purposes*, 36, 76–85. <https://doi.org/10.1016/j.jeap.2018.09.004>
- 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>

## A corpus study of headless compound *chi-zao* in Chinese: A constructionalization approach

Xiaolong Lu

University of Arizona, USA

[charmander@email.arizona.edu](mailto:charmander@email.arizona.edu)

Compounding is a morphosyntactic phenomenon in the connection of grammar and lexicon in Chinese. Li and Thompson (1989: 81) would predict that the headless compound *chi-zao* 'late-early=sooner or later' is an adjective-adjective adverb concerning time. Its meaning "sooner or later" indicates a gradable property which has the two adjectives (*chi* and *zao*) as extreme poles, and speakers very often use this word in the discourse to show something will absolutely happen.

This research reports on the historical change of the compound *chi-zao* using the constructionalization framework (Traugott & Trousdale, 2013). The approach incorporates Construction Grammar, arguing that "constructionalization is the creation of form new-meaning new (combinations) of signs. It forms new type nodes, which have new syntax or morphology and new coded meaning, in the linguistic network and is accompanied by changes in degree of schematicity, productivity, and compositionality" (p. 22). Previous studies have examined compounds in Dutch and English (e.g., Booij, 2010; Booij & Audring, 2017), yet there is a lack of studies discussing constructionalization theories examining the historical change of Chinese compounds. This study aims to address two questions:

- i. What are the semantic and syntactic changes of *chi-zao* in its path to constructionalization?
- ii. What accounts for the constructionalization of *chi-zao*?

The CCL Chinese corpus ([http://ccl.pku.edu.cn:8080/ccl\\_corpus/](http://ccl.pku.edu.cn:8080/ccl_corpus/)) was used to trace historical data of *chi-zao*. I examined secondary literature to complement my data collection, including Tang poems and parts of novels in the Ming dynasty and Qing dynasty. My frequency count is based on the historical database of CCL corpus as the era (which dynasty) and source (which text) for the two variants of constructions (*chi-zao* and *chi zao*), which was clearly annotated and categorized. R statistics was employed to compute its token frequency at different stages, as is shown in Figure 1.

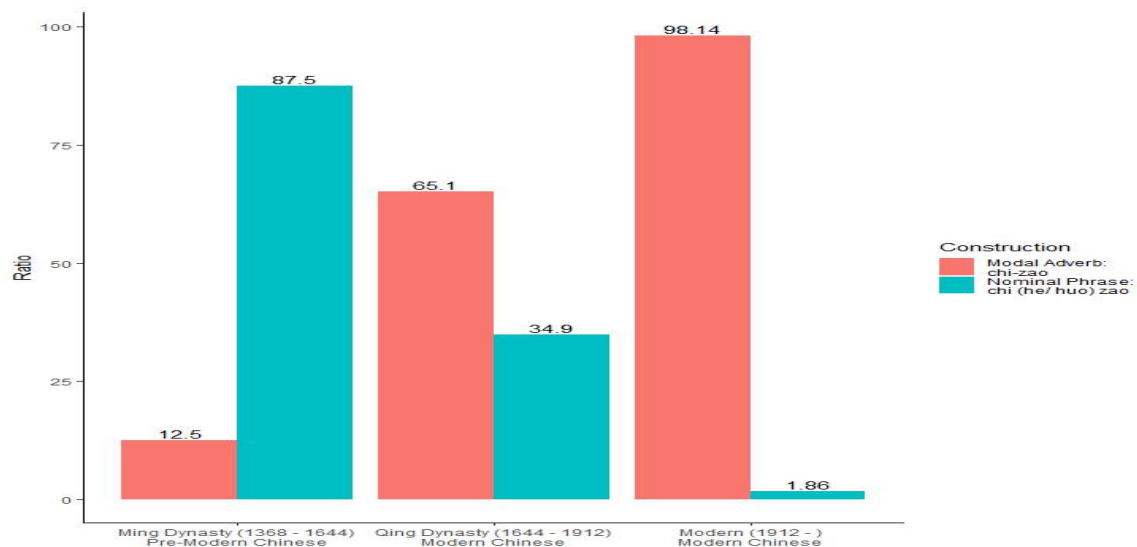


Figure 1. Token frequency of modal adverb (chi-zao) and nominal phrase (chi zao) in history  
Based on corpus analysis, I argue that the diachronic change of chi-zao provides evidence that constructionalization is at work at the compound level. Figure 2 tells that the form change is a gradual process: the antithesis structure from Medieval Chinese as the source construction is condensed into a nominal phrase (NP) as the intermediate construction, and then constructionalized into a compound word in Modern Chinese. Meanwhile, the meaning and function change reveals that the temporal meaning of chi zao (lateness and earliness) has been bleached into the epistemic modality found in “sooner or later” denoted by chi-zao. The whole process indicates that the creation of chi-zao as a new grammatical construction is not just a change of one morpheme, but a change in the construction as a whole in different stages.

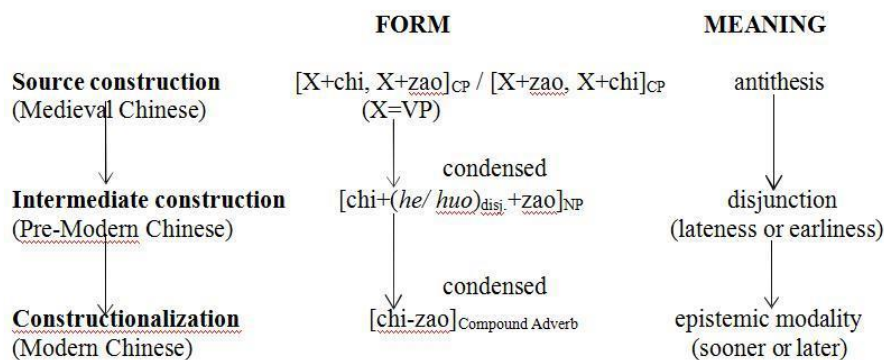


Figure 2. Grammatical constructionalization of *chio-zao* in Chinese

Subjectification (Traugott, 1995), frequency (Bybee, 2003) and metaphor (Lakoff & Johnson, 1980) are considered to explain the diachronic change and conventional usage of chi-zao as a modal adverb today. Taken together, the study of antonymous compounds in Mandarin can shed light on the historical change of lexicogrammar phenomena in other languages, including *sooner or later* (English) and *osokare hayakare* (Japanese) from a typological perspective.

## References

- Booij, G. (2010). *Construction Morphology*. Oxford: Oxford University Press.
- Center for Chinese Linguistics PKU (CCL corpus). [http://ccl.pku.edu.cn:8080/ccl\\_corpus/](http://ccl.pku.edu.cn:8080/ccl_corpus/) [Accessed from January to April, 2022].
- Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: IL: University of Chicago Press.
- Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar* (Vol. 3). University of California Press.
- Traugott, E. C. & Trousdale, G. (2013). *Constructionalization and Constructional Changes*. OUP.



## A dictionary of legal collocations: Looking back and looking forward

Tomasz Michta

University of Białystok, Poland

[t.michta@gmail.com](mailto:t.michta@gmail.com)

The motivation behind exploring collocations is at least two-fold. First, they are of theoretical interest and can be invoked to explain the fact that certain words tend to habitually occur in the company of other words (see e.g. Firth 1957, Sinclair 1991, Hoey 2005). Seen from this perspective, collocations appear to be a perfect illustration of the non-random nature of language (Kilgariff 2005), with lexical and grammatical factors contributing to their distinct nature. Second, and related to the first, the centrality of collocations in language and their unpredictability make them a potential minefield for language learners and sometimes for native speakers as well (Benson 1989: 4, Frankenberg-Garcia 2018).

The importance of collocations and the challenges related to their use are by no means limited to the context of general language (Benson 1989: 4, Frankenberg-Garcia 2018, Saber et al. 2020: 106, Giczela-Pastwa 2021: 191). Yet, the use of collocations in many LSPs, including legal English, continues to be underresearched. Moreover, while there is a considerable number of legal English teaching materials that focus on terminology, the same can hardly be said of collocations.

The overarching aim of this work-in-progress report is to explore selected challenges relevant to the compilation process of a monolingual corpus-based dictionary of legal English collocations and to discuss the solutions that were adopted in the exploratory phase of an ongoing lexicographic project (to be completed in 2022). While occasional references to existing dictionaries will be made, the bulk of the paper will be concerned with selected features of legal collocations and the principles guiding the compilation of a dictionary of legal collocations.

In order to investigate legal collocations, a purpose-built corpus was used. It consisted of 636 judgments handed down by the UK Supreme Court. It totalled 9.5 million tokens (8.1 million words) and spanned a ten-year period from the Court's inception in 2009 to 2018. To interrogate the corpus, Sketch Engine was used. First, a total of 100 nodes (all nouns) were selected on the basis of their frequency, legal importance as well as collocational productivity. The next stage involved extracting collocates for the chosen nodes. Qualitative and quantitative criteria were used to operationalise the concept of collocation for the purposes of the project. These criteria were rooted in both the frequency-based and the phraseological approaches, and included the frequency, range, as well as semantic and syntactic properties of such word combinations. Collocations were classified according to 6 patterns (1) premodifier + noun; (2) noun + noun; (3) verb + noun; (4) noun + verb; (5) preposition + noun; (6) noun + preposition. Both single-word as well as multi-word collocates were included. The latter were found to play an important role in legal phraseology due to their frequency. More importantly, however, they seem to provide counterevidence against treating collocations as structures consisting of two words. The study also found that a number of collocations showed various grammatical properties that should be accounted for in a dictionary of collocations. Some premodifying adjectives tended to be used in the comparative or superlative degree only. Certain nominal collocates were typically used in the plural. Finally, several verbal collocates revealed a preference for the passive voice or the imperative mood. While the study is not without its limitations, it does point to the rich variety of collocations found in legal English.

## References

- Benson, M. (1989). "The Structure of the Collocational Dictionary." *International Journal of Lexicography* 2 (1): 1-14.
- Firth, J. R. (1957). "Modes of Meaning." In: J. R. Firth. *Papers in linguistics, 1934-1951*, London: Oxford University Press, 190-215.
- Frankenberg-Garcia, A. (2018). "Investigating the collocations available to EAP writers." *Journal of English for Academic Purposes* 35 (September): 93–104.
- Giczela-Pastwa, J. (2021). "Developing phraseological competence in L2 legal translator trainees: a proposal of a data mining technique applied in translation from an LLD into ELF." *The Interpreter and Translator Trainer* 15 (2): 187–204.
- Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Kilgariff, A. (2005). "Language is never ever ever random." *Corpus Linguistics and Linguistic Theory* 1 (2): 263-276.
- Saber, A., A. Cartron, C. Kloppmann-Lambert, and C. Louis (2020). "Towards a typology of linguistic and stylistic errors in scientific abstracts written by low-proficiency geoscience and mechanical engineering doctoral students in France." *Fachsprache* 42 (3–4): 90–114.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

## Concessive conjunctions in Serbian: subordinators and/or coordinators?

Aleksandra Milosavljević

Institute for the Serbian Language of SASA, Serbia

[aleksandrasjovanovic@hotmail.com](mailto:aleksandrasjovanovic@hotmail.com)

This study is a corpus-based analysis of the concessive conjunctions *iako*, *mada* and *premda* ‘although’ in Serbian aimed at identifying semantic and syntactic differences between them. I test the hypothesis that each of the three concessive conjunctions is specialized (the strong version), or at least shows a tendency (the weak version) for some of the three main concessive relations distinguished in the literature. These concessive relations are usually referred to as standard, rectifying, and rhetorical concessives (cf. König 1988, 2006; Mizuno 2007), illustrated in (1–3), respectively, from König 1988:

- (1) **Although** John had no money, he went into this expensive restaurant.
- (2) He is very intelligent, **although** some of the things that he says are a bit silly.
- (3) **Even though** this solution would be harmful to our enemies, the damage done to us would be even greater.

Different types of concessives manifest various degree of subordination to the main clause: standard concessives are considered a typical instance of subordination, whereas rectifying and rhetorical concessives are closer to parataxis (cf. König 2006; Günthner 2000). In this sense, the hypothesis I test in this research predicts that concessive clauses marked by different conjunctions in Serbian show distinct syntactic behavior w.r.t. the integration into the main clause. In other words, there is a continuum ranging from concessive conjunctions as typical subordinators to concessive conjunctions as coordinators and discourse markers.

The research is coached within the usage-based framework (cf. Langacker 1988) and conducted on a random sample of 2.400 contextualized examples of concessives (800 examples per conjunction) extracted from the *Corpus of Contemporary Serbian* and the *Serbian web corpus*. The examples are manually annotated for various parameters, among which:

- (a) the type of conjunction;
- (b) the type of concessive relation;
- (c) the domain of use of the conjunction (in the sense of Sweetser 1990 and Crevels 2000 – i.e. content, epistemic, speech-act, or text domain);
- (d) the level of the introduced syntactic structure (clauses/NPs);
- (e) the position of the concessives w.r.t. the main clause;
- (f) the (lack of) illocutionary force of the concessive clause (cf. Lakoff 1984; Lehman 1988);
- (g) the foreground/background status of the concessive clause (cf. Tomlin 1985; Langacker 2008);
- (h) the possibility of the conjunction to be under the scope of a focus particle (cf. König 2006).

The results indicate an overall significant correlation between (I) the concessive conjunction, (II) the type of concessive relation it marks, and (III) its syntactic behavior. Most notably, the conjunction *iako* shows a strong tendency to mark the standard concessive relation, whereas *mada* and *premda* mark all three concessive relations almost equally. In the case of rectifying and rhetorical concessive uses, the conjunctions *mada* and *premda* behave unusually for subordinators – e.g. they can introduce a lower-level syntactic structure, they are obligatorily in a medial position, and never under the scope of focus particles. The concessive clauses marked

by these conjunctions can have their own illocutionary force and be foregrounded, which brings them closer to parataxis.

## References

- Crevels, M. (2000). Concessives on different semantic levels: A typological perspective. In Couper-Kuhlen, E., and B. Kortmann, eds. *Cause-condition-concession-contrast: Cognitive and discourse perspectives*. Walter de Gruyter, Vol. 33, 313–340.
- Gunthner, S. (2000). From concessive connector to discourse marker: The use of *obwohl* in everyday German interaction. In Couper-Kuhlen, E., and B. Kortmann, eds. *Cause-condition-concession-contrast: Cognitive and discourse perspectives*. Walter de Gruyter, Vol. 33, 439–468.
- König, E. (1988). Concessive connectives and concessive sentences: cross-linguistic regularities and pragmatic principles. In Hawkins, J. A. ed., *Explaining language universals*, Oxford: Blackwell, 145–166.
- König, E. (2006). Concessive clauses. In Brown, K., ed., *Encyclopedia of language and linguistics*. Elsevier. 820–824.
- Lakoff, G. (1984). Performative subordinate clauses. In *Annual Meeting of the Berkeley Linguistics Society*, Vol. 10, 472–480.
- Langacker, R. W. (1988). A usage-based model. In Rudzka-Ostyn, B., ed., *Topics in cognitive linguistics*. John Benjamins, 127–161.
- Langacker, R. W. (2008). Subordination in Cognitive Grammar, In Lewandowska-Tomaszczyk, B. (ed.), *Asymmetric Events*, John Benjamins Publishing Company: Amsterdam–Philadelphia, 137–149.
- Lehmann, C. (1988). Towards a typology of clause linkage, *Clause combining in grammar and discourse*, 18, 181–225.
- Mizuno, Y. (2007). *“Although” clauses in English Discourse: A functional analysis*. Doctoral dissertation. Hokkaido University.
- Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Vol. 54. Cambridge University Press. T
- Tomlin, R. S. (1985). Foreground-background information and the syntax of subordination, *Text-Interdisciplinary Journal for the Study of Discourse*, 5/1–2, 85–122.

## Corpora

*Corpus of Contemporary Serbian*.

<http://www.korpus.matf.bg.ac.rs>

*Serbian Web Corpus*.

[https://www.clarin.si/noske/all.cgi/corp\\_info?corpname=srwac&struct\\_attr\\_stats=1&subcorpora=1&fbclid=IwAR3](https://www.clarin.si/noske/all.cgi/corp_info?corpname=srwac&struct_attr_stats=1&subcorpora=1&fbclid=IwAR3)

**Word-formation as a complex product of lexico-grammatical relations:  
A study of Croatian nouns prefixed with *nad-* and *pod-***

Marta Petrak  
University of Zagreb, Croatia  
[mvlaic@ffzg.hr](mailto:mvlaic@ffzg.hr)

In this paper we present an analysis of Croatian nouns formed through simple prefixation (e.g. Körtvelyessy and Genči 2017) with the antonymous prefixes *nad-* ‘over-’ and *pod-* ‘under-’. We approach these prefix-formations as doubly motivated (e.g. Raffaelli 2013), both formally and semantically. More precisely, we hypothesize that all grammatical (word-formation) mechanisms always entail semantic change. This hypothesis is in line with the general Cognitive Linguistic tenet that symbolic units are the basic units of natural languages (Taylor 2003:38-58), or put differently, that there cannot be form without meaning in language (Tesnière 1976:39). Our main study question in this paper is, therefore, what occurs on the semantic level of these prefix-formations in Croatian.

In order to answer that question, we have extracted all nouns beginning with *nad-* and *pod-* from hrWaC (Baroni et al. 2009), the largest extant Croatian corpus, with a minimal frequency of 10 occurrences. This is in line with the general usage-based approach adopted in Cognitive Linguistics (e.g. Evans and Green 2006). The corpus results were manually checked for typos, words that are not a result of prefixation and other types of noise. The final list consisted of a total of 258 nouns, 58 formed with *nad-* and 200 with *pod-*. All the prefixed nouns were analysed semantically on the basis of the context in which they are used in the corpus and meanings attested in the Croatian Language Portal (HJP).

Our analysis has demonstrated that, first of all, when the prefixes *nad-* and *pod-* are attached to nominal bases, they construe a polysemous semantic network resembling that of categories traditionally labelled “lexical” such as nouns or verbs. Concrete spatial meanings make 24% of all the meanings of the prefix *nad-* (e.g. *nadlak* ‘lit. over-polish; topcoat’) and 10% of the meanings of the prefix *pod-* (e.g. *podzid* ‘lit. subwall’). In other words, the meanings of the two prefixes are predominantly extended, mostly based on orientational metaphors of the general type MORE/BETTER IS UP (e.g. *nadbiskup* ‘archbishop’) and WORSE/LESS IS DOWN (e.g. *potpukovnik* ‘lit. under-lieutenant; lieutenant colonel’) (e.g. Lakoff and Johnson 1980), by which more positive phenomena are conceptualized as being located higher in space, while less positive phenomena are thought of as being located in inferior spatial positions. Not only does prefixation with *nad-* and *pod-* often entail metaphorization, but it also frequently leads to hyponymy. For instance, *nadstolnjak* ‘lit. above-tablecloth; table runner’ is a type of *stolnjak* ‘tablecloth’ and *pothodnik* ‘underpass’ is a type of *hodnik* ‘pass’. Thus *nad-* and *pod-* formations contribute to lexical enrichment by building a number of nouns a majority of which are hyponyms of their bases. In addition to hyponymy, the prefixes *nad-* and *pod-* also build antonymy relations: there are 13 pairs of antonymous nouns (e.g. *nadnaslov* ‘supertitle’ – *podnaslov* ‘subtitle’) and 7 pairs of antonymous prefixal meanings (e.g. ‘superiority’ – ‘inferiority’, ‘control’ – ‘lack of control’), altogether.

This study has demonstrated that word-formation phenomena such as simple prefixation cannot be analysed thoroughly without considering both the formal (grammatical) and semantic level, and therefore prove the necessity of the grammar-lexicon continuum.

## References

- Baroni, Marco, Bernardini, Silvia, Ferraresi, Adriano i Eros Zanchetta. 2009. „The WaCky wide web: a collection of very large linguistically processed web-crawled corpora“. *Language Resources and Evaluation* 43:209-226.
- Evans, Vyvyan i Melanie Green. 2006. *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Hrvatski jezični portal (HJP), Novi Liber, SRCE, <http://hjp.novi-liber.hr/>
- Körtvélyessy, Livia and Ján Genči. 2017. „Cross-linguistic research into affixation processes in European languages“. *SKASE Journal of Theoretical Linguistics* 14(2):18-34.
- Lakoff, George and Johnson, Mark. 1980. *Metaphors We Live By*. Chicago : London: The University of Chicago Press.
- Raffaelli, Ida. 2013. „The model of morphosemantic patterns in the description of lexical architecture“. *Lingue e linguaggio* 1:47-72.
- Taylor, John R. 2003. *Cognitive Grammar*. New York: Oxford University Press.
- Tesnière, Lucien. 1976. *Éléments de syntaxe structurale*. Pariz: Librairie C. Klincksieck.

**Not all subjects are “raised” equally:  
The alternation between the constructions with *sembrare* (*seem*)  
in written and spoken Italian**

Flavio Pisciotta  
University of Bologna  
[flavio.pisciotta2@unibo.it](mailto:flavio.pisciotta2@unibo.it)

Traditionally, *sembrare* (*seem*) has gained interest for its peculiar syntactic and semantic behaviour. In particular, scholars have claimed that the raising construction with *sembrare* (a) is truth-conditionally equivalent to the impersonal construction in (b) (Davies & Dubinsky 2004: 4):

- a. Gianni sembra amare Piero.  
‘Gianni seems to love Piero’
- b. Sembra che Gianni ami Piero.  
‘It seems that Gianni loves Piero’

This has led to a transformational analysis of (1a), where the subject *Gianni* is selected by the subordinate verb and then raised to be the syntactic subject of *sembrare*. This perspective implicitly maintains a free variation between the two constructions. Nonetheless, many scholars working in constructionist frameworks have found that a number of formal alternations are motivated by differences in the functional properties of the alternating constructions (e.g., Gries 1999; 2003; Stefanowitsch 2003; Perek 2015; De Vaere et al. 2018). Following a Construction Grammar approach, we hypothesize that *sembrare*+infinitive (1a) and *sembra che* (1b) are allostructions, i.e., formal variants of a more abstract construction that differ with respect to some semantic or pragmatic properties (Cappelle 2006).

To test this hypothesis, we focus on selection properties by analyzing whether any subject is a good candidate to appear in the constructions under exam. 216 occurrences of the two constructions were selected from the written Italian corpus CORIS (Rossini Favretti et al. 2002), and 189 occurrences from two spoken Italian corpora: LIP (De Mauro et al. 1993) and KIParla (Mauri et al. 2019). We tested four variables related to “raised” and “unraised” subjects, namely length, animacy, accessibility, and type of anaphoric link represented by the subject (Nissim et al. 2004). We used the Mann-Whitney test to assess the length of the subjects, and the chi-squared test with Monte Carlo simulation of p-values for each of the other three variables. Written and spoken occurrences were tested separately.

Our results show that there are, in fact, differences in the selection of the subject. In the data from the written corpus, the subjects of *sembrare*+infinitive tend to be longer than the subjects of *sembra che* (p-value = 0.005374). In the data from the spoken corpora, the type of anaphoric link is highly significant (p-value = 9.999e-05): the standardized residuals show that *sembra che* rarely selects a relative pronoun as its subject, while it is very commonly selected by *sembrare*+infinitive. Furthermore, *sembrare*+infinitive never selects a 1<sup>st</sup>/2<sup>nd</sup> person pronoun as its subject, unlike *sembra che*. These two restrictions on the anaphoric links proved to be valid also for our written data, even if they did not provide a significant result; this is probably due to their lower frequency in the written corpus.

These results support the hypothesis that *sembrare*+infinitive and *sembra che* are allostructions, although the functional differences between the two constructions need to be inquired more thoroughly. Moreover, this study highlights how the analysis of spoken data can be valuable for syntactic theory, providing examples of phenomena that would otherwise be lost.

## References

- Cappelle, B. (2006) Particle placement and the case for 'allostructions'. *Constructions*, SV 1(7), pp. 1-28.
- Davies, W. D. & Dubinsky, S. (2004) *The grammar of raising and control: A course in syntactic argumentation*. Malden: Blackwell Publishing.
- De Mauro, T., Mancini, F., Vedovelli, M. & Voghera, M. (1993) *Lessico di frequenza dell'italiano parlato*. Milano: Etaslibri. Available at: <http://badip.uni-graz.at/it> [accessed 13 April 2022].
- De Vaere, H., De Cuypere, L. & Willems, K. (2018) Alternating constructions with ditransitive 'geben' in present-day German. *Corpus Linguistics and Linguistic Theory*, 17(1), pp. 73-107.
- Gries, S. Th. (1999) Particle movement: A cognitive and functional approach. *Cognitive Linguistics*, 10(2), pp. 105-145.
- Gries, S. Th. (2003) *Multifactorial analysis in corpus linguistics: A study of particle placement*. London: Continuum.
- Mauri, C., Ballarè, S., Gorla, E., Cerruti, M. & Suriano, F. (2019) KIParla corpus: a new resource for spoken Italian. In: Raffaella Bernardi, Roberto Navigli & Giovanni Semeraro (eds.), *Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it, November 13-15, 2019, Bari, Italy*. Available at: <https://kiparla.it/search/> [accessed 13 April 2022]
- Nissim, M., Dingare, S., Carletta, J. & Steedman, M. (2004) An Annotation Scheme for Information Status in Dialogue. In: Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (eds.), *Proceedings of the 4th Conference on Language Resources and Evaluation (LREC2004), May 2004, Lisbon, Portugal*. Available at: <http://www.lrec-conf.org/proceedings/lrec2004/pdf/638.pdf> [accessed 13 April 2022].
- Perek, F. (2015) *Argument Structure in usage-based Construction Grammar*. Amsterdam: John Benjamins.
- Rossini Favretti, R., Tamburini, F. & De Santis, C. (2002) CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In: Andrew Wilson, Paul Rayson & Tony McEnery (eds.), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*, pp. 27-38. München: Lincom-Europa. Available at: <http://corpora.dslo.unibo.it/TCORIS/> [accessed 13 April 2022]
- Stefanowitsch, A. (2003) Constructional semantics as a limit to grammatical alternation: The two genitives of English. In: Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of Grammatical variation in English*, pp. 413-443. Berlin-Boston: De Gruyter Mouton.



## Valency Patterns of TAKE in Hong Kong English A Work-in-progress report

Yating Tao

Université catholique de Louvain, Belgium

[yating.tao@uclouvain.be](mailto:yating.tao@uclouvain.be)

Pattern grammar (Hunston & Francis, 2000) has always been viewed as the first corpus-driven grammar model, which tries to systematically establish a connection between lexis and grammar. However, it only emphasizes parts of speech and the necessary lexis in its descriptive model, which usually leads to different interpretations for one and the same pattern. “*The hatred of a million coolies*” (*N of N* in pattern grammar), for example, can not only be interpreted as a) the colonizers who hate the coolies but also as b) the coolies who hate the colonizers (Teubert, 2007: 224). Some researchers (e.g. Reichardt, 2014; Zhen & Yang, 2015), therefore, put forward valency pattern, which combines valency grammar with pattern grammar, with a view to overcoming the ambiguity of pattern grammar and to clearly reflecting the dependency relationships between the headword and its governed lexical items (e.g. *sub V obj vb(to-inf)*).

Within this framework, this study aims to investigate the structural features of the highly polysemous verb TAKE in Hong Kong written English. Student writing samples from the Hong Kong component of the International Corpus of English (ICE) are used, together with the British university student essays from the Louvain Corpus of Native English Essays (LOCNESS) as reference samples. Due to the lack of phraseological descriptions in the valency pattern model, this study will analyze the valency patterns of TAKE at two levels. At the higher level, all the cases including TAKE will be analyzed in the manner of valency pattern, namely by distinguishing between monovalency, divalency, trivalency and quadrivalency, and their subcategories. At the lower level, I will examine phraseological uses including phrasal verbs (PVs), light verb constructions (LVCs) and idioms, which have been shown to remain a stumbling block to non-native English speakers (e.g. Gouverneur, 2008). To identify the potential variations of valency patterns of TAKE in Hong Kong English (HKE), I use the chi-squared test to detect the differences in valency patterns and phraseological uses between the two corpora.

The results show that there is a significant difference ( $p < 0.01$ ) between British English (BrE) and HKE regarding the overall valency patterns of TAKE. The BrE data include 4 types of valency pattern with 52 subcategories, while the HKE data include 4 types with 22 subcategories, among which 5 subcategories are not found in the British counterpart. Interestingly, 4 of 5 non-standard uses in HKE are related to *take place* (e.g. *be taken place*).

A significant difference also appears at the phraseological level ( $p < 0.01$ ), but the overall proportions of the phraseological uses among all patterns in BrE (58.7%) and in HKE (58.8%) are similar, which is an echo to the phraseological tendency (Sinclair, 1991) in language use. However, the PVs, LVCs and idioms in BrE are far more diverse than those in HKE. HKE speakers tend to repeat the same patterns more often than BrE speakers (standard type-token ratio = 66.1% and 69.5%, respectively). More nuanced variations regarding the phraseological uses (e.g. new verb+noun combinations in the LVCs) will be discussed further in my presentation.

## References

- Gouverneur, C. (2008). The phraseological patterns of high-frequency verbs in advanced English for general purposes: A corpus-driven approach to EFL textbook analysis. In F. Meunier & S. Granger (eds) *Phraseology in Foreign Language Learning and Teaching* (pp. 223-243). Amsterdam: John Benjamins.
- Hunston, S. & Francis, G. (2000). *Pattern Grammar: A Corpus-Drive Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Reichardt, R. (2014). Valency Sentence Patterns and Meaning Interpretation. Unpublished PhD thesis. Birmingham: University of Birmingham.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Teubert, W. (2007). Sinclair, pattern grammar and the question of hatred. *International Journal of Corpus Linguistics*. 12: 223-248.
- Zhen, F. & Yang, F. (2015). Verb valency in learner English: A corpus-driven study of CONSIDER. *Journal of Foreign Languages* 6: 57-67.

## The applicability of reading illustrated books with Learners of Spanish as a foreign language: A corpus analysis

Paulina Valenzuela-Lagos & Sy-Ying Lee  
National Taiwan University of Science and Technology, Taiwan  
[m10917808@gapps.ntust.edu.tw](mailto:m10917808@gapps.ntust.edu.tw)

Extensive reading (Krashen, 2004) has multiple positive impacts on L2 acquisition. However, its success depends on whether suitable reading materials are accessible. Some researchers argue that simplified texts (e.g., graded readers) distort the language, making it unsuitable for learners (Nation & Wang, 1999). On the other hand, authentic texts may overwhelm learners with a high number of challenging vocabulary (Nation, 1983), which may impede reading fluency and comprehension (Herman, 2019). Our study thus argued for using both types of texts in teaching and learning Spanish as a foreign language in and out of the classroom.

This study compared graded illustrated books (GRB) and authentic illustrated books (ACB) in Spanish from a lexico-grammatical perspective to determine their respective impact on beginning learners. Two corpora, thirteen GRB (9836 tokens) and twenty-nine ACB (9790 tokens), were analyzed using *AntConc*, *Venny*, and *Jasp* to compare the variety and frequency of verbs in general, as well as the use of moods, tenses, and verboids. Another analysis was conducted to determine the difficulty level of the verbs used in the two corpora by comparing them to a frequency dictionary of Spanish (Davies, 2006).

Results using z-test showed a significant difference between the variety of verbs utilized in both texts ( $p < 0.0002$ ), with ACB containing more different types of verbs than GRB. An independent sample T-test, however, showed that the frequency of verbs is not statistically different ( $t = 1.67$ ,  $p < 0.096$ ). In terms of moods, tenses, and verboids, it was found that: 1) GRB contains more indicative tenses but no subjunctive mood, which is a more advanced grammatical structure often used in daily conversations. 2) GRB uses more present tenses, fewer past tenses, and no future or conditional tenses, indicating an adaptation to the Spanish curricular plan for beginning learners to avoid the structures learners haven't learned. 3) GRB has significantly more infinitives and fewer gerunds, with the latter used in more complex sentences in ACB (temporal, causal, modal, conditional, concession). 4) A *Venn* diagram revealed 165 (43.3%) shared verbs in both corpora, 48 (12.6%) distinctive verbs in GRB, and 168 (44.1%) in ACB. Among the common verbs, no statistical difference was found between their frequency in each corpus ( $p < .91$ ), which means that GRB holds highly frequent verbs in fewer texts, while ACB use more diverse verb forms rather than repeating the words. Moreover, 88% of the verbs in GRB, 76% in ACB, and 93% of the shared verbs are in the first 3000 frequent words, with GRB providing simpler and more highly frequent vocabulary.

In conclusion, graded readers provide easier and more frequent vocabulary and grammar structures, while authentic texts contain more diverse vocabulary and complex structures. This implies that for beginners: 1) GRB are proper materials because they follow the lexico-grammatical recommendations for A1 and A2 students, and they offer equally highly repeated verbs for easier Spanish vocabulary acquisition. 2) With a similar pattern, ACB can still be used to expand the quality and quantity of language input. 3) GRB and ACB can serve as partnering bridge materials that invite learners to begin the journey of reading in Spanish while aiming toward reading more authentic and challenging texts.

## References

- Davies, M. (2006). *A Frequency Dictionary of Spanish*. Routledge.
- Herman, E. (2019). *The Relationship between Lexical Coverage and Levels of Reading Comprehension: Extensive Reading of Graded Readers by L2 Spanish Beginners*. Florida State University.
- Krashen, S. (2004). *The power of reading: Insights from the research* (2nd ed.). Libraries Unlimited.
- Krashen, S. (2013). *Second language acquisition: Theory, applications, and some conjectures*. Cambridge University Press.
- Nation, P. (1983). Testing and teaching vocabulary. *Guidelines*, 5(1), 12–25.
- Nation, P., & Wang Ming-tzu, K. (1999). *Graded Readers and Vocabulary*. 355–380.
- Rodrigo, V. (2016). Graded readers: Validating reading levels across publishers. *Hispania*, 66–86.

## How formulaic sequences link to larger rhetorical patterns: The case of the German formulae “Fakt ist” and “Tatsache ist”

Mirjam Weder

University of Basel, Switzerland

[mirjam.weder@unibas.ch](mailto:mirjam.weder@unibas.ch)

There is abundant research about the fact that lexico-grammatical features such as patterns and multiword units such as lexical bundles, collocations, constructions, or formulaic sequences relate to register and genre (cf. for an overview Hunston, 2022, p. 186ff.). There is less research, however, about how they are embedded in and contribute to larger rhetorical moves (cf. Swales, 1990).

The proposed paper will present a corpus-linguistic case study on the formulae «Fakt ist» and «Tatsache ist» (in English: ‘The fact is’, ‘as a matter of fact’). Instances will be drawn from various German reference corpora, such as DWDS, COSMAS, and Sketch Engine (German Web Corpus). The formulae “Fakt ist” and “Tatsache ist” are fully lexicalized and highly restricted patterns with little variation. They tend to appear in media texts and related text types. A formal analysis reveals the following structures:

FAKT/TATSACHE IST [+ADV]: full clause.

FAKT/TATSACHE IST [+ADV], DASS [‘that’] + subordinate clause.

The “Fakt/Tatsache ist”-formulae feature some striking formal restrictions: the NP FAKT or TATSACHE cannot be preceded by a determiner (cf. example 1-3); they are either followed by a that-clause or colon and a full sentence, and a corpus linguistic analysis shows that there are instances in which the verb is omitted altogether (cf. 3). All in all, this formula appears to be very condensed.

- (1) **Fakt ist**, dass Fahrzeuge sich die meiste Zeit nicht bewegen und dass die Produktion von erneuerbaren Energien Schwankungen aufweist, die ausgeglichen werden müssen. (corpus: DWDS; source: Die Zeit, 02.12.2016 online)  
*Engl. translation: The fact is that vehicles don't move most of the time and that renewable energy production has fluctuations that need to be balanced.*
- (2) **Fakt ist:** Trump ist mit einer völlig faktenfreien Kampagne zum Präsidenten gewählt worden. (Corpus: DWDS; source: Die Zeit, 23.01.2017 online)  
*Engl. translation: The fact is, Trump was elected president on a completely fact-free campaign.*
- (3) **Fakt:** der Vatikan befindet sich auf der Intensivstation. (Corpus German: Sketch Engine, Source iknews.de)  
*Engl. translation: Fact: the Vatican is in intensive care.*

The lexical meaning of the “Fakt/Tatsache ist”-formula seems to be quite transparent: it is a so-called stance noun (Gray & Biber, 2015, p. 228), i.e. an epistemic marker which frames the following proposition as true. However, as Wray points out, formulae tend to have “a second layer of pragmatic meaning” (Wray, 2002, p. 58) that can only be revealed through the careful analysis of the context and the communicative intent as expressed in the text (Wray, 2002, pp. 58–59; Coulmas, 1981, pp. 16–17). In the case of the “Fakt/Tatsache ist”-formula, the pragmatic meaning is rather complex and does not necessarily align with the lexical semantics of the

formula. The examples above illustrate, for example, that the meaning varies as far as the degree of factivity and objectivity of the embedded clause is concerned. Furthermore, it is important to look at the preceding text. The formula is often applied after contradictory facts and opinions are presented or after an unclear situation is established. The formula then serves to introduce a new rhetorical move (cf. Swales, 1990) which may indicate what can count as an established fact (or is at least perceived as such in the eyes of the writer), or expresses the stance of the writer in the contested situation, or introduces new aspects, or offers a conclusion and so on. The lowest common denominator of these various types of rhetorical moves seems to be discursive insofar as it contributes to text cohesion (Wray, 2002, p. 97). The work-in-progress report will discuss how the “Fakt/Tatsache ist”-formula is embedded in larger discourse patterns and how the context impacts the pragmatic meaning of the formula.

## References

- Coulmas, F. (1981). Introduction: Conversational Routine. In F. Coulmas (Ed.), *Rasmus Rask Studies in Pragmatic Linguistics, Volume 2, Conversational Routine* (pp. 1–18). De Gruyter Mouton. <https://doi.org/10.1515/9783110809145.1>
- Gray, B., & Biber, D. (2015). Stance markers. In K. Aijmer & C. Rühlemann (Eds.), *Corpus Pragmatics* (pp. 219–248). Cambridge University Press. <https://doi.org/10.1017/CBO9781139057493.012>
- Hunston, S. (2022). *Corpora in applied linguistics* (Second edition). Cambridge University Press.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.