Corpus Approaches to Lexicogrammar 2021

2-3 July 2021

http://ehu.ac.uk/lxgr

LxGr2021 SYMPOSIUM ARCHIVE

PROGRAMME
COMMITTEE
ABSTRACTS

LxGr2021 SYMPOSIUM PROGRAMME

Day 1: Friday 2 July 2021

9.30am – 10am   WELCOME

10am – 11am   GUEST PRESENTATION
Sebastian HOFFMANN (University of Trier)
Rhythm in World Englishes: A look at the interaction of phonology and lexico-grammar from a corpus-based perspective

11am – 11.30am   BREAK / CHAT

11.30am – 12.05pm
Ling LIN (Shanghai Jiao Tong University) & **Ming LIU** (Hong Kong Polytechnic University)
Part-of-speech patterns in research introductions: A cross-disciplinary study

12.10pm – 12.45pm
Christina Sook Beng ONG (Universiti Tunku Abdul Rahman)
Nativised light verb constructions in Malaysian English

12.45pm – 1.15pm   BREAK / CHAT

1.15pm – 1.50pm
Souhaila MESSAOUDI, Claire BRIERLEY & **James DICKINS** (University of Leeds)
Investigating French interference in Algerian students' English-Arabic / Arabic-English translations   of   collocations

1.55pm – 2.30pm
Alex CARR (Cardiff University)
An exploration into the relationship between nominal form and temporal semantics

2.30pm – 3pm   BREAK / CHAT

3pm – 3.35pm
Joyce LIM (University of Cambridge)
Exploring the development of 'verb 4-gram' sequences and grammar patterns in L2 writing: A comparative study of L1 Korean speakers and L2 speakers of various L1

3.40pm – 4.15pm
Jim LAW (Brigham Young University)
A lexicogrammatical approach to participant role alternations in the French spending frame

4.15pm – 4.45pm   BREAK / CHAT

4.45pm – 5.20pm
Lucy CHRISPIN (Cardiff University)
A corpus investigation into lexical aspect as a lexicogrammatical feature of the intransitive construction

5.25pm – 6pm
Milntra RAKSACHAT (University of Oregon)
Serial verb ʔaw 'take' with instrumental meaning in Isaan: A distinct construction?

Day 2: Saturday 3 July 2021

9am – 9.30am   WELCOME

9.30am – 10.05am
Susanne DeVORE (University of Hawaii)
Usage-based indices of proficiency: Verb-vac and phrasal constructions in Mandarin

10.10am – 10.45am
Chen-Yu Chester HSIEH (National Taiwan University of Science and Technology)
Categorization for evaluation: a corpus-based study on the use of zhong *'kind; type' with constructions of evaluation in Mandarin conversation*

10.50am – 11.25am
Daniel JACH (Southwest Jiaotong University)
The case of German two-way prepositions

11.25am – 11.55am   BREAK / CHAT

11.55am – 12.30pm
Viktoria ZHUKOVSKA (Zhytomyr State Ivan Franko University)
English with/with-less-subjpartᵢ *constructions: A case of a quantitative corpus-based analysis*

12.35pm – 1.10pm
Piotr TWARDZISZ (University of Warsaw)
A corpus-based study of affixation tendencies in related academic disciplines.

1.10pm – 1.40pm   BREAK / CHAT

1.40pm – 2.15pm
Eva ZEHENTNER (University of Zurich)
Alternations emerge and disappear: The network of dispossession constructions in the history of English

2.20pm – 2.55pm
Carlos KAUFFMANN (São Paulo Catholic University)
Multi-dimensional analysis of literary style from a lexicogrammatical perspective

2.55pm – 3.25pm   BREAK / CHAT

3.25pm – 4pm
Irene CHECA-GARCIA (University of Wyoming)
What the correlation between syntactic complexity and lexical development measures can tell us about modes of acquisition

4.05pm – 4.40pm
Lucia BUSSO (Aston University)
The blended nature of legal-lay language in Italian and English: A corpus-based contrastive analysis

4.40pm – 5.40pm   DISCUSSION & FEEDBACK

LxGr2021 PROGRAMME COMMITTEE

Federica Barbieri (University of Swansea)

Tine Breban (University of Manchester)

Belén Díez Bedmar (University of Jaén)

Eva Duran Eppler (University of Roehampton)

Lise Fontaine (Cardiff University)

Gaëtanelle Gilquin (Université catholique de Louvain)

Christopher Gledhill (Université Paris-Diderot)

Nick Groom (University of Birmingham)

Glenn Hadikin (University of Portsmouth)

Andrew Hardie (Lancaster University)

Sebastian Hoffmann (University of Trier)

Jennifer Hughes (Lancaster University)

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)

Pascual Perez-Paredes (University of Cambridge)

Paul Rayson (Lancaster University)

Ute Römer (Georgia State University)

Benet Vincent (Coventry University)

John Williams (University of Portsmouth)

Stefanie Wulff (University of Florida)

LxGr2021

SYMPOSIUM ABSTRACTS

The blended nature of legal-lay language in Italian and English:
A corpus-based contrastive analysis

Lucia Busso

Aston Institute for Forensic Linguistics, Aston University

l.busso@aston.ac.uk

The present contribution proposes a contrastive analysis of lexico-grammatical features in English and Italian legal-lay language (henceforth: LLL) – defined here as any legal text type aimed at a non-specialist audience (Tiersma, 1999; Bhatia, 2010). Studies on LLL as an independent genre are still scarce in the linguistic literature (Van Boom et al., 2016; Conklin et al., 2019), which predominantly focus on legal language simplification (*inter alia*: Cortelazzo, 2008; Adler, 2012; Mori, 2019).

The study investigates a specialistic corpus of LLL using collostructional analysis (Römer, 2009; Stefanowitsch, 2013) and a comparative frequency analysis with other specialised corpora. The research has two primary aims: first, to provide an exploratory account of the lexico-grammatical features of LLL and second, to ascertain whether such features can be considered idiosyncratic. Particularly, the research sets off from two interrelated hypotheses: that LLL exhibits idiosyncratic lexico-grammatical characters, different from specialised legal jargon and non-specialised written prose alike; we further hypothesize that LLL will show a 'blended' nature, with mixed characters between these two genres. To analyse both the lexical and the grammatical end of the lexico-grammar continuum (Halliday, 1991; Gabrielatos, 2018), the study takes the constructionist standpoint that language is formed by *constructions*, holistic pairs of form (syntax) and function (semantics) (Goldberg, 2006, 2019). Construction Grammar is in fact increasingly applied in both synchronic and diachronic corpus-based studies (Gries, 2013; Hilpert, 2013). A growing body of literature has also been using constructionist tenets for the analysis of genre (Hoffmann & Bergs, 2018).

The research is conducted on a self-compiled specialised corpus that comprises several textual types ascribable to LLL: CorIELLS (CORpus of Italian and English Legal-lay textS, Busso [accepted]). To analyse grammatical patterns and the lexical items they tend to co-occur with in CorIELLS, collostructional analysis is adopted. This family of quantitative methods finds statistically associated subcategorization preferences (i.e., lexical items) for a given abstract grammatical construction. Simple and covarying collexeme analyses are employed to analyse four grammatical constructions: nominalizations heading prepositional phrases, participial constructions, modal verbs, and passive constructions. These constructions were selected at different abstractedness levels, based on previous research on legal and bureaucratic grammatical features in both Italian and English (Garavelli, 2001; Brunato and Venturi, 2014; Coppolella, 2014; Brunato, 2015; Mori, 2019).

Statistically associated collexemes found with the collostructional analysis are further contrasted to the same structures in different corpora of legal and non-specialised written prose: the specialised legal and written prose subcorpora of the Italian reference corpus CORIS (Rossini-Favretti, 2000) and for English the imaginative subcorpus of the BNC and an ad-hoc created subcorpus of EurLEX-English. Data are analysed with linear mixed-effect-modelling to highlight differences in usage. Results from the analyses of both languages will be further compared and discussed to find differences and similarities in the lexico-grammatical profile of LLL in Italian and English.

The analysis of English is still ongoing, but findings from Italian appear to preliminary validate our hypotheses: subcategorization preferences show a mixture of specifically legal and more colloquial lexical choices. Moreover, constructions are used significantly differently than in both non-specialised prose and legal jargon (Busso, submitted).

References

Bangalore, S. & Joshi, A.K. (1999) Supertagging: An approach to almost parsing. Computational linguistics, 25(2), 237-265.

Adler, M. (2012). *The Plain Language Movement*. Oxford University Press.

Bhatia, V.K. (2010). Specification in Legislative Writing: Accessibility, Transparency, Power and Control. In M. Coulthard, & A. Johnson, (eds) *The Routlege Handbook of Forensic Linguistics* (p.50). Routledge.

Brunato, D. (2015). Operationalizing linguistic complexity from a NLP perspective: The computational assessment of text readability. *Unpublished Phd Thesis University of Siena.*

Brunato, D., and Venturi, G. (2014). Le tecnologie linguistico-computazionali nella misura della leggibilità di testi giuridici [computational linguistic technologies in legal texts readability measure]. *Informatica e diritto*, XL (XXIII), 111-142.

Busso, L. (accepted at the *Società Linguistica Italiana Conference* 2020, postponed to September 2021). *CorIELLS:* A specialised bilingual corpus of lay legal communication.

Busso, L. (submitted). Lexicon and grammar in legal-lay language: a quantitative corpus study on Italian

Coppolella, M. (2014). Modal values of verbal forms in the European Charter for Regional or Minority Languages. A linguistic comparative analysis of the English, Italian and Spanish versions. *Comparative Legilinguistics*, *18*, 24-46.

Conklin, K., Hyde, R., & Parente, F. (2019). Assessing plain and intelligible language in the Consumer Rights Act: a role for reading scores? *Legal Studies,* 1-20.

Cortelazzo, M. (2008). È meglio essere chiari e leggibili [It is best to be clear and readable]. *Etica* 10(3), 37-45.

Gabrielatos, C. (2018). The lexicogrammar of BE interested: description and pedagogy. In *Corpora and Lexis* (pp. 240-276). Brill Rodopi.

Garavelli, B. (2001). Le parole e la giustizia. Divagazioni grammaticali e retoriche su testi giuridici italiani [Words and justice. Grammatical and rhetorical digressions on Italian legal texts]. Einaudi.

Goldberg, AE. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.

Goldberg, AE. (2019). *Explain Me This*. Princeton University Press.

Gries, S. Th. (2013). Data in construction grammar. In T. Hoffmann and G. Trousdale (Eds.), *The Oxford handbook of Construction Grammar* (pp. 93-108). Oxford University Press.

Halliday, M.A.K. (1991) Corpus studies and probabilistic grammar. In Aijmer, K. & Altenberg, B. (eds.), English Corpus Linguistics: Studies in honour of Jan Svartvik. London: Longman. 30-40.

Hilpert, M. (2013). *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax* (Studies in English Language). Cambridge University Press.

Hoffman, T., & Bergs, A. (2018). A construction grammar approach to genre. *CogniTextes* (18), (open access version https://journals.openedition.org/cognitextes/1032).

Mori, L. (2019). Complessità sintattica e leggibilità. Un monitoraggio linguistico per la valutazione dell'accessibilità dei testi legislativi europei e italiani [Syntactic complexity and readability. Linguistic monitoring for the assessment of the accessibility of European and Italian legislative texts]. *Studi Italiani di Linguistica Teorica e Applicata* [Italian Studies of Theoretical and Applied Linguistics] (48), 627-657.

Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, *7*(1), 140-162.

Rossini-Favretti, R. (2000). Progettazione e costruzione di un corpus di italiano scritto: CORIS/CODIS [Design and construction of a written Italian corpus: CORIS / CODIS]. In R. Rossini-Favretti (ed.), *Linguistica e informatica. Multimedialità, corpora e percorsi di apprendimento*[Linguisticsand Computer science: Multimediality, corpora, and learning](pp. 39-56). Bulzoni. URL:http://corpora.dslo.unibo.it/TCORIS/

Stefanowitsch, A. (2013). Collostructional analysis. InT. Hoffmann & G. Trousdale (eds.), *TheOxford handbook of construction grammar*(pp.1-16). OxfordUniversity Press.

The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries,University of Oxford, on behalf of the BNC Consortium. URL:http://www.natcorp.ox.ac.uk/

Tiersma, P. M. (1999).*Legal language*. University of Chicago Press.

Van Boom,W. H., Desmet, P., & Van Dam, M. (2016). "IfIt's Easy toRead, It's Easy to Claim"—The Effect of the Readability of Insurance Contracts on Consumer Expectations and Conflict Behaviour. *Journal of Consumer Policy*, 39(2), 187–197.

An Exploration into the Relationship Between Nominal Form and Temporal Semantics

Alex Carr

Centre for Language and Communication Research ENCAP, Cardiff University

Carra1@cardiff.ac.uk

Traditionally, nominals have been semantically associated with notions of 'time stability', considered to typically denote objects which do not belong to the temporal domain, e.g. table (Langacker 1991:298; Givon 2001:51). Nevertheless, nominals have proven the capability to express temporal meaning (Vendler 1967). For example, the nominal fire construes a dynamic event, as a fire can 'take place', and 'be observed over time' (Vendler 1967:141). Recent research into the temporal semantics of nominals has identified that particular deverbal nominalizations (Balvet et al 2011), deadjectival nominalizations (Arche and Marín 2014) and underived nominals (Huyghe et al 2017) can express temporal meaning. However, while studies have largely focused on the identification of temporal meaning in different nominal forms, significantly less work has been devoted to the examination of the specific character of these temporal meanings in relation to nominal form. Thus, the aims of this research are twofold: (1) to evaluate the extent to which different nominal forms vary in relation to temporal semantic behaviour, and (2) to determine how object, state, and event meaning come to be expressed in nominal form.

Using part-of-speech tagging, a random sample of 5000 nominal instances was extracted from the British National Corpus (BNC). To avoid ambiguity in the coding process, the nominals were first analysed for their function in use, using a Systemic Functional Linguistic 'experiential analysis' (Halliday 2014). Nominal instances which did not function in the syntactic head position of the nominal group were discarded. Repetition instances were also discarded from the sample, to circumvent potential inflation of relationships between certain nominal forms and temporal semantic distributions. The removal of the above mentioned instances left 1664 nominals instances in the sample. These nominal instances were analysed for 'Lexical Aspect' (i.e. Aktionsart, Ontological/Situation Aspect) (summarised in Smith 1991; Declerck 2006), count/mass status and abstract/concrete status. This process involved applying diagnostic syntactic tests to the nominals (Vendler 1967; Dowty 1979; Smith 1991; Declerck 2006), e.g. did NOUN take place?. Six types of lexical aspect are associated with this research: 'State', 'Activity', 'Accomplishment', 'Achievement', 'Semelfactive' and 'Object'. Through consulting the Oxford English Dictionary (2020), the nominals were then further classified into eight categories based on their etymology: 'Borrowing'; 'Compound'; 'Morphologically derived from adjective' (MDA); Morphologically derived from noun' (MDN); 'Morphologically derived from verb' (MDV); 'Other'; 'Transcategorization' (TC), and 'Underived'. The results from this research indicate an overall significant difference between the distribution of temporal semantics expressed by the different word formation types. Most notably, MDAs, MDVs and TCs displayed a considerably stronger association with temporal semantics than the other word formation types. Despite this significant difference however, the results from this research also suggest that abstract/concrete status is the most influential variable overall in predicting the distribution of temporal semantics expressed by the nominal instances.

References

Arche, J, M. and Marín, R. (2014) Edgy nominalizations. In Proceedings of the 31st West Coast Conference on Formal Linguistics. Cascadilla Proceedings Project. 21-30.

Balvet, A., Barque, L., Condette, H, M., Haas, P., Huyghe, R., Marín, R. and Merlo, A. (2011) Nomage: an electronic lexicon of French deverbal nouns based on a semantically annotated corpus. In WoLeR 2011 at ESSLLI, International Workshop on Lexical Resources: 8-15.

Declerck, R., Reed, S. and Cappelle, B. (2006) The grammar of the English verb phrase, vol. 1. Berlin; New York: Mouton de Gruyter.

Dowty, R, D. (1979) Word meaning and montague grammar: the semantics of verbs and times in generative semantics and in Montague's PTQ. Dordrecht, Holland; London: D. Reidel.

Givón, T. (2001) Syntax: an introduction, Volume 1. Amsterdam, Philadelphia: John Benjamins.

Huyghe, R., Barque, L., Haas, P. and Tribout, D. (2017) The semantics of underived event nouns in French. Italian Journal of Linguistics 29(1): 117-142.

Langacker, W, R. (1991) Foundations of cognitive grammar. Vol 2, descriptive application. Stanford: Stanford University Press.

Oxford English Dictionary (2020) Oxford English dictionary (online). Available (22.11.2020) at https://www.oed.com/.

Smith, C. (1991) The parameter of aspect. Kluwer: Dordrecht. Vendler, Z. (1967) Linguistics in philosophy. Ithaca N.Y.: Cornell University Press.

Vendler, Z. (1967) Linguistics in philosophy. Ithaca N.Y.: Cornell University Press.

What the correlation between syntactic complexity and lexical development measures can tell us about modes of acquisition

Irene Checa-García

University of Wyoming

irene.checa@uwyo.edu

It is a common practice in the study of heritage language learners (HLL) to talk of how their language skills compare to those of L2 learners (L2L) and monolingual native speakers (MNS), particularly in the case of Spanish in the US. Most studies tend to concentrate in a specific construction, such as gender agreement or verb aspect when dealing with grammar (Montrul, Foote, & Perpiñán, 2008; Montrul & Perpiñán, 2011), or on accuracy or transfer when dealing with the lexicon (Fairclough & Garza 2018; Chávez, 2017). Few have dealt with more encompassing measures of lexical and syntactic abilities (Marqués-Pascual & Checa-García, Under Review; Checa-García & Marqués-Pascual, 2020; Abchi & De Mier, 2017; Dengub, 2012). These works have used lexical deployment measures that have been frequently -though not always- successful in characterizing lexical development in both L1 and L2, such as Lexical Density, Lexical Diversity, Lexical Sophistication, and Lexical Accuracy. The syntactic elaboration measures most frequently used in these works about heritage speakers have been the T-unit Length, the Index of Subordination, and a little less frequently the Clause length, all indexes introduced by Hunt (1965).

None of these works, however, have investigated if there is a relationship between lexical development and syntactic development measures. In this paper, I study the correlation between 4 lexical deployment measures: Density, Diversity of content words, Diversity of all words, and Sophistication, and 5 syntactic measures: number of T-units, Syntactic Errors per T-unit, T-unit length, Clause length, and Subordination Index, in three different groups: HLL, L2L, and MNS. The corpus analyzed consisted of one 350-450 words composition by participant (total of 90) with the same prompt that was then tagged for all syntactic units as well as lemmatized using CLAN, which also tagged parts of speech. For the sophistication, the Davies (2006) list of frequency words was used.

Preliminary results show very few significant correlations between syntactic and lexical measures in all three groups, that are usually not very high. However, those few present correlations show more similarities between L2L and HLL. There are also marked differences among the lexical indexes, such that diversity of content and diversity of all words are the most commonly correlated with syntactic measures in all three groups.

For L2L, having longer T-units means having more syntactic errors, and also less diversity in their content words. Similarly, for HLL more syntactic errors are correlated with more less diversity. Thus, those L2L and HLL which have less vocabulary have also lower syntactic development. This could be due to a parallel development of both abilities. However, while the L2L formal acquisition setting and its sequencing may easily explain this parallel development, for HLL, whose vocabulary tends to be more developed before arriving even to a classroom setting, this explanation does not seem sufficient. I propose that in the case of HLL with no formal grammar training, the development of grammar goes hand in hand with the development of vocabulary, giving support to the idea of how closely linked the two are.

References

Abchi, V. S., & De Mier, V. (2017). Syntactic complexity in narratives written by Spanish heritage speakers. *Vigo International Journal of Applied Linguistics*, 14, 125-148.

Chávez, Y. (2017). Lexical Exams in Spanish as Second Language and as a Heritage Language: A Tentative Guideline for Improvement. *International Multilingual Journal of Contemporary Research, 5*(1), 1-3. https://doi.org/10.15640/imjcr.v5n1a1

Checa-García, I. & Marqués-Pascual, L. (2020). Lexical Richness Measures in Spanish heritage, native and L2 learners. Paper presented at *49th LASSO Virtual Conference*.

Davies, M. (2006). *A frequency dictionary of Spanish: Core vocabulary for learners*. New York, NY: Routledge.

Dengub, E. (2012). *Investigating syntactic and lexical complexity, accuracy, and fluency in the writing of heritage speakers of Russian* (Doctoral dissertation, Bryn Mawr College).

Fairclough, M. & Garza, A. (2018). The lexicon of Spanish heritage language speakers. In Potowski, K. & Muñoz-Basols, J. *The Routledge Handbook of Spanish as a Heritage Language.* Routledge. https://doi.org/10.4324/9781315735139

Hunt, K.W. (1965). Grammatical structures written at three grade levels (NCTE Research Rep. No. 3). Champaign, IL: National Council of Teachers of English. (ERIC Document Reproduction Service No. ED 113 735)

Marqués-Pascual, L. & Checa-García, I. (Under review). Lexical Development of Spanish Heritage and L2 Learners in a Study Abroad Setting.

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language learning*, *58*(3), 503-553. https://doi.org/10.1111/j.1467-9922.2008.00449.

Montrul, S., & Perpiñán, S. (2011). Assessing differences and similarities between instructed heritage language learners and L2 learners in their knowledge of Spanish tense-aspect and mood (TAM) morphology. *Heritage Language Journal*, *8*(1), 90-133. https://doi.org/https://doi.org/10.46538/hlj.8.1.

A corpus investigation into lexical aspect as a lexicogrammatical feature of the intransitive construction

Lucy Chrispin

Cardiff University

chrispinl@cardiff.ac.uk

Lexical aspect (i.e aktionsart, ontological/situation aspect) concerns the construal of inherent temporal structures by situations (Vendler 1967; Smith 1991). Though it is widely acknowledged that the grammatical properties of an eventuality in a clause are essential in determining the lexical aspect (Smith 1991; Declerck et al. 2006), certain lexical verbs are commonly representative of specific situation types e.g. typical verbs of 'states' include be, have and resemble (Van Rompaey 2013, p.198). In terms of intransitives, literature has provided fundamental contributions on the link between the two intransitive verb types 'unergative' and 'unaccusative', and lexical aspect (Tenny 1987; Dowty 1991; van Gelderen 2018). For example, unaccusatives have been identified as telic whereas unergatives are atelic (Dowty 1991; van Gelderen 2018, p.10). However, unacusativity research appears to focus predominantly on the unaccusative side as opposed to unergatives (also known as 'pure' intransitives). Additionally, whilst the focus of lexical aspect has been predominantly directed towards verbs alone, less attention has been paid to whether the wider constructions denote specific lexical aspect categories themselves i.e. whether certain aspectual types can be considered a lexicogrammatical feature of pure intransitive constructions. Therefore, this paper aims to investigate the relationship between lexical aspect and constructions, with a focus on pure intransitive constructions in particular, in order to determine the variation of situation types within and between constructions.

This study was based on 10 verbs that commonly occur in pure intransitive constructions: sneeze, meditate, frown, converse, stare, resign, jog, walk, compete and climb. The verbs were selected using several resources (Levin 1993; Francis et al. 1996; Fontaine 2013; Hartmann et al. 2013; Bartlett 2014; Banks 2015), then used in a query search in the Corpus of Contemporary American English (COCA; Davies 2008-) to obtain the wider constructions. A total of 2500 concordance lines (250 lines per verb) were analysed using Hanks' (2004) Corpus Pattern Analysis (CPA), an approach that reveals how meaning is mapped onto verbal patterns (ibid, p.87). In CPA, patterns are grouped according to their syntactic structure and an associated meaning or 'implicature'. Each concordance was also analysed for lexical aspect (summarised in Van Rompaey 2013, pp.181–219); the five main situation types concerned with this research are activity, semelfactive, accomplishment, culmination (i.e. achievement) and state. In line with this study's aims, the patterns revealed from CPA were investigated in relation to their lexical aspect. Results reveal both homogenous patterns in terms of situation type, as well as intransitive patterns that have multiple situation types. Predominant influences on the lexical aspect of the constructions include verbal semantics, the head of the prepositional complement, the noun phrase in the prepositional complement, and adverbials. Overall, results reveal multiple intransitive patterns with prototypical aspectual meaning, and point towards an inherent prototypical meaning of the intransitive. This paper shows how a corpus approach can appropriately reveal the typical relationship between specific intransitive constructions and aspectual type.

References

Banks, D. 2015. On the (non) necessity of the hybrid category behavioural process. In: Bayley, P. ed. Hybridity in Systemic Functional Linguistics: Grammar, Text and Discursive Context. London: Equinox, pp. 21–40.

Bartlett, T. 2014. Analysing power in language: a practical guide. London and New York: Routledge.

Davies, M. 2008. The Corpus of Contemporary American English (COCA): 560 million words, 1990-present. Available at: https://corpus.byu.edu/coca/.

Declerck, R. et al. 2006. The Grammar of the English Verb Phrase. Volume I: The Grammar of the English Tense System. Berlin: Mouton de Gruyter. doi: 10.1515/9783110199888.

Dowty, D. 1991. Thematic Proto-Roles and Argument Selection. Language 67(3), pp. 547–619.

Fontaine, L. 2013. Analysing English Grammar: A Systemic Functional Introduction. Cambridge: Cambridge University Press.

Francis, G. et al. 1996. Collins COBUILD Grammar Patterns 1: Verbs. London: Harper Collins Publishers.

van Gelderen, E. 2018. The Diachrony of Verb Meaning. New York: Routledge.

Hanks, P. 2004. Corpus Pattern Analysis. In: Williams, G. and Vessier, S. eds. Proceedings of the 11th Euralex International Congress. Lorient, France: UBS, pp. 87–98.

Hartmann, I. et al. 2013. Valency Patterns Leipzig (ValPal). Available at: http://valpal.info.

Levin, B. 1993. English Verb Classes and Alternations: A Preliminary Investigation. Chicago: Chicago University Press.

Van Rompaey, T. 2013. The development of P + NP + of/to + V(ing) progressive aspect markers. Katholieke Universiteit Leuven.

Smith, C. 1991. The Parameter of Aspect (Studies in Linguistics and Philosophy Vol. 43). Dordrecht: Kluwer Academic Publishers.

Tenny, C.L. 1987. Grammaticalizing Aspect and Affectedness. massachusetts Institute of Technology.

Vendler, Z. 1967. Linguistics in philosophy. Ithaca [N.Y.]: Cornell University Press

Usage-based indices of proficiency:
Verb-VAC and phrasal constructions in Mandarin

Susanne DeVore

University of Hawaiʻi at Mānoa

srdevore@gmail.com

Syntactic complexity has long been studied as a key construct in language learner development (Wolfe-Quintero et al., 1998). Recent work using strength of association between verb-argument constructions and the verbs that fill them in a native speaker corpus as indices of learners' syntactic complexity has shown that they are more effective than general measures (ex: mean length of T-unit, clauses per T-unit, etc.) at predicting English learners' proficiency (Kyle & Crossley, 2017). Similarly, the strength of association between lexical items in specific constructions (ex: adj + N, V + Object) has been shown to distinguish between English learner levels (Paquot, 2019). In both cases, higher level learners use combinations that have higher strengths of association.

Previous work on Mandarin has found that the frequency of one type of phrase, resultative verb compounds, overall distinguished between intermediate-low, Intermediate-high, and Advanced learners and native speakers (Zhang, 2014). In another study, Zhang & Lu (2013) found that both high and low intermediate learners used more classifiers than native speakers, and both groups used the general classifier (个 ge) significantly more than native speakers.

Both Zhang (2014) and Zhang & Lu (2013) use target language norms as a point of comparison between groups. In the present study, I extend this by directly using the target language norms as indices for learner data, as Kyle & Crossley (2017) and Paquot (2019) do. Specifically, this study asks: To what extent does the strength of association between the lexical items in noun-classifier pairs and resultative verb compounds in an L1 reference corpus account for writing proficiency in learners of Chinese? And how does this compare with other syntactic, lexical, and lexico-grammatical indices?

References

HSK动态作文语料库. (n.d.).

Kyle, K., & Crossley, S. A. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. Language Testing, 1–23. https://doi.org/10.1177/0265532217712554

Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. Second Language Research, 35(1), 121–145. https://doi.org/10.1177/0267658317694221

Wolfe-Quintero, Kate., Inagaki, S., & Kim, H.-Y. (1998). Second language development in writing: Measures of fluency, accuracy, & complexity. Second Language Teaching & Curriculum Center, University of Hawaii at Manoa.

Zhang, J. (2014). A Learner Corpus Study of L2 Lexical Development of Chinese Resultative Verb Compounds1. Journal of the Chinese Language Teachers Association, 49(3), 1–24.

Zhang, J., & Lu, X. (2013). Variability in Chinese as a Foreign Language Learners' Development of the Chinese Numeral Classifier System. The Modern Language Journal, 97(S1), 46–60. https://doi.org/10.1111/j.1540-4781.2012.01423.x

Rhythm in World Englishes
A look at the interaction of phonology and lexico-grammar
from a corpus-based perspective

Sebastian Hoffmann
Trier University
hoffmann@uni-trier.de

This talk investigates the connection between stress and rhythm in World English. More specifically, it demonstrates that the impact of rhythmically different L1-contexts can be detected in the (written) output of institutionalised second-language varieties of English. In a second step, I will be asking the question whether these differences have an impact on the lexico-grammatical inventory of English world-wide.

It is a well-established fact that languages have rhythmic properties. Following Pike (1945) and Abercrombie (1965, 1967), languages have traditionally been categorised as stress-timed (e.g. English) or syllable-timed (e.g. Spanish); in addition, a number of languages have been classified as being mora-timed (e.g. Japanese, see e.g. Han 1962). More recent experimental research (e.g. Dauer 1983, 1987) has shown that these rhythmic classes are not clearly defined and that we are instead dealing with a continuum of rhythmic variation.

For English, there is a considerable body of research on what has been termed the Principle of Rhythmic Alternation ('PRA', Sweet 1876) – i.e. the general tendency to maintain an alternation of stressed and unstressed syllables. The bulk of this research is on written data (or on orthographically transcribed speech) and focuses on preferences in lexical or grammatical choice (e.g. *drúnken sáilor* instead of *drúnk sáilor*) or word ordering preferences (e.g. *compléte and únabridged* instead of *únabridged and compléte*) that are interpreted as resulting from stress-clash – or stress-lapse – avoidance strategies (see. e.g. Schlüter 2005; Shih 2017). Complementing this work, there is a growing body of corpus-based research in phonology assessing the status of metrical constraints on a more global scale. Based on simple bigram probabilities in a large variety of corpora comprising more than 10 million words, Breiss & Hayes (2020) show that metrically critical bigrams – i.e. phonetic contexts deemed less preferable by the PRA – are underrepresented in their data.

A number of varieties of English have been claimed to exhibit clear tendencies towards syllable timing (see e.g. the list in Mesthrie & Bhatt 2008: 129). For Singapore English, for example, this classification is supported in studies by Low and colleagues (e.g. Low 1998, Low & Grabe 1995, Low et al. 2000); for a book-length study of speech rhythm in acrolectal Indian English, see Fuchs (2016). All inner-circle varieties of English (cf. Kachru 1985), however, are said to be stress-timed. Given the differences between inner and many outer circle varieties of English, it can be expected that the PRA should apply to different degrees. By applying the approach taken by Breiss & Hayes (2020) to GloWbE corpus data (Davies & Fuchs 2015), I demonstrate that this indeed appears to be the case. However, some of my findings are inconclusive and further analysis is clearly needed to link these general variety-specific co-occurrence preferences to individual lexical and/or lexico-grammatical patterns.

This study is thus exploratory in nature. In addition to the linguistic results I present, I will focus on some methodological issues as well as the theoretical implications of my findings for the study of so-called native and non-native varieties of English.

References

Abercrombie, D. (1965) *Studies in Phonetics and Linguistics.* London: Oxford University Press.

Abercrombie, D. (1967) *Elements of General Phonetics.* Edinburgh: Edinburgh University Press.

Breiss, C. & Hayes, B. (2020) "Phonological Markedness Effects in Sentence Formation." *Language* 96(2), 338–370.

Davies, M. & Fuchs, R. (2015) "Expanding Horizons in the Study of World Englishes with the 1.9 Billion Word Global Web-based English Corpus (GloWbE)." *English World-Wide* 36(1): 1–28.

Dauer, R.M. (1983) "Stress-timing and Syllable-timing Re-analysed." *Journal of Phonetics*, 11: 51–62.

Dauer, R.M. (1987) "Phonetic and phonological components of language rhythm." *Proceedings of the XIth International Congress of Phonetic Sciences*, Tallinn, Estonia, 447–450.

Fuchs, R. (2016) *Speech Rhythm in Varieties of English: Evidence from Educated Indian English and British English*. Singapore: Springer.

Han, M. S. (1962) "The Feature of Duration in Japanese." *Onsei no kenkyuu*, 10: 65–80.

Kachru, B.B. (1985) "Standards, Codification, and Sociolinguistic Realism: The English Language in the Outer Circle." In R. Quirk & H.G. Widdowson, eds. *English in the World: Teaching and Learning the Language and the Literature*. Cambridge: Cambridge University Press. 11–30.

Low, E.L. (1998) *Prosodic Prominence in Singapore English.* Doctoral dissertation, University of Cambridge.

Low, E.L., & Grabe, E. (1995) "Prosodic Patterns in Singapore English." *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, 13 – 19 August: Vol. 3: 636 – 639.

Low E.L., Grabe, E. & Nolan, F. (2000) "Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English." *Language and Speech* 43(4): 377–401.

Mesthrie, R. and Bhatt, R.M. (2008) *World Englishes: The Study of New Linguistic Varieties*. Cambridge: Cambridge University Press.

Pike, K. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Schlüter, J. (2005) *Rhythmic Grammar: The Influence of Rhythm on Grammatical Variation and Change in English.* Berlin/New York, NY: de Gruyter Mouton.

Shih, S.S. (2017) "Phonological Influences in Syntactic Alternations." In V. Gribanova & S.S. Shih, eds. *The Morphosyntax-Phonology Connection*. Oxford: Oxford University Press. 223–252.

Sweet, H. (1876) "Words, Logic, and Grammar." *Transactions of the Philological Society, 1875-1876*: 470–503.

Categorization for Evaluation:
A corpus-based study on the use of *zhong* 'kind; type' with constructions of evaluation in Mandarin conversation

Chen-Yu Chester Hsieh

National Taiwan University of Science and Technology

chesterhugues@gmail.com

As categorization is an essential part of human cognition and language, linguistic devices that serve this function such as classifiers have received much attention, especially in languages like Mandarin Chinese (e.g., Tai, 1994; Huang & Ahrens, 2003; Her, 2012). While a majority of past studies treat categorization as a basic cognitive phenomenon (e.g., Rosch, 1978; Lakoff, 1987; Taylor, 1995), linguists have started to study it as an essential part of social interaction (e.g., Mayes & Tao, 2019; Tao, 2020). For example, Tao (2020) approaches categorizing in Mandarin conversation as social activities and suggests that categorizing activities in spoken Chinese demonstrate features of interaction and serve particular socio-interactional functions, including expressing subjectivity and achieving intersubjectivity.

Despite such insight, relatively little has been done to investigate the use of categorization for interactional purposes, especially in Mandarin Chinese. To contribute to this line of research, the current study analyzes the instances of the Chinese kind classifier *zhong* 'kind/type' retrieved the NCCU Corpus of Spoken Taiwan Mandarin (Chui, 2018), using the framework of Interactional Construction Grammar (ICxG) (Linell, 2009; Imo, 2015; Hsieh & Su, 2019), which combines Interactional Linguistics (Couper-Kuhlen & Selting, 2018) and Construction Grammar (Goldberg, 1995) to investigate the sequential patterns and interactional functions of form-meaning pairings in spoken discourse.

Results show that while *zhong* is found in a variety of patterns or constructions, most of the instances involve the public expression of evaluation and affective stance in one way or another (cf. Hunston, 2010; Hunston & Su, 2019). Most common patterns/constructions of evaluation in which *zhong* is utilized include [*shi*/*you* 'be/have' (+ *nazhong* 'that kind') + evaluation + *na*/*zhezhong* 'that/this kind' (+ NP)], [Subject + emotion verb + *nazhong* (+ NP)], and [*nazhong* (+ NP) + evaluation], as well as using the noun phrase [*nazhong* + NP] as an evaluative response, among others.

Moreover, the kind classifier *zhong* and the constructions in which it often occurs may not only enable speakers to express their assessment of a stance object (Du Bois, 2007), but also allow co-conversationalists to jointly construct their intersubjective stance (cf. Tao, 2020). Speakers may use *zhong* phrases to refer to a topic in the prior speaker's turn and express or imply their own (usually negative) evaluation. The second speaker can also use *zhong* phrases to echo and even intensify the first speaker's assessment. We argue that the evaluation function of *zhong* may arise from the generalization involved when categorization is made (Scheibman, 2007) and the construction in which it is deployed (cf. Schmid, 2000; Mahlberg, 2005).

This research demonstrates the interaction between a lexico-grammatical item and the co-occurring evaluation patterns in Chinese conversation and between cognition and social interaction. It thus holds implications for research on categorization, evaluation, and the interface between lexis, grammar and discourse.

References

Chui, K. (2018). *The NCCU Corpus of Spoken Taiwan Mandarin*. https://spokentaiwanmandarin.nccu.edu.tw.

Couper-Kuhlen, E., & Selting, M. (2018). *Interactional Linguistics: An Introduction to Language in Social Interaction*. Cambridge: Cambridge University Press.

Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (Ed.), *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*. Amsterdam: John Benjamins.

Goldberg, A. E. (1995). *Construction: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Her, O-S. 2012. Structure of classifiers and measure words: A lexical functional account. *Language and Linguistics* 13(6), 1211-1511.

Hsieh, C-Y. C. & L. I-W. Su*. (2019). Construction in conversation: An Interactional Construction Grammar approach to the use of *xiangshuo* 'think' in spoken Taiwan Mandarin. *Review of Cognitive Linguistics* 17(1), 132-155.

Huang, C-R., & Ahrens, K. (2003). Individuals, kinds and events: classifier coercion of nouns. *Language Sciences 25*(4), 353-373.

Hunston, S. (2010). *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. London: Routledge.

Hunston, S., & Su, H. (2019). Patterns, constructions, and local grammar: A case study of 'evaluation'. *Applied Linguistics*, 40(4), 567-593.

Imo, W. (2015). Interactional construction grammar. *Linguistics Vanguard,* 1(1), 69-77.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press.

Linell, P. (2009). Grammatical constructions in dialogue. In A. Bergs & G. Diewald (Eds.), *Contexts and Constructions*, (pp. 97-110). Amsterdam: John Benjamins.

Mahlberg, M. (2005). *English General Nouns: A Corpus Theoretical Approach.* Amsterdam: John Benjamins.

Mayes, P. and Tao, H. (2019). Referring expressions in categorizing activities: Rethinking the nature of linguistic units for the study of interaction. *Studies in Language* 43(2), 329-363.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Eds), *Cognition and Categorization* (pp. 27-48). Hillsdale: Lawrence Erlbaum.

Scheibman, J. (2007). Subjective and intersubjective uses of generalizations in English conversation. In R. Englebretson (Ed.), *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction* (pp. 111-138). Amsterdam: John Benjamins.

Schmid, H. (2000). *English Abstract Nouns as Conceptual Shells: From Corpus to Cognition*. Berlin: Walter de Gruyter.

Tai, James HY. (1994). Chinese classifier systems and human categorization. In M. Y. Chen & O. J. L. Tseng (Eds.), *In Honor of William S.-Y. Wang: Interdisciplinary Studies on Language and Language Change* (pp. 479-494). Taipei: Pyramid.

Tao, H. (2020). Categorizing activities in Mandarin conversation and some theoretical and language pedagogical implications. *Yuyan Jiaoxue yu Yanjiu* [Language Teaching and Linguistic Studies] 2020.1 (201): 1-12.

Taylor, J. R. (1995). *Linguistic Categorization: Prototypes in Linguistic Theory*. New York: Oxford University Press.

The case of German two-way prepositions

Daniel Jach

Southwest Jiaotong University

danieljach@protonmail.com

In case languages, case is assigned to nominal phrases by their heads, typically verbs assigning case to their arguments to indicate syntactic relations. In German prepositions, too, assign cases to their complements, normally either accusative or dative case independent of context of use. A notorious exception to this rule are the so-called two-way prepositions that occur with both accusative and dative case depending on context, as in (1).

(1)   a.   Ich stelle den Reis auf den-ACC Tisch
               "I place the rice onto the table."
        b.   Der Reis steht auf dem-DAT Tisch.
               "The rice is on the table."

The current study investigates what determines the case of two-way prepositions in a large corpus sample of authentic language use. The variation is often attributed to different underlying meaning construals (Langacker, 1999) but recent exploratory corpus studies propose that the variation in addition depends on multiple co-occurring contextual features and specific lexical items (Willems, De Cuypere, and Rys, 2018). A sample of 9.332 sentences containing two-way prepositions was extracted from a large corpus of present-day German (Goldhahn, Eckart, and Quasthoff, 2012), automatically lemmatized and parsed, and then annotated for relevant lexical and structural features such as individual prepositions and surrounding lexical items, collocation strength, syntactic role of the prepositional phrase, type of head and complement, form of the preposition, transitivity of the verb, and others. The annotated features were used as predictor variables in multivariate logistic regression and collostructional analyses to predict case (dependent variable) from context of use.

Contrary to expectations, the results of the regression analysis indicated that none of the structural features had much influence on case, suggesting that the effects observed in the literature only hold for specific contexts. Instead, estimates of item-specific effects and the results of the collostructional analysis indicated strong associations of accusative and dative with individual prepositions and specific lexical items in the context. Case varied with verbs from two larger meaning clusters (accusative with telic verbs such as *stellen* "to place", *legen* "to put", and *setzen* "to sit down" vs. dative with atelic verbs such as *stehen* "to stand", *liegen* "to lie", and *sitzen* "to sit"). Moreover, accusative and dative were often used as parts of relatively fixed lexical verb-preposition strings (e.g., *sich freuen über* ACC "be happy about", *jemanden vor etwas*-DAT *warnen* "to warn somebody about something"). Last, dative frequently occurred in a handful of fixed preposition-(pro)noun strings (e.g., *vor allem*-DAT "especially", *in der*-DAT *Regel* "as a rule").

Framed in terms of usage-based construction grammar (Diessel, 2019), this is interpreted as a network of item-specific constructional prototypes that emerge from usage patterns and, once established, determine case based on form-meaning overlap with the current context of use. Frequent and invariant strings are stored as lexical wholes. At this level of representation case markers are little more than forms that complete routinized idiomatic patterns. More schematic constructions are gradually build up as language users generalize across lexical strings of similar form and meaning. From the two verb meaning clusters two

lexical prototypes emerge which gradually break away from individual verbs and become available for meaning construal in other contexts.

References

Diessel, H. (2019). The Grammar Network: How Linguistic Structure is Shaped by Language Use. Cambridge, England: Cambridge University Press.

Goldhahn, Dirk, Thomas Eckart, and Uwe Quasthoff (2012). "Building Large Monolingual Dictionaries at the

Leipzig Corpora Collection: From 100 to 200 Languages." In: Proceedings of the 8th International Language Resources and Evaluation (LREC'12). Ed. by Nicoletta Calzolari et al. Istanbul, Turkey: European Language Resources Association (ELRA), pp. 759–765.

Langacker, Ronald W. (1999). "Assessing the cognitive linguistic enterprise". In: Cognitive Linguistics: Foundations, Scope, and Methodology. Berlin, Germany: De Gruyter, pp. 13–59.

Willems, Klaas, Ludovic De Cuypere, and Jonah Rys (2018). "Case alternation in argument structure constructions with prepositional verbs: A case study in corpus-based constructional analysis". In: Constructional Approaches to Syntactic Structures in German. Ed. by Hans C. Boas and Alexander Ziem. Berlin, Germany: De Gruyter, pp. 85–130.

Multi-Dimensional analysis of literary style from a lexicogrammatical perspective

Carlos Kauffmann
São Paulo Catholic University
chkauffmann@corpuslg.org

In the area of Corpus Linguistics, studies that employed Multi-Dimensional (MD) analysis initially focused on the grammatical end of lexicogrammatical continuum (Berber Sardinha, 2020) to identify co-occurrent morpho-syntactic categories interpreted as dimensions of register variation (Biber, 1988). More recently, a new lexis-based MD framework has been introduced (Crossley & Louwerse, 2007; Berber Sardinha, 2019; Zuppardi, 2020), which uses lexical units as data. As a result, two major approaches to MD analysis exist: the grammatical MD analysis providing the grammatical dimensions, and the lexical MD analysis providing the lexical dimensions. As far as I am aware, just one MD study has attempted to integrate grammatical and lexical dimensions (Mayer, 2018), enabling a cline between lexis on one hand and structure on the other.

In this paper, I present the results of a study whose goal is to take a lexicogrammatical approach to MD analysis that brings together grammatical dimensions and lexical dimensions, thereby providing a holistic view of the use of lexicogrammatical resources in a corpus of literary texts. Most MD studies have characterized literary style from a grammatical perspective (Opas, 1992; Biber & Finegan, 1994; Egbert, 2012; Ali & Ahmad, 2016), and therefore a description of the full spectrum of lexicogrammatical resources used by literary authors is missing in the MD literature. In order to carry out the research, a corpus of novels and short stories by the most celebrated Brazilian fictional author, Machado de Assis (1839–1908), was collected. The corpus (called CLIMA) comprises 9 novels and 76 short stories, totaling 859,521 words. It was tagged for part of speech with the PALAVRAS parser, and lemmatized with TreeTagger for Portuguese. For reasons of space, readers are referred to Berber Sardinha and Veirano Pinto (2019) for a description of the methods used for both the grammatical and lexical MD analyses.

The grammatical MD analysis resulted in five dimensions: Abstract discourse *vs*. Orality, Narration, Hedging, Contextual *vs*. Dramatic discourse, and Thought presentation *vs*. Ornamentation. And the lexical MD analysis resulted in nine lexical dimensions: Sentimental expression, Romantic reference, Condition of man, Social representation, Urban setting, Patriarchal authority, Opposition, Metalanguage, and Uncertainty *vs*. Epistolary.

The integration of these grammatical and lexical dimensions was obtained through canonical correlation analysis (Afifi, May, & Clark, 2012; Mayer, 2018), which identified four significant canonical correlations, three of which were interpreted as aesthetic dimensions (Kauffmann & Berber Sardinha, 2021). Dimension 1, Formal Introspective Romanticism, emphasized hedging by the use of existence and modal verbs in tandem with romantic vocabulary (e.g., *heart*, *love*). Dimension 2, Sentimental Oralized Narrative, combined action verbs and conjunctions, which injected orality to the discourse, with lexis related to body parts and patriarchal authority (e.g., *sir*, *man*). Dimension 3, Dramatization, united thought presentation (mental verbs, first personal pronouns), orality (discourse markers), and metalinguistic vocabulary (e.g., *chapter*, *write*).

In conclusion, this study offered a lexicogrammatical-based interpretation about the unique style of Machado de Assis through Corpus Linguistics procedures that could be used as a blueprint for future stylistic studies about other authors, in different languages.

References

Afifi, A., May, S., & Clark, V. A. (2012) Practical Multivariate Analysis (Fifth ed.). CRC Press.

Ali, S. & Ahmad, S. (2016) Discourse style variation among the leading novelists of Pakistani fiction in English: A Multidimensional Analysis. ELF Annual Research, 18, 227-246.

Berber Sardinha, T. (2019) Using Multi-Dimensional Analysis to detect representations of national cultures. In T. Berber Sardinha & M. Veirano Pinto (Eds.), Multi-Dimensional Analysis: Research methods and current issues (pp. 231-258). Bloomsbury.

Berber Sardinha, T. (2020) Lexicogrammar. In C. A. Chapelle (Ed.), The Concise Encyclopedia of Applied Linguistics (pp. 701-706). Wiley Blackwell.

Berber Sardinha, T. & Veirano Pinto, M (Eds.), Multi-Dimensional Analysis: Research methods and current issues (pp. 231-258). Bloomsbury.

Biber, D. (1988) Variation across speech and language. Cambridge University Press.

Biber, D. & Finegan, E. (1994) Multi-dimensional analyses of authors' styles: some case studies from the 18th century. Research in Humanities Computing, 3, 3-17.

Crossley, S. & Louwerse, M. M. (2007) Multi-dimensional register classification using bigrams. International Journal of Corpus Linguistics, 12(4), 453-478.

Egbert, J. (2012) Style in nineteenth century fiction: A Multi-Dimensional analysis. Scientific Study of Literature, 2(2), 167-198.

Kauffmann, C. & Berber Sardinha, T. (2021) Brazilian Portuguese literary style. In E. Friginal & J. A. Hardy (Eds.), The Routledge handbook of corpus approaches to discourse analysis (pp. 354-375). Routledge.

Mayer, C. (2018) O que e como escrevemos na Web: Um estudo multidimensional de variação de registro em língua inglesa [What and how we write on the Web: A multi-dimensional study of register variation in English]. (Ph.D. dissertation.) Graduate Program in Applied Linguistics, São Paulo Catholic University, São Paulo.

Opas, L. L. (1992) A multidimensional analysis of style in Samuel Beckett's prose works. Research in Humanities Computing, 4, 81-14.

Zuppardi, M. C. (2020) Collocation dimensions in academic English. (Ph.D. dissertation.) Graduate Program in Applied Linguistics, São Paulo Catholic University, São Paulo.

A lexicogrammatical approach to participant role alternations in the French Spending frame

James Law
Brigham Young University
jimlaw@byu.edu

This study presents a diachronic corpus analysis of a semantic role alternation observed in the valence patterns of 28 French lexical items related to spending (e.g. *coût* 'cost', *dépenser* 'spend'). The analysis is grounded in Construction Grammar, which views grammar and lexicon as a continuum of form-meaning pairings of varying specificity known as constructions (Fillmore 1988). The constructional alternation observed in the Spending frame allows the participant role of Goods (e.g. *the cost of **materials***) to be replaced by other roles: the Purpose (e.g. *the cost of **the operation***), the Seller (e.g. *the cost of **suppliers***), or the Cause of Expense (e.g. *the cost of **the delays***). These roles index each other and their alternation often reflects a difference of focus rather than a difference of denotation.

In Halliday's (1961) view of a lexicogrammatical continuum, we would expect this alternation to be affected or constrained by traits specific to the lexical items heading the construction. For example, we find more frequent realization of the Purpose in constructions involving *débourser* 'spend' than *coûter* 'cost' because these words provide different perspectives on the Spending scenario and usually take different types of complements. We might also expect shifts in these patterns over time, as usage impacts constructional probabilities (Halliday 1991).

From the Frantext and MCVF diachronic corpora (ATILF - CNRS & Université de Lorraine, Martineau 2008), I draw a balanced data set of 1,629 tokens representing 28 French lexical items of the Spending frame from the 12th-20th centuries. Annotation of participant roles and statistical analysis using multinomial logistic regression reveals that replacement of Goods in constructional slots by Purpose declined over time while replacement by Cause of Expense increased. Replacement by Seller is consistently rare across the data set.

The decline in Purpose realization is tied to a number of lexical factors, including a decline among the nouns in prepositional support constructions such as *à grand coût* 'at great cost' and a corresponding increase in the frequency of genitive constructions such as *le coût de* 'the cost of'. The former allow nouns to modify a VP representing the Purpose, while the latter take as their complement an NP more frequently representing the Goods. The rise in Cause of Expense realization is related to lexical items that indicate negative sentiment towards the spending scenario, such as *onéreux* 'pricey', affecting other lexical items of the frame. Both shifts are also related to higher levels of Goods replacement in the valence patterns of low-frequency items. Text genre likewise impacts the use of this alternation. For example, replacement of the Goods by the Cause of Expense occurs more frequently in narrative genres where the sentiment of characters is especially relevant.

This case study describes an under-documented type of grammatical alternation that is relevant to other semantic domains and languages. The narrow focus on the French Spending frame demonstrates the degree to which usage of this alternation is lexically variable. Such constructions are thus best analysed under a lexicogrammatical perspective where a number of semantic, textual, lexical, and grammatical features are considered together.

References

ARTFL-Frantext corpus. (2019) ATILF - CNRS & Université de Lorraine. http://www.frantext.fr.

Halliday, M. A. K. (1961) Categories of the Theory of Grammar. WORD, 17(2), 241–292.

Halliday, M. A. K. (1991) Towards probabilistic interpretations. In Eija Ventola (ed.), Functional and Systemic Linguistics : Approaches and Uses, Berlin: De Gruyter, 39–61.

Martineau, France (2008) Un Corpus pour l'analyse de la variation et du changement linguistique. Corpus, 7(1), 135–155.

Exploring the development of 'verb 4-gram' sequences and grammar patterns in L2 writing: A comparison study of L1 Korean speakers and other L1 speakers

Joyce Lim

University of Cambridge

dol20@cam.ac.uk

Usage-based approaches (Ellis *et al.*, 2016) claim that the human's ability to construct language can be better understood by observing our cognitive sensitivity to frequency (Tyler and Ortega, 2016). This notion has made a significant contribution to the learner corpus approach to identify candidate 'constructions' and their varying levels of abstractions. The current study compared the verb sequences that shape the transition from beginner (A2) to near-native proficiency (C2) in a corpus of essays written by L1 Korean speakers (n=1,973; 734,300 words) and other L1 speakers (n=231,701; 53,819,650 words; 121 nationalities) from the Cambridge Learner Corpus. Using a bottom-up approach, I observed the frequency and distribution of 4-slot verb sequences in the English essays contributed between 1992 and 2011. In the analyses, it was found that across all proficiency levels in both the L1 Korean and other L2 speaker sub-corpora, there were 'core' sequences (e.g. VV DT NN IN, VV IN DT NN, VVN IN DT NN) that have already been acquired at the beginning level and remain stable in essays during the transition towards near-native-like proficiency. Furthermore, six 'emerging' sequences were identified, which are sequences that gradually start to surface, making it to the top 10 ranking at the C2 level.

From these 'core' and 'emerging' sequences, the grammar pattern approach (Hunston & Francis, 2000; Hunston 2019) was used to conduct case studies to regroup the identified patterns into broader categories. By drawing on the taxonomy of pattern grammar, I was able to observe the dynamic restructuring process of learners at each proficiency level and also make generalizations about how there is a transition towards more variety of complex 'constructions' and a wider group of meanings that become more prevalent in higher levels within the same verb sequence. Furthermore, it was clear that even within the CEFR level, the learner language is impacted by the L1, and that this was especially prevalent in the lower levels. However, even in the highest level, C2, there were noticeable differences in the use of verb sequences between L1 Korean speakers and other L2 speakers.

Overall, this study offers insight into the residual acquisition of form and meaning and demonstrates that the learners' implicit adjustment of their production is impacted by their mother tongue at all levels.

References

Ellis, N., Römer, U., & O'Donnell, M. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar.* Oxford: Wiley.

Hunston, S. (2019). Patterns, constructions, and applied linguistics. *International Journal of Corpus Linguistics, 24*(3), 324-353.

Hunston S. & Francis, G. (2000). *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English.* Amsterdam/Philadelphia, PA: John Benjamins.

Tyler, A., & Ortega, L. (2016). Usage-based approaches to language and language learning: An introduction to the special issue. *Language and Cognition, 8*, 335-345.

Part-of-speech patterns in research introductions:
A cross-disciplinary study

(Kathy) Ling LIN (Shanghai Jiao Tong University) & Ming LIU (Hong Kong Polytechnic University)

kathyll@sjtu.edu.cn, ming1.liu@polyu.edu.hk

This study applies the Part-of-Speech-gram (PoS-gram) procedure to the examination of language patterning and variability in a largely conventionalized part-genre (i.e., research introductions). 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. 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).

Based on 400 article introductions of computer engineering (CE) and cognitive linguistics (CL) collected from AntCorgenGen 1.1.2 (Anthony, 2019), the study has identified key PoS-grams and their associated lexico-grammatical frames, using the written academic component of British National Corpus as the reference corpus. In the identification and concordance search of key PoS-grams, Sketch Engine with their modified English TreeTagger PoS tagset was adopted (Kilgarriff et al., 2014).

Findings are summarized as follows. First of all, the PoS-grams with high keyness scores have been successfully identified for introductions of both disciplines, with their representative lexicogrammatical frames and phraseologies highlighted, which has empirically validated the phraseological tendency and idiomaticity of language use in academic genres (Sinclair, 1996). Second, the analysis reveals key PoS-grams shared in CE and CL introductions, e.g., those associated with the step "purposive announcement" (viz., *IN DT JJ NN VBD TO* and *DT JJ NN VBD TO VV*), as well as the discipline-specific ones such as the PoS-gram for structure-outlining only found in CE introductions (viz., *DT NN VBZ VVN RB VVZ*). In addition to identifying sets of characteristic lexicogrammatical frames and phraseologies that could be directly transformed into EAP pedagogical input, the PoS-gram analysis has also helped revealing contrasting language styles in introductions of the two disciplines. The apparently more compact language use has been noted in CE introductions than in CL introductions, as evidenced in the total absence of the *that*-clause but the strong presence of the *to*-infinitive clause and the prepositional phrase instead in tokens of top-ranking key PoS-grams identified in CCE. Contrastingly, in CCL, the use of the *that*-clause is far more frequent, e.g., three out of the four key PoS-grams for realizing the step of topic summarization do contain it. The more compressed language style of academic introductions in CE could also be perceived from the particularly intensive use of the construction "noun +noun(+noun) …" as well as the pre-modifications and/or post-modifications of noun phrases in them.

Compared to various forms of multi-word sequences like n-grams, the PoS-gram has the unique strength of grouping phraseologies with similar or identical structure and discursive functions and yet either recurrent or varying lexical choices under the co-selected grammatical categories.

The advantage enriches analyses and helps yield pedagogically useful findings, in that patterning and variability is revealed not only in the overall function, structure and composition of PoS-grams but in such aspects of their recurrent or diversified tokens. This study illustrates the innovative application of corpus-based PoS-gram procedure to academic genres, which may inspire a promising new line of inquiry and the current genre pedagogy.

Acknowledgements

References

Anthony, L. (2019). AntCorGen (Version 1.1.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Clear, J. (1993). From Firth principles: Computational tools for the study of collocation. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 271-292). Amsterdam: John Benjamins.

Kilgarriff, 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.

Pinna, A., & Brett, D. (2018). Constance and variability: Using PoS-grams to find phraseologies in the language of newspapers. In J. Kopaczyk, & J. Tyrkkö (Eds.), *Applications of pattern-driven methods in corpus linguistics* (pp. 107-130). Amsterdam: John Benjamins.

Sinclair, J. M. (1996). The search for units of meaning. *Textus, 9*(1), 75-106.

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.

Investigating French Interference in Algerian Students' English-Arabic / Arabic- English Translations of Collocations

Souhaila Messaoudi, Claire Brierley, James Dickins
University of Leeds
{mlsme, C.Brierley, J.Dickins}@leeds.ac.uk

This study investigates the interference of French as Second Language (SL) on English as Foreign Language (FL) for 89 Algerian translation students in their English Arabic/ Arabic English translation of test sentences involving adjective+ noun collocations. Based on existing definitions from (Sinclair, 1991; Evert, 2008; McEnery and Hardie, 2012) and given the adopted corpus-based and lexico-grammatical approach, we define collocations as follows: the tendency of lexical items of a particular class to significantly and exclusively re-appear in the company of another word within specific grammatical patterns at a specified proximity in a given corpus data.

The study, not only analyses translation between two languages (Arabic and English) but also focuses on interference from a third language (French) which is seldom covered in existing literature. It investigates how these students: (i) render into Arabic, English collocations involving adjectives which are themselves false friends with corresponding French adjectives; (ii) translate Arabic collocations, the nodes of which are synonymous nouns in Arabic to English nouns that are false friends with French. These English nouns, as is the case with the adjectives, are obtained from a compiled list of false friends between English and French (Thody and Evans, 1985). The list categorises the false friends into weak and strong based on whether they share some meaning/s or not respectively between the two languages. This research uses strong false friends as they will clearly reveal instances of French interference.

The research adopts a mixed-methods approach in which both a self-reporting questionnaire, adapted from other studies (Magno, 2009; Ahmed, 2012); and a devised 30-item translation of two parts, English and Arabic collocations, have been used. To extract both frequent and exclusive general English and Arabic collocations, we used English Web 2015 (enTenTen15) and Arabic Web Corpus (arTenTen) which are both available through the Sketch Engine tool (Kilgarriff et al, 2014).

In order to extract the English collocations, twenty strong adjectival false friends were explored. For each of the nodes in the list, the best noun collocate was found in the top ten collocates in both the Log-Likelihood and the Log Dice score lists in Sketch Engine. Function words, technical terms, and punctuation were considered in the analysis. The same process was replicated to form the Arabic collocations but with ten focal nodes representing synonymous nouns, in Arabic, to English nouns which are strong false friends with French.

The preliminary analysis reveals that French interference is more likely to occur when translating from English to Arabic rather than vice versa. This can be demonstrated through the better performance in translating Arabic collocations into English than translating from English to Arabic. The reasons for Algerian students 'collocational errors were lexical to a greater extent and grammatical to a lesser extent, which showcases how collocation depicts the inseparability of lexis and grammar even if they do not contribute equally to lexical cores (Gabrielatos, 2019).

References

Ahmed, Z. (2012). *English lexical collocation knowledge of Libyan university students*. PhD. thesis, Bangor University. [Accessed 1November 2017and thereon]. Accessible from: http://e.bangor.ac.uk/4538/1/590646.pdf.

Evert, S. (2008). Corpora and collocations. *Corpus linguistics*. *An international handbook*, 2, pp.1212-1248.

Gabrielatos, C. (2019). *Lexicogrammar: Lexical Grammar or Construction Grammar? Two corpus-based case studies*. Invited talk, UCREL Research Seminar, Lancaster University, 25 April 2019. [Accessed 1February 2020]. Available from: https://research.edgehill.ac.uk/ws/portalfiles/portal/20342065/Gabrielatos.UCREL.2019.pdf.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. and Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), pp.7-36. [Accessed 1November 2017 and thereon]. Available from: https://www.sketchengine.eu/.

Magno, C. (2009). Assessing the level of English language exposure of Taiwanese college students in Taiwan and the Philippines. *Asian EFL Journal*. Accessible from: https://www.asian-efl-journal.com/main-journals/assessing-the-level-of-english-language-exposure-of-taiwanese-college-students-in-taiwan-and-the-philippines/.

McEnery, T. and Hardie, A. (2012). *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Thody, P., Evans, H. (1985). *Faux amis and key words: A dictionary-guide to French language, culture and society through lookalikes and confusables*. London: The Athlone Press.

Nativised Light Verb Constructions in Malaysian English

Christina Ong Sook Beng
Universiti Tunku Abdul Rahman, Malaysia
ongsb@utar.edu.my

This study aims to investigate the structural and functional patterns of light verb constructions (LVCs) in Malaysian English, i.e. combinations of a semantically "empty" verb (e.g. GIVE, TAKE, and MAKE) with a typically isomorphic noun derived from the corresponding verb (e.g. look); see for example (1) and (2):

(1) She has to **take a look** at her biography again.

(2) Young entrepreneurs are afraid to **make unpopular decision**.

Due to their semantically light nature, LVCs are interchangeable with their corresponding simplex forms (Hoffmann, Hundt, & Mukherjee, 2011; Mehl, 2017). To detect potential grammatical innovations, I created a general corpus consisting of threads from Lowyat.NET, a popular Internet forum in Malaysia and used the British National Corpus (BNC) as the reference corpus. Most researchers (Wierzbicka, 1982; Algeo, 1995; Dixon, 2005) exclusively focus on LVCs of the format 'verb + indefinite article + isomorphic deverbal noun'. However, I take a more inclusive approach and also consider structures without articles and derived nouns to be relevant for my analysis – cf. example (2) above.

The structural analysis reveals that Malaysians tend to overuse non-isomorphic nouns, zero article LVCs and those taking determiners as well as descriptive adjectives (e.g. 'unpopular' in example 2) in the modifier slots. The functional analysis shows that most Malaysian English LVCs are atelic; this is due to a great number of LVCs being used in declarative and interrogative forms. The latter finding contradicts hypotheses of related work which suggest that the function of LVCs is mainly to convert aimless actions into achievements (Leech, Hundt, Mair, & Smith, 2009; Wittenberg & Levy, 2017; Bonial & Pollard, 2020).

In comparison with the BNC data, TAKE and MAKE LVCs are more dominant in the Malaysian English corpus. To identify nativised LVCs in Malaysian English, I use the log-likelihood test to evaluate the difference between each structural and functional LVC pattern in the two corpora. Nativised LVC structures are defined as those where the log-likelihood scores are greater than 100. This is the case for most zero article LVCs. Their high frequency could be explained by the absence of articles in the substrate languages (Malay, Chinese, and various Indian languages). As for functional patterns, nativisation can be shown for TAKE_care, MAKE_report, and MAKE_decision LVCs that indicate goalless activities and incomplete tasks; their log-likelihood scores range from 130 to 1976. In contrast to the findings of past studies, the function of most Malaysian English LVCs is atelic. This corpus study thus highlights the flexible nature of LVCs and the resulting patterns of nativisation in non-native varieties of English.

References

Algeo, J. (1995). Having a look at the expanded predicate. In B. Aarts & C. F. Meyer (Eds.), *The verb in contemporary English: Theory and description* (pp. 203–217). Cambridge University Press.

Bonial, C., & Pollard, K.A. (2020). Choosing an event description: What a PropBank study reveals about the contrast between light verb constructions and counterpart synthetic verbs. *Journal of Linguistics, 2020*, 1–24. https://doi:10.1017/S0022226720000109

Dixon, R. M.W. (2005). She gave him a look, they both had a laugh and then took a stroll: Give a verb, have a verb and take a verb constructions. In R. M.W. Dixon (Ed.), *A semantic approach to English grammar* (pp. 459–483). Oxford University Press.

Hoffmann, S., Hundt, M., & Mukherjee, J. (2011). Indian English – an emerging epicentre? A pilot study on light verbs in web-derived corpora of South Asian Englishes. *Anglia*, *129*(3–4), 258–280. https://doi.org/10.1515/angl.2011.083

Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). *Change in contemporary English: A grammatical study*. Cambridge University Press. https://doi.org/10.1017/CBO9780511642210

Mehl, S. (2017). Light verb semantics in the International Corpus of English: Onomasiological variation, identity evidence and degrees of lightness. *English Language & Linguistics, 23*(1), 55–80. https://doi.org/10.1017/S1360674317000302

Wierzbicka, A. (1982). Why can you have a drink when you can't *have an eat? *Language, 58*(4), 753–799. https://doi.org/10.2307/413956

Wittenberg, E. & Levy, R. (2017). If you want a quick kiss, make it count: How choice of syntactic construction affects event construal. *Journal of Memory and Language, 94*(2017), 254–271. http://dx.doi.org/10.1016/j.jml.2016.12.001

Serial verb *ʔaw* 'take' with instrumental meaning in Isaan:
A distinct construction?

Milntra Raksachat
University of Oregon
Email: milntrar@uoregon.edu

Contrary to previous hypotheses about the instrumental meaning of *ʔaw* 'take' in Thai and Lao serial verb constructions (SVCs), I argue that in the grammar of closely related Isaan (ISO: tts; Enfield 2002a) the instrumental meaning of *ʔaw* within an SVC belongs to a distinct 'instrumental' construction, and does not qualify as an instance of a more general SVC. The study relies on collexeme analysis within an Isaan corpus.

Serial verb constructions, according to one definition, comprise "a sequence of verbs which act together as a single predicate, without any overt marker of coordination, subordination or syntactic dependency" (Aikhenvald 2006: 1). When used in SVCs, some verbs like 'give', 'say', and 'take' have been claimed to function as grammatical markers that introduce causer, beneficiary, goal, theme, or instrument roles into the structure (Stine 1968; Lord 1973; Bickerton 1981; Sebba 1987). Previous discussions on SVCs in Tai-Kadai languages mostly concern two major languages: (Central or Bangkok) Thai and (Vientiane) Lao (Stine 1968; Thepkanjana 1986; Sudmuk 2005; Enfield 2002b; 2007), but not the Isaan variety of northeastern Thailand. Within this literature, there are proposals that SVCs with *ʔaw* 'take' and instrumental meaning are instances of a more generalized SVC that can also express purpose or motion/direction of an item. This is partly because the use of *ʔaw* in SVCs seems to correlate with multiple functions, e.g., from the fully lexical meaning 'take hold of' to grammatical 'instrumental' meaning. Previous analyses on Thai and Lao SVCs largely ignore corpus data and contextual uses which help clarify the lexical versus grammatical status of *ʔaw* and which augment understanding of syntactic and semantic structures of SVCs.

This study examines the syntactic properties and functions of serial verb clauses with *ʔaw* 'take' in Isaan from a Construction Grammar perspective (Langacker 1987; Goldberg 1995; Croft 2001; Diessel 2019). The goal is to determine whether there is a dedicated construction for the instrumental meaning in Isaan that involves serial verbs. By "dedicated construction", I mean an idiosyncratic syntactic pattern that in this case is partly phonologically filled, and that is learned as a whole unit used for a specific function. The data are extracted from the Spoken Isaan Corpus which I have been building since 2018 with an aim to provide a resource for establishing the similarities and differences between Isaan, Thai and Lao.

I specifically argue that the pattern *ʔaw* NP VP does not qualify as an instance of the "purposive" SVC nor of the "handling-dispatch" SVC. The analysis uses a combination of semantic tests, discourse analysis and analysis of corpus frequency patterns. I follow the logic of the distinctive collexeme analysis developed by Gries & Stefanowitsch (2004) to differentiate distinct SVC constructions that can involve *ʔaw*. The collocation analyses suggest that serial verb uses of *ʔaw* with instrumental meaning not only have a syntactic pattern distinguishable from the purposive and handling-dispatch constructions, but also exhibit different discourse-informational structure; the instrument participant is typically new or contextually non-recoverable information, which is not true for the handling-dispatch construction.

References

Aikhenvald, Alexandra Y. (2006) Serial verb constructions in typological perspective. In Robert M. W. Dixon & Alexandra Y. Aikhenvald (eds.), *Serial Verb Constructions: A Cross-Linguistic Typology*. Oxford: Oxford University Press.

Bickerton, Derek. (1981) *Roots of Language*. Ann Arbor: Karoma Publishers.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Diessel, Holger. (2019) *The Grammar Network: How Linguistic Structure is Shaped by Language Use*. 1st edn. New York: Cambridge University Press.

Enfield, Nicholas J. (2002a) How to define "Lao", "Thai", and "Isan" language? A view from Linguistic Sceince. *Tai Culture* 7(1). 62–66.

Enfield, Nicholas J. (2002b) Functions of 'give' and 'take' in Lao complex predicates. In Robert S. Bauer (ed.), *Collected Papers on Southeast Asian and Pacific Languages*, 13–36. Canberra: Pacific Linguistics.

Enfield, Nicholas. J. (2007) *A Grammar of Lao*. Berlin: Mouton de Gruyter.

Goldberg, Adele E. (1995) *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on "alternations." *International Journal of Corpus Linguistics* 9(1). 97–129. https://doi.org/10.1075/ijcl.9.1.06gri.

Langacker, Ronald W. (1987) *Foundations of Cognitive Grammar*. California: Stanford University Press.

Lord, Carol. (1973) Serial verbs in transition. *Studies in African linguistics*. Los Angeles: Dept. of Linguistics and the African Studies Center, University of California, Los Angeles 4(3). 269–296.

Sebba, Mark. (1987) *Syntax of Serial Verbs: An investigation into serialisation in Sranan and other languages*. Amsterdam: John Benjamins Publishing Company.

Stine, Philip Clare. (1968) *The Instrumental Case in Thai: A Study of Syntax and Semantics in a Generative Model.* University of Michigan Ph.D. Dissertation.

Sudmuk, Cholthicha. (2005) *The Syntax and Semantics of Serial Verb Constructions in Thai*. University of Texas at Austin PhD Dissertation.

Thepkanjana, Kingkarn. (1986) *Serial Verb Constructions in Thai*. University of Michigan PhD Dissertation.

A corpus-based study of affixation tendencies in related academic disciplines
A work-in-progress report

Piotr Twardzisz

University of Warsaw

p.twardzisz@uw.edu.pl

This work-in-progress report presents the results of an ongoing project focusing on morphologically complex words in English as used in academic texts in the humanities and social sciences. Morphologically complex words are affixed (derived) words. Such lexical items still remain understudied in applied linguistics. Our research question is whether relatively similar academic disciplines attract their own specific affixes and derivations. Also, we enquire whether it is possible to establish a stable lexical "core" characteristic of the humanities and social sciences. In this, we seek both morphological individuation as well as stability in relatively homogenous texts.

Numerous analyses have shown that textual features such as grammatical constructions, lexical bundles (Hyland 2008: 7), or phraseology (Vincent 2013: 44), vary systematically across academic disciplines (Cunningham 2017: 72). Individual disciplines frequently develop their own patterns of discourse which tend to depart from those found in general English (Montero-Fleta 2011: 4). As for vocabulary in academic discourse, the existence of some lexical core, common to a wide range of disciplines, has also been questioned by some scholars. The behaviour of individual lexical items has been claimed to vary across disciplines as for their range, meanings, collocations they enter or frequencies that they show (Hyland & Tse 2007: 235).

Our preliminary research establishes a certain amount of stability across disciplines involving affixed words. It seems inevitable that certain cores of complex words are used by individual authors for all kinds of purposes in specialist writing across disciplines (cf. Brezina & Gablasova 2015: 17). We indicate candidate affixes which constitute the core of lexical complexity of academic texts in the humanities and social sciences. Some high-frequency, but also – unpredictably – some low-frequency, affixed words appear to be common to more varieties. We also identify affixation types which are characteristic of only certain disciplines within the humanities and social sciences.

The data are obtained from the Corpus of Contemporary American English (COCA). In COCA's academic genre (ACAD), there are ten sub-divisions. For the purpose of this study, we have selected three: history, education and geography/social sciences. Given numerous search sub-strings and three discipline-based sub-corpora, the analyses result in high numbers of discipline-based wordlists. Cutting-off points for high- and low-frequency items are established on the basis of the overall numbers of word types for each affix in each sub-corpus. The items retained after manual cleaning of the raw-data enter lists of morphologically complex word types for each discipline studied. Quantitative similarities and differences regarding individual affixed words in different lists are identified. Objective criteria are tested for recognizing certain word types as characteristic of the stable lexical core.

Our findings have the potential of informing both theoretical and applied morphology. The former receives systematic data and insights for measuring morphological productivity. The latter is informed about morphological (ir)regularities applicable to academic writing across disciplines.

References

Brezina, V. & Gablasova, D. (2015) Is there a core general vocabulary? Introducing the *New General Service List*. Applied Linguistics, 36(1), 1-22.

Cunningham, K. J. (2017) A phraseological exploration of recent mathematics research articles through key phrase frames. Journal of English for Academic Purposes, 25, 71-83.

Hyland, K. (2008) *As can be seen*: Lexical bundles and disciplinary variation. English for Specific Purposes, 27, 4-21.

Hyland, K. & Tse, P. (2007) Is there an "Academic vocabulary"? TESOL Quarterly, 41(2), 235-53.

Montero-Fleta, B. (2011) Suffixes in word-formation processes in scientific English. LSP, Professional Communication, Knowledge Management and Cognition, 2(2), 4-14.

Vincent, B. (2013) Investigating academic phraseology through combinations of very frequent words: A methodological exploration. Journal of English for Academic Purposes, 12, 44-56.

Alternations Emerge and Disappear:
The network of dispossession constructions in the history of English

Eva Zehentner

University of Zurich

eva.zehentner@es.uzh.ch

This paper takes a diachronic construction grammar approach to changes in syntactic alternations in the history of English; specifically, it focusses on the question of how to model the disappearance of alternation relationships from a constructionist perspective. This is illustrated by zooming in on the history of a set of ditransitive verbs, viz. dispossession verbs like *steal* or *rob*, in comparison to the well-known English dative alternation.

In Present Day English, dispossession verbs are typically used in either of two prepositional constructions: in the 'prepositional deprivee construction' (PDC, 1a), the PP denotes the victim or source, while in (1b), a 'prepositional theme construction' (PTC), it is the theme that is marked by a preposition. Importantly, despite expressing similar events, the two constructions cannot be said to alternate in the strictest sense, as individual verbs do not vary between them (Levin 1993; Goldberg 1995). The constructions are furthermore associated with different prepositions (*from* vs *of*) and complementary object orders.

(1) a. They stole <u>cake</u> **from the students**.
     b. They robbed **the students** <u>of cake</u>.

However, interchangeability between the two prepositional patterns seems to have still been given in earlier stages of English (Visser 1963). Moreover, dispossession verbs could also be used in a double-object nominal construction in earlier English (DOC, 2), linking this verb class' history to the development of the English dative alternation. As shown in e.g. Colleman & De Clerck (2011), this nominal option has since been lost.

(2) For dronkenesse bireveth **hym** <u>the discrecioun of his wit</u>.
     'For drunkenness robs him [of] the discretion of his wit.'
     (CMCTPARS,316.C2.1212)

The precise development of this verb class and particularly the PP-patterns associated with it is nevertheless to-date largely unaccounted for. The present paper aims to address this lack by means of a quantitative corpus analysis of tokens of a set of dispossession verbs in the *Penn-Parsed Corpora of Historical English* (PPCME2, PPCEME, PPCMBE), covering a time span from 1150 to 1914.

I first show that DOC uses of dispossession verbs greatly decrease in Middle English already; in a second step, I then use Multiple Correspondence Analysis (Greenacre 2017) to investigate the overlap between the three construction types in terms of a range of features such as verb lemma, order, or preposition type, as well as semantic-pragmatic variables of the arguments like animacy or definiteness, and changes therein. The results suggest that (a) the loss of the DOC may have been motivated by the great functional similarity between this pattern and the PTC, and (b) that from Early Modern English onwards, PDC and PTC increasingly differentiate from each other, to a point where each pattern comes to be exclusively associated with particular verbs, prepositions and other features, and there is no variation to speak of anymore. I model these developments as changes in the network of dispossession constructions over time, giving particular attention to the emergence and loss of 'allostructional' relationships as outlined in

Cappelle (2006) and Perek (2015), as well as the concept of 'niche construction' (Traugott & Trousdale 2013).

References

Cappelle, B. (2006) Particle placement and the case for "allostructions". In Schönefeld, D. (ed.), *Constructions Special Volume 1 – Constructions all over: Case studies and theoretical implications*.

Colleman, T. & De Clerck, B. (2011) Constructional semantics on the move: On semantic specialization in the English double object construction. Cognitive Linguistics, 22(1), 183-209.

Goldberg, A. (1995) *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Greenacre, M. (2017) *Correspondence analysis in practice* (3rd edn.). Boca Raton, FL: Chapman and Hall.

Levin, B. (1993) *English verb classes and alternations: A preliminary investigation*. Chicago, IL: University of Chicago Press.

Kroch, A. & Taylor, A. (2000) Penn–Helsinki Parsed Corpus of Middle English, second edition. www.ling.upenn.edu/hist–corpora/PPCME2–RELEASE–3/index.html .

Kroch, A., Santorini, B. & Delfs, L. (2004) Penn-Helsinki Parsed Corpus of Early Modern English. www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-3/index.html .

Kroch, A., Santorini, B. & Diertani, A. (2010) Penn Parsed Corpus of Modern British English. www.ling.upenn.edu/hist-corpora/PPCMBE2-RELEASE-1/index.html .

Perek, F. (2015) *Argument structure in usage-based construction grammar: Experimental and corpus–based perspectives*. Amsterdam: Benjamins.

Traugott, E. & Trousdale, G. (2013) *Constructionalization and constructional changes*. Oxford: OUP.

Visser, F. (1963) *An historical syntax of the English language*. Leiden: Brill.

English *with/with-less-SubjPart₁* constructions:
A case of a quantitative corpus-based analysis

Victoria Zhukovska

Zhytomyr State Ivan Franko University, Ukraine

victoriazhukovska@gmail.com

This study is a quantitative corpus-based analysis aimed at identifying semantic and functional differences between two English alternative grammatical constructions (*with* and *with-less* Participle I clauses with the explicit subject). These syntactic patterns represent secondary predication of syntactically independent configuration, attached to the matrix clause by the augmentor *with* or asyndetically. In a sentence, the patterns perform the general syntactic role of an adverbial modifier elaborating, extending, or enhancing the matrix proposition. Regarding the form, the obligatory slots of the pattern are schematically represented as [øaug/aug][SBJ][PRED$_{PI}$].

Adopting the theoretical and methodological assumptions of usage-based construction grammar and quantitative corpus linguistics [2; 4; 6; 9; 10], the study analyzes the distributional and functional dissimilarities between the *with-SubjPart₁-cxn* and *with-less-SubjPart₁–cxn* applying the quantitative corpus methods of simple collexeme and distinctive collexeme analyses [7; 8; 11; 12] to the sample collected from the BNC-BYU [1]. The output of the conducted quantitative analyses serves as the basis for considering the underlying semantic factors that motivate the distribution of nouns in the subject slot of the investigated syntactic structures and thereby define semantic and functional contrast between them. The obtained results suggest that the analyzed constructions differ in terms of

1) *productivity*, i.e. the *with-SubjPart₁–cxn* proves to be more productive in modern English usage than the *with-less-SubjPart₁–cxn*;

2) *semantics of nouns in the subject slot* (the common nouns occurred in the subject slot of the *with-less-SubjPart₁–cxn* evoke BODY_PARTS, WEATHER, KINSHIP, PURPOSE, REASON, EMPHASIZING, INCLUSION, and SIMILARITY semantic frames (as in [3; 5]). The distinctive collexemes refer to semantic frames BODY_PARTS, WEATHER, KINSHIP, and PURPOSE. The nouns in the subject slot of the *with-SubjPart₁–cxn* instantiate PEOPLE, PEOPLE_BY_VOCATION, COMMERCE_SCENARIO, LEADERSHIP, PEOPLE ALONG POLITICAL SPECTRUM, LAW_ENFORCEMENT_AGENCY, BODY_PARTS, WEATHER, PRECIPITATION, CALENDRIC UNIT, NATURAL FEATURES, and ATTENTION semantic frames. The distinctive collexemes evoke the semantic frame PEOPLE;

3) *agentivity of the subject's referent:* the subject of the *with-less-SubjPart₁–cxn* is typically inanimate, acts as PATIENT of state/process expressed by [V$_{PARTICIPLE\ I}$], coreferent (PARTITIVE) with the subject of the matrix; the subject of the *with-SubjPart₁–cxn* is typically animate, acts as AGENT of a process/state expressed by [V$_{PARTICIPLE\ I}$], not coreferent with the matrix subject;

4) *discourse function*: the *with-less-SubjPart₁–cxn* performs one prominent (depictive function) and one peripheral function (support function). Used in fiction with body part nouns in the subject slot, this construction adds new details to the matrix event by describing the personages, their outward and inward characteristics. With general factual nouns in the subject slot, it provides supplementary information to the matrix event in written narrative non-fiction texts. The *with-SubjPart₁–cxn* typically

implements support function. Prevailing in newspapers and magazines, the construction attracts general people nouns in the subject slot. It elaborates the event in the matrix, stressing the significance of a human being in general.

From the quantitative-corpus research that has been carried out, we conclude that the *with-SubjPartI–cxn* and *with-less-SubjPart₁–cxn* are semantically distinct patterns that perform different discourse functions.

References

BNC-BYU. https://www.english-corpora.org/bnc/

Croft, W.: Construction Grammar. In: Geeraerts D., Cuyckens H. (eds.) The Oxford Handbook of Cognitive Linguistics, Oxford University Press, 463-508 (2008).

Fillmore, C.J., Lee-Goldman, R.R., Rhodes, R.-S.: The FrameNet constructicon. In: Boas H.C., Sag I.A. (eds.) Sign-Based Construction Grammar. CSLI Publications, Stanford, 283-299 (2012).

Fillmore, Ch.: The Mechanisms of "Construction Grammar". In: Proceedings of the Fourteenth Annual Meeting of the Berkley Linguistic Society 14, 35-55 (1988).

FrameNet. https://framenet.icsi.berkeley.edu/fndrupal/WhatIsFrameNet.

Goldberg, A. E.: Constructions at Work: The Nature of Generalization in Language. Oxford University Press (2006).

Gries, S. Th., Stefanowitsch, A.: Extending collostructional analysis: A corpus-based perspective on 'alternations'. In: International Journal of Corpus Linguistics 9(1), 97-129 (2004).

Gries, S. Th.: More (old and new) misunderstandings of collostructional analysis: On Schmid and Küchenhoff (2013). In: Cognitive Linguistics 26(3), 505-536 (2015).

Hilpert, M.: Constructional Grammar and its application to English. Edinburgh University Press (2019).

Hoffmann, Th., Trousdale, G.: Construction grammar: Introduction. In: Hoffmann T., Trousdale G. (eds.) The Oxford Handbook of Construction Grammar. Oxford University Press, 15-31 (2013).

Stefanowitsch, A., Gries, St. Th.: Collostructions: Investigating the interaction between words and constructions. In: International Journal of Corpus Linguistics 8(2), 209-243 (2003).

Stefanowitsch, A.: Collostructional analysis. In: Hoffman Th., Trousdale G. (eds.) The Oxford Handbook of Construction Grammar. Oxford University Press, 290-307 (2013).