Introduction

Current trends in analyzing syntactic variation

Ludovic De Cuypere and Gert De Sutter
Ghent University – Vrije Universiteit Brussel | Ghent University

Variation is a key feature of any living language. The present issue brings together a collection of papers dealing with variation in syntax. All studies were originally presented at *New Ways of Analyzing Syntactic Variation 2* (Ghent University, 19–20 May 2016), a conference which specifically invited original research on variation in the syntax of any (present or historical) language or language variety.

Syntactic variation encompasses a broad spectrum of syntactic phenomena and may refer to alternating morphosyntactic constructions, or constructions that exhibit minimal formal differences, such as a different word/constituent patterning and/or the optional usage of a grammatical marker (e.g., morphological case, prepositions, complementizers, etc.). Examples of such alternating constructions that have received much attention in the literature include the dative alternation (e.g. *John gave the book to Mary* vs. *John gave Mary the book*) (e.g., Bresnan and Ford 2010, as well as many many others), particle placement in English (e.g., Gries 2003, Cappelle 2006), the English genitive alternation (e.g. *John's book* vs. *the book of John*) (e.g., Heller et al. 2017), English zero vs. *that*-complementizer (e.g. *I think (that) he is smart*) (e.g., Shank et al. 2016, Wulff 2016), various differential marking phenomena (e.g., Spanish Differential Object Marking, Portuguese Differential Infinitival Marking, Hindi Differential Subject Marking, et cetera (e.g., Vanderschueren & Diependaele 2013). Several other examples of alternating constructions are studied in the papers collected in this issue.

Syntactic variation has driven the research agenda of different linguistic paradigms, including, but not limited to, construction grammar and related usage-based approaches, formalist approaches to grammar, variationist sociolinguistics (cf. the advent of socio-syntax), psycholinguistics, language acquisition research, and computational linguistics/NLP. Syntactic variation has accordingly been approached from quite divergent theoretical and methodological perspectives.

The linguistic research on morphosyntactic alternations has been driven by a common interest as to what motivates the speakers to choose between the different variants. One approach that was often entertained in the past was to

look for subtle semantic differences between the alternants (either on a lexical or on a more general constructional level), that could then explain speakers' choices of options deterministically. For instance, several interpretations of a semantic difference have been suggested for the dative alternation. One interpretation (e.g., Wierzbicka 1986, Newman 1996) holds that the Double Object Construction (DOC) (*John gives Mary a book*) emphasizes the Recipient, while the to-dative construction the Theme. Another interpretation relates the formal closeness between the Recipient and the verb in the DOC to their semantic closeness; the teaching would thus have been more effective in *I taught John Greek* than in *I taught Greek to John* (e.g., Lakoff & Johnson 1980, Rohdenburg 2003). Another approach was taken by generative scholars who tried to find derivational rules that would relate the formally different but semantically similar surface patterns to one underlying deep structure (e.g., Katz-Postal 1964, Jackendoff and Culicover 1971, Larson 1988 and, more recently, Bruening 2010). What both approaches have in common is a shared assumption that morphosyntactic choices depend on a single determining factor.

However, as Stefan Gries observes in his contribution to this volume, there appears to be a growing consensus in the research field that morphosyntactic alternations are motivated by multiple factors including intra-linguistic, i.e., a combination of syntactic, semantic or pragmatic factors, and extra-linguistic ones (e.g., age, gender, medium, genre, time, etc.). This multifactorial approach is commonly grounded in a general probabilistic view on syntax. Different factors are taken to operate simultaneously, but yielding a different "effect", meaning that they have a different but shared impact on the probability of one of both alternants. Taken together, multiple factors can have small incremental effects that result in preferring one construction over the other, given specific factor loadings.

It has become a common research objective in the study of morphosyntactic alternations to examine and estimate the effects of multiple factors by means of corpus data. This corpus-based line of research typically deploys sophisticated statistical modelling tools – mixed-effects binary logistic regression has become the bread and butter of the corpus linguist – and may additionally be triangulated with other methods, including grammaticality judgment experiments (e.g., the 100-split task pioneered by Bresnan et. al in the context of the English dative alternation) or psycholinguistic experiments (e.g., continuous lexical decision and sentence completion tasks in the context of the English dative alternation, Bresnan & Ford 2010). Taken together, these methodological advances offer strong empirical and converging evidence in favour of a multifactorial, probabilistic approach towards syntactic alternations.

Stefan Gries' multifactorial work has inspired and shown the way to many a corpus linguist. In the opening paper "Syntactic alternation research: taking stock

and some suggestions for the future", Gries discusses two possible pathways for further improvement. One set of improvements includes making more accurate model building choices, such as testing specific contrasts or allowing for non-linear covariates (e.g., splines), taking into account autocorrelation effects, showing more caution when performing model selection or using alternatives to regression modelling (e.g. clasification trees or ensemble methods). A second pathway for improvement suggested by Gries consists of testing predictors that remain understudied in corpus-based research but that have proven worthwhile in psycholinguistics (e.g., potential predictor variables relating to information status, phonetics, rhythm, etc.).

Jeroen Claes combines existing spoken corpus data with data gathered from *Twitter* in his paper "Cognitive and geographic constraints on morphosyntactic variation. The variable agreement of presentational *haber* in Peninsular Spanish". Additional data processing included geographical and part-of-speech tagging. Claes also used multiple data analyses: collocational analysis mixed-effects logistic regression and generalized additive mixed modeling. The analysis shows that *haber* pluralization (e.g., *hubo/hubieron fiestas* 'there was/were parties') is spreading geographically and that its use is preferred with nouns that refer to typical action-chain heads, when negation is absent, and with tenses other than the preterite.

Artemis Alexiadou's paper "Language variation and change: a case study of the loss of genitive Case in (Heritage) Greek", offers yet another way to advance the study of syntactic alternations. First of all, the object of study is a relatively new field of investigation. Following Rothman (2009, 156), Alexiadou defines a heritage language as "a language that is spoken at home or otherwise readily available to young children, but, crucially, this language is not a dominant language of the larger (national) society". Alexiadou makes the case that, in contrast to what is commonly assumed, heritage grammars are not suboptimal systems – either because of attrition or a lack of full grammatical development –, but the result of systematic language changes that are present in the homeland variety as well. What is interesting, from a methodological perspective, is that the argumentation is built on a triangulation of data sources. The study draws on Greek diachronic, dialectal and language acquisition data, as well findings on Old French.

Another innovative way of data collection is offered in "A crowdsourcing approach to the description of regional variation in French object clitic clusters", by Mathieu Avanzi and Elisabeth Stark. They studied regional distribution of the ordering patterns of object clitic clusters (e.g., *il le lui montre*, 'he shows it to him') in European French (France, Belgium and Switzerland). Data was collected based on questionnaire-based surveys. Using mailing lists and social media, they gathered data from over 20,000 participants, which allows to draw very fine-grained

geographical maps of the regional distribution of the clitic ordering preferences. Their results suggest that the different ordering variants are generally not associated with specific dialect regions. Language contact and analogical levelling are additionally discussed as possible explanations for the findings.

From a formal-linguistic perspective, Eugenia Mangialavori focuses on causative-inchoative alternations in Spanish, and proposes an innovative theoretical way to understand the so-called Stative Causative Construction. Under the assumption that the syntactic projection of arguments is related to event structure in a non-trivial manner, the author gathers empirical (corpus-illustrated) support for the central theory proposed in this paper.

Gosse Bouma's paper "Agreement Mismatches in Dutch Relatives" is also based on *Twitter* data; no less than 723,470 relevant tweets were gathered. In Dutch, two relative pronouns *die* vs. *dat* are selected based on the grammatical gender of the head noun. The pronoun *dat* is used with neuter nouns (e.g., *het huis dat …* 'the house that…'), the pronoun *die* with common nouns (*de woning die* 'the house that…'). However, in informal Dutch, disagreement is observed, with *die* being used where *dat* is expected. Bouma found that linguistic factors are involved in this mismatch (e.g., animacy and lexical class), but also extra-linguistic variables. More specifically, more mismatches occurred in the speech of young, female users.

The following four papers study alternating constructions from multiple methodological perspectives. Elisabeth Verhoeven's study "Scales or features in verb meaning? Verb classes as predictors of syntactic behavior" zooms in on the role of agentivity as a verbal feature in distinguishing verb classes and the effect of agentivity on the syntactic behaviour of verbs. Her basic research question – is the factor "agentivity" a discrete or gradient notion – not only dovetails well with Stefan Gries's appeal for more thoughtful model building choices, it also sheds light on the alleged probabilistic nature of grammatical knowledge. Verhoeven's experimental and corpus data offer mixed results. While the grammaticality judgement task seems to justify the assumption of a gradual agentivity scale – speakers have fine-grained knowledge about the likelihood of verbs to occur in agentivity context ', a scalar notion of agentivity did not increase the predictive power of the model for the syntactic variation in word order. A binary notion of agentivity (±agentive and non-agentive) was associated with the best model fit.

Brandt and Fuß's study "A corpus-based analysis of pronoun choice in German relative clauses" combines a qualitative analysis based on introspection with a cluster analysis of corpus data. They make the case that the alternation between the German relativizers *das* and *was* is determined by categorial properties of the antecedent of the relative clause. In general, it appears that *das* is used when a lexicalized head noun is present, while *was* is used in all other instances. Their corpus

evidence further suggests that the alternation is motivated by semantic and pragmatic properties related to the head noun word.

Laurence Romain's paper "Measuring the alternation strength of causative verbs: a quantitative and qualitative analysis of the interaction between verb, theme and construction" combines distinctive collostructional analysis, Shared Index Type, and distributional analysis to examine the alternation associate with labile verbs, i.e., verbs that take both a transitive and intransitive construction. Shared Index Type is a new method proposed by Maarten Lemmens and is used by Romain to estimate the semantic overlap between the themes shared by the two constructions. Romain's findings indicate that there is little overlap in the themes that occur in both constructions.

Anton Granvik combines collostructional analysis, distinctive colleme analysis, cluster analysis and mixed-effects logistic regression in his study "Accounting for syntactic variation in diachrony: the presence vs. absence of *de* in finite nominal complement clauses in 16th and 17th century Spanish". The alternation under analysis involves the optional use of the preposition *de* in combination with the complementizer *que*, as in *la posibilidad (de) que…* 'the possibility that…'. Granvik's corpus findings indicate that the use of *de que* is preferred with independent nouns, whereas *que* is preferred with nouns following a verb with which it forms a complex predicate construction.

The paper by Borja Herce also zooms in on a Spanish cluster of constructions, viz. time constructions involving *hacer* 'to make'. The results from a quantitative investigation reveal an important hitherto unnoticed contrast between the properties of the so-called clausal and adverbial constructions. Whereas *hacer* has properties not significantly different from other verbs in the former construction, it shows a strong erosion of its verbal and clausal properties like TAM morphology, negation, time adjunction or word order freedom in the latter. This contradicts earlier research on the topic and argues against those formal proposals positing a synchronic derivational relation between the two constructions.

David Tizon-Couto, finally, examines syntactic alternation from a diachronic perspective in his paper "Exploring the Left Dislocation construction by means of multiple linear regression: complexity and orality of Modern English left-dislocated NPs". Once a rather frequent phenomenon in Old and Middle English (with an estimated frequency of 1/100 sentences), Left-Dislocated noun-phrases (e.g., *John* this morning, *he* …) have steadily decreased over the past centuries (Modern English: 1500–1914), particularly in written language. Using corpus data, Tizon-Couto estimates the effects of different factors (e.g., genre, style and complexity constraints) on the length of dislocated items. The association between short left-dislocated items and speech-related genres is additionally explored.

References

Bruening, Benjamin. 2010. "Double Object Constructions Disguised as Prepositional Datives". *Linguistic Inquiry* 41: 287–305. https://doi.org/10.1162/ling.2010.41.2.287

Bresnan, Joan W., and Marilyn Ford. 2010. "Predicting syntax: Processing dative constructions in American and Australian varieties of English." *Language* 86 (1): 168–213. https://doi.org/10.1353/lan.0.0189

Cappelle, Bert. 2006. "Particle placement and the case for 'allostructions'". *Constructions*

Gries, Stefan Th.. 2003. *Multifactorial analysis in corpus linguistics: a study of Particle Placement*. London & New York: Continuum Press.

Heller, Benedikt, Benedikt Szmrecsanyi & Jason Grafmiller. 2017. "Stability and fluidity in syntactic variation world-wide: the genitive alternation across varieties of English". *Journal of English Linguistics* 45(1): 3–27.

Jackendoff, Ray, and Perter William Culicover. 1971. "A Reconsideration of Dative Movement." *Foundations of Language* 7: 397–412.

Katz, Jerrold, and Paul Postal. 1964. *An integrated theory of linguistic descriptions*. Cambridge, MA: MIT Press.

Lakoff, George, and Mark Johnson. 1980. *Metaphors we live by*. Chicago (Ill.): University of Chicago press.

Larson, R. K. 1988. "On the Double Object Construction". *Linguistic Inquiry* 19: 335–393.

Newman, John. 1996. *Give: a cognitive linguistic study*. Cognitive linguistics research 7. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110823714

Rothman, Jason. 2009. "Understanding the Nature and Outcomes of Early Bilingualism: Romance Languages as Heritage Languages." *International Journal of Bilingualism* 13: 155–163. https://doi.org/10.1177/1367006909339814

Rohdenburg, Günther. 2003. "Aspects of grammatical iconicity in English." In *From Sign to Signing*, edited by Wolfgang G. Müller and Olga Fischer, 263–285. Amsterdam: Benjamins. https://doi.org/10.1075/ill.3.19roh

Shank, Christopher, Van Bogaert Julie, and Koen Plevoets. 2016. The diachronic development of zero complementation: A multifactorial analysis of the that/zero alternation with think, suppose, and believe. *Corpus Linguistics and Linguistic Theory*. 12(1): 31–72.

Vanderschueren, Clara, and Kevin Diependaele. 2013. "The Portuguese Inflected Infinitive: An Empirical Approach." *Corpus Linguistics and Linguistic Theory* 9 (1): 161–186. https://doi.org/10.1515/cllt-2013-0013

Wulff, Stefanie. 2016. "A friendly conspiracy of input, L1, and processing demands: *that* – variation in German and Spanish learner language". In *The usage-based study of language learning and multilingualism* (Proceedings of GURT 2014), edited by Lourdes Ortega, Andrea E. Tyler, Hae In Park and Mariko Uno, 115–136. Georgetown: Georgetown University Press.

Authors' addresses

Ludovic De Cuypere
Department of Linguistics
Ghent University – Vrije Universiteit Brussel
Blandijnberg 2
9000 Gent
Belgium

Ludovic.DeCuypere@UGent.be

Gert De Sutter
Department of Translation, Interpreting and
Communication
Ghent University
Groot-Brittanniëlaan 45
BE-9000 Gent
Belgium

gert.desutter@ugent.be