

GO (*a*)round and V vs. GO (*a*)round Ving

A multivariate distinctive collexeme analysis based on association rules

Kim Ebensgaard Jensen¹ and Stefan Th. Gries^{2,3}

¹University of Copenhagen | ²UC Santa Barbara | ³JLU Giessen

Distinctive collexeme analysis has proven very useful in identifying distinctive patterns of lexemic attraction among multiple constructions. However, if construction grammar is to take seriously the usage-based tenet that context is a crucial component of the language system, multivariate methods are required. In this article, one such analytical approach is proposed as we apply an extension of distinctive collexeme analysis – named multivariate collexeme analysis – in an analysis of the GO (*a*)round and V and GO (*a*)round Ving constructions. Based on the data-mining technique known as association rules, multivariate collexeme analysis can identify not just singular distinctive features but also establish entire ‘collo-profiles’ of multiple features going far beyond individual collexemes. Our analysis takes into account no less than eight features (including collexemic, colligational, pragmatic, and discursive ones), and it is found that this approach offers much more informative accounts of the two constructions than a traditional distinctive collexeme analysis would.

Keywords: association rules, collo-profiling, internal/external constructional properties, constructional multidimensionality, usage-based construction grammar

1. Introduction

One of the most widely used corpus methods in construction grammar, collostructional analysis (Stefanowitsch & Gries, 2003, 2005; Gries & Stefanowitsch, 2004a, 2004b, Gilquin & Gries, under review) has proven to be extremely valuable to usage-based construction grammarians in the study of lexeme-construction interaction. It has contributed to the identification of constructions and subconstructions and the mapping of underlying cognitive-semantic patterns.

However, given that a key tenet of usage-based linguistic theory is the inseparability of the language system and recurring contexts of use (e.g., Kemmer & Barlow, 2000), there is a need to take collostructional analysis beyond the lexeme and the syntactic configuration and systematically include contextual factors.

Schönefeld's (2013) register-based distinctive collexeme analysis and Hilpert's (2012) diachronic distinctive collexeme analysis, both of which are limited to two elements (lexeme + register and lexeme + year within one construction respectively) are testimony to this need. Stefanowitsch and Flach (2020) propose a variant dubbed distinctive covarying collexeme analysis which measures the interaction of three elements (construction + adjective + adjective in the *too* ADJ *to* V and ADJ *enough* *to* V constructions).

Going much beyond previous work, this paper explores the possibility of expanding collostructional analysis into including even more features. More specifically, we apply a method that we call multivariate collexeme analysis in a study of the GO (*a*)*round and* V and GO (*a*)*round Ving* constructions as observed in the 2010–2019 portion of Davies' (2008-) *Corpus of Contemporary American English* (henceforth, COCA). In a previous study of these two constructions and in an attempt to account for semantic- and discourse-prosodic differences, Jensen (2024) applies a series of monofactorial distinctive collexeme analyses to the following dimensions: collexemes in the V-positions, discourse and semantic prosody, colligation, and speech acts. While this approach provides some useful insights into the semantic- and discourse-prosodic differences between the two constructions, Jensen (2024) already argues for the preferability of a multivariate analysis seeing that taking multiple factors into account would offer a richer account. That is exactly what multivariate collexeme analysis is. Based on a data-mining technique known as association rules (Hahsler et al., 2005), this method allows the analyst to address multiple factors simultaneously such that collexemic patterns can be addressed alongside contextual ones. In this paper, we apply this multivariate kind of analysis to the two above-mentioned constructions, incorporating eight features (collexemes in the V-positions, colligation, discourse prosody, semantic prosody, modality, polarity, register, and speech acts) and addressing how the intricate patterns of feature interactions differentiate between the two constructions.¹

1. Readers might wonder, as did one reviewer, whether the case for multivariate methods still needs to be made. Indeed, the higher degree of utility of multivariate approaches should be self-evident by now *in theory*, but it is just as clear that, especially in the domain of collostructional applications, multivariate approaches remain extremely rare. Therefore, we consider our case study to still provide a valuable contribution and exemplification.

Since we build on Jensen (2024), it is useful to briefly discuss how this study relates to that prior work. The present study has a completely different *raison d'être* from Jensen (2024) and is distinct in a number of ways. First, Jensen (2024) focuses on semantic prosody and discourse prosody. In contrast, the present study seeks to offer a richer account of the overall distinctive behaviors of the two constructions without focusing on any particular dimension of usage. Second, Jensen (2024) applies a range of monofactorial analyses and only accounts for distinctiveness one dimension at the time while this study applies a single multivariate method, thus accounting for distinctiveness in interactions across, and competition between, dimensions. Third, this study includes more feature dimensions than Jensen (2024), as we include modality, polarity, and register, all of which feed into a richer account of the two constructions. Finally, this study has a methodological purpose completely absent from Jensen (2024): the application of multivariate analysis based on association rules as a way to take construction grammar towards a multidimensional and hence richer understanding of constructions.

2. Constructions and their external properties

The basic tenets of construction grammar (Fillmore et al., 1988; Goldberg, 1995; Croft, 2001; Hilpert, 2019) are as follows (see also Traugott & Trousdale, 2014, p.258):

- the basic unit of grammar is the construction, a conventionalized pairing of form and meaning;
- grammar is non-modular so no linguistic domain is primary;
- constructional form includes syntax, morphology, and phonology;
- constructional meaning includes semantics and discourse-pragmatics;
- constructions may be formally specific or formally schematic to varying degrees.

In usage-based construction grammar, constructions are further viewed as entrenched routines in the speech community (Croft, 2005, p.274) emerging as regularities from recurring interactions. Patten (2014, p.91) neatly summarizes this:

humans are not innately programmed with grammatical knowledge; instead, all aspects of language are learned from the input (or rather from the speaker's linguistic experiences). Both language learning and language change involve the speaker inductively generalizing over instances to form mental schemas (or constructions) which are represented in the language system. On a usage-based model then, constructions are simply conventionalized chunks of linguistic

knowledge... From this, it follows that the storage and organization of grammatical knowledge is dependent on, and can change according to, patterns of use

Contextual information is part of a construction: features of contexts in which the construction recurs may be entrenched as constructional knowledge on par with formal-structural information.

While early iterations of construction grammar were not rigorously usage-based in the above sense, there was an awareness of the importance of context, discourse-pragmatics always having been part and parcel of constructional meaning. A distinction was made between internal and external properties of constructions (Fillmore, 1988, pp.36–37). The former covers formal, semantic, and symbolic structures, and the latter subsumes recurring contextual patterns. This distinction is not a matter of one set of properties being more inherent to a construction than another set. Rather, it highlights that structural knowledge of constructions is not the only kind of knowledge in speakers. Knowledge of contexts of use is also *bona fide* constructional knowledge. In a usage-based perspective, external properties range from co-textual ones (including traditional ‘collostructions’) to situational, social, and intertextual ones. We consider colligational and collostructional features as external properties because, while they appear in positions within a construction, these positions are schematic and characterized by variability just like collocates and other co-textual elements. Only in cases of lexical fixedness (as with GO in both constructions) or morpho-syntactic fixedness (as with *Ving* in GO (*a*)*round Ving*) does it make sense to consider such elements internal properties. This inclusion of contextual features is in line with Kemmer and Barlow’s (2000) usage-based tenet of the crucial role of context in the operation of the language system; consequently, what determines subconstruction- and constructionhood may be a complex intersection of internal and external properties. In other words, it makes sense to assume that many constructions are essentially multidimensional entities, and application of multivariate methods, such as association rules, is thus a useful, if not even necessary, way of identifying how external properties interact and, in the case of comparing two or more constructions, the extent to which such interactions are distinctive.

3. A preliminary look at the two constructions

Consider the following instances of GO (*a*)*round Ving* (1–3) and GO (*a*)*round and V* (4–6):

- (1) Well I don’t go around assaulting people (COCA 2019 TV Warrior)

- (2) Unlike a lot of older women with money to burn, she doesn't go around buying fabulous designer fashions that were created for twenty-five-year-olds but look ridiculous on a sixty-five-year-old. (COCA 2010 FIC Bk:ThreadSoThin)
- (3) You shouldn't go around fighting strangers in the forest.
(COCA 2010 TV Adventure Time)
- (4) So what they're doing is, they're going around and paying off all of these countries to sign onto the deal. (COCA 2015 SPOK Fox: Fox Hannity)
- (5) In my mind, as I saw it happen, she was just going to go around and shoot everyone. (COCA 2010 NEWS NYTimes)
- (6) You know he invented a little camera. He's going around and spying on us.
(COCA 2012 MOV 3 Day Test)

In both constructions, an inflectionally schematic GO is followed by (*a*)round. However, in (1–3) (*a*)round is followed by an inflectionally fixed present participle in an adverbial *ing*-clause while, in (4–6), it is coordinated with a clause in which the verbal is inflectionally concordant with GO.

Both constructions carry compositional and non-compositional meaning and serve pragmatically as stance-markers in which the activities expressed by the V and Ving slots are assessed negatively in what Stefanowitsch (2000, p.262) calls a DISAPPROVAL construal. In (1–6), GO *around* arguably expresses ITERATIVITY as a metaphorical extension of the literal meaning of GO (*a*)round. Stefanowitsch (2000, p.262) suggests that a basic image schema of MOTION ALONG A NON-DIRECTIONAL PATH, conveying AIMLESSNESS, INTENTIONALITY, and HABITUALITY, is involved. Note that the hypothetical mass shooting in (5), while not habitual, is iterative in how the event of the mass shooting as a whole consists of the event of shooting at one or more victims, i.e. shooting multiple times.

English has more constructions expressing DISAPPROVAL of aimless and iterative activities, the most famous one in construction grammar circles probably being the V *the TIME away* construction as described by Jackendoff (1997), which also expresses an excessive atelic situation that may be iterative if the verb expresses a bounded but repeatable event. In (7–8), the V and Ving elements express behaviors evaluated as bad or unpleasant through the DISAPPROVAL construal although the situations involved would otherwise not be seen as bad or unpleasant:

- (7) We don't exchange heart-shaped boxes of chocolates or glossy cards with manufactured endearments inside, and we don't go around kissing in public or saying "I love you" twenty times a day. (To my mind, couples like that are always suspect – really, who are they trying to fool?) (COCA 2018 FIC New Yorker)

- (8) Going around and promising a whole bunch of new ideas and new projects and big ideas – that was fine. And maybe we need it; we can do that. I supported the doubling of the National Institutes of Health. But we didn’t have a \$ 1.2 trillion deficit. We didn’t – we weren’t at over – we are now going to reach \$ 16 trillion, which is more than our whole GDP. We were not in that situation 20, 15 years ago. (COCA 2012 SPOK CNNLIVEEVENT/S)

However, compared to our target constructions, the V *the TIME away* construction does not seem to add the same sense of unpleasantness. Jackendoff (1997, p.537) himself suggests an implication in the construction “that the subject is in some sense ‘using’ the time, or even better ‘using the time up’”. Thus, the sense of DISAPPROVAL in this construction might be the excessiveness of the situation rather than the unpleasantness associated with GO (*a*)round Ving and GO (*a*)round and V.

Let us return to the two constructions studied in this paper. Initially, they appear functionally very similar. However, given the principle of no synonymy (Goldberg, 1995, p.67), one would expect that, for both constructions to ‘stick around’ in contemporary usage, there must be some functional difference, which would motivate the continued existence of two structurally very similar yet not identical constructions. In our paper, we are deploying the new multivariate extension of (distinctive) collexeme analysis with various different kinds of features to identify what featural distributional differences might suggest about what that functional difference might be. A note on terminology: we are using *distinctive* in the previous sentence because we are here comparing two constructions with each other, as in the typical case of distinctive collexeme analysis (see also the beginning of Section 4.2) – however, the multivariate extension we are proposing here can also be applied in contexts that are not distinctive in that sense, as when a *single* construction is explored with many different features.

Why address these constructions in a multivariate perspective? There are at least three reasons. First, as mentioned above, if we are to take seriously the usage-based tenet mentioned above that context plays a central role in the operation of the language system (Kemmer & Barlow, 2000), then we can expect recurring contextual features to become entrenched as constructional knowledge in speakers in addition to purely linguistic ones. This in itself calls for analyses that draw on more dimensions than just one.

Second, if we want to gain a sufficient understanding of the two constructions as outlined above, multiple dimensions are required. As mentioned above, the two constructions are essentially stance markers. Extant research in semantic prosody and related phenomena (e.g., Hunston & Francis, 1999; Stubbs, 2001; Jensen, 2017) has shown that the stance function is often reflected in patterns among the lexemes that the stance marker co-occurs with in actual usage-events

wherefore semantic prosody should be taken into account. Moreover, the two constructions may be used not only to express disapproval of behavior deemed unpleasant or otherwise socially unacceptable but also to discourage people from engaging in such behavior as well as to commit oneself not to engage in it. Consequently, it is relevant to take into account the speech acts that the constructions are used in as well as colligational VP patterns, as these (i) are often indicative of sentence types, which in turn are also linked to speech acts, (ii) often contribute semantic adjustments to the content expressed by the main verb, and (iii) may be attracted not just to the constructions but also to particular verbs occurring in the construction.

However, there is a need to go beyond these features and also include modality as a factor, because some of the main functions of deontic modality are the expression of obligation, permission, volition, and commitment. Clearly, then, modality and speech acts together are vital to our understanding of the two constructions. By the same token, polarity is relevant, as polarity markers signal whether the action or behavior in question is encouraged or discouraged.

We see many of these dimensions at play in the examples presented so far. Consider, for instance, (3) where the lexeme in the *Ving* position expresses an arguably undesirable activity, and the modal verb *should* and the negator *n't* signal a negative obligation such that the speaker is discouraging the addressee from fighting strangers in the forest. Thus, (3) is arguably also a directive speech act. Compare this to (1) which has more of a commissive function.

Third, decades of variationist research in sociolinguistics and neighboring disciplines – including construction grammar (e.g., Hilpert, 2012; Schönefeld, 2013) – have taught us that linguistic phenomena may occur more frequently in some contexts or varieties than others and that the same linguistic phenomenon may have quite different functions and exhibit different types of behaviors in different contexts or varieties. Circling back to usage-based language theory, this type of information is likely to be part of speakers' linguistic competence. Consequently, where possible (as is the case of COCA which includes information on register), it makes sense to take this into account as well.

4. Method

The usage-based perspective adopted in this study naturally has methodological consequences. As mentioned above, if recurring contextual patterns are part of speakers' operational knowledge of constructions, then contextual features are part of constructions too and entrenched in speakers as external constructional properties. Therefore, studying recurring contextual features equals studying the

construction itself. Thus, our application of a multivariate method that addresses contextual features on a par with collexemes is not just valid but arguably seems necessary if one's goal is to offer a richer account of the usage of the construction.

4.1 Data and annotation

This study draws on data from COCA (Davies, 2008-). More specifically, we used a 248,145,425-word portion of the corpus covering the registers ACADEMIC, FICTION, MAGAZINES, MOVIES, NEWS, SPOKEN, and TV from the time period 2010–2019. Thus, all data are recent, all COCA registers are represented, and the corpus portion is sufficiently sized as it contains almost 250 million tokens.

Data were retrieved using the following search strings:

- GO around _V?G: GO in any form followed by *around* followed by a verb in the *ing*-form;
- GO round _V?G: GO in any form followed by *round* followed by a verb in the *ing*-form;
- GO around and VERB: GO in any form followed by *around and* followed by a verb in any form;
- GO round and VERB: GO in any form followed by *round and* followed by a verb in any form.

The search yielded altogether 544 instances: 443 instances of GO (*a*)*round Ving* and 101 instances of GO (*a*)*round and V*. Note that these searches ignore discontinuous instances of the two constructions like *So if you go around publicly pirating wedding china...* (COCA 2007 FIC Fantasy/SciFi) or *She can't go around and personally remedy it for everyone* (COCA 2017 MOV 13 Reasons Why: Beyond the Reasons). Searches for discontinuous forms using wildcard functions before *Ving* and *V* resulted in large numbers of false positives and numbers of genuine instances so small that their inclusion would make no difference for the purpose of the current study.² For future research, of course, adverbials (often causing dis-

2. Weeding out false positives – i.e., patterns that meet a structural description but are nonetheless not instances of the target construction – is extremely widespread in corpus-linguistic studies (and often an important difference to fully computational approaches that maximize recall at the expense of precision). Here are some examples of false positives: *And then there's another date that keeps going around and that is in December* (COCA 2011 SPOK Fox_Beck), *And go around and then touching my breast* (COCA 2013 MOV Ghost Team One), *A bad batch going around was catching a lot of people* (COCA 2018 FIC Confrontation), and *Certainly, there is enough blame to go around in explaining out addiction to oil* (2011 NEWS CSMonitor).

continuity) could be interesting and potentially valuable to include in multivariate analysis of the two constructions. Consequently, this study cannot be said to – nor do we claim that it does – provide the full picture regarding the two constructions.

The instances were subsequently manually annotated for features within the following feature categories:

- COLLEXEME: Lexemes in the *Ving*- and V-positions (214 lexemes altogether), i.e. the one and only feature that just about all previous distinctive collexeme analyses have relied on, and many such studies have shown that collexemes are often distinctive features seeing apart otherwise seemingly synonymous constructions;
- COLLIGATION: Grammatical categories that the constructions co-occur with – more specifically, all the colligates of GO observed in the data: *simple present*, *simple past*, *present participle* (i.e. verbal in *ing*-clause), *progressive aspect*, *excessive aspect*, *inceptive aspect*, *modalized VP*, *mandative subjunctive*, *imperative*, *infinitive* (i.e. verbal in infinitive clause), *controlled future tense*, *perfective aspect*;
- MODALITY: Types of modality where applicable: *epistemic* (possibility, probability, likelihood, logical necessity, and ability), *deontic* (obligation, compulsion, volition, permission and prohibition), and *none* (used when instances of the construction were not used in a modal context);³
- POLARITY: Clause polarity categories: *negative* polarity (instances where a negator is present) and *affirmative* polarity (instances where a negator is absent); while technically also a matter of colligation, this is given its own category, as it is separate from verbal colligative categories;
- SEMANTIC PROSODY: Emergent semantic categories among the observed collexemes. The semantic prosodies in this study were manually identified, and there are 120 prosodies altogether;
- DISCOURSE PROSODY: Emergent attitudinal patterns among the observed collexemes: *positive*, *negative*, and *neutral discourse prosody*. These were manually identified (a caveat is in its place here: judgments of positive and negative assessment may be subject to unwitting subjectivity on the part of the analyst);
- REGISTER: All registers of the portion of COCA used for this study: *academic*, *fiction*, *magazines*, *movies*, *news*, *spoken*, and *TV*;

3. We are aware that there are more complex models of modality in the extant research which make more fine-grained distinctions, but – for the sake of simplicity – we opted for the traditional bipartite model, noting that a more fine-grained model might result in a more detailed picture than what is provided here.

- **SPEECH ACT:** Speech act functions of utterances containing the constructions: *statements*, *questions*, *directives*, and *commissives*. This list is exhaustive as it subsumes all types of speech acts observed in the data.

SEMANTIC PROSODY and DISCOURSE PROSODY require some explication. According to Hunston and Francis (1999, p.137), “a word may be said to have semantic prosody if it can be shown to co-occur typically with other words that belong to a particular semantic set”. For instance, *undergo* prefers nouns from the semantic domains MEDICINE, EDUCATION AND ASSESSMENT, and CHANGE as direct objects. These semantic domains constitute its semantic prosody; moreover, many of the semantic preferences of *undergo* are associated with unpleasant experiences, and consequently the verb has negative discourse prosody (Stubbs, 2001, pp.89–95). SEMANTIC PROSODY and DISCOURSE PROSODY are typically associated with lexemes but have also been shown to be constructional features (Jensen, 2017). The traditional terminological distinction between semantic prosody and semantic preference is, in our opinion, misguided. Both phenomena are prosodic, as they extend beyond the individual unit (Stubbs, 2001, p.65), and only semantic preference has to do with semantics; semantic prosody is a matter of discourse-pragmatics. Therefore, we use ‘semantic prosody’ with reference to the former, and ‘discourse prosody’ – the term that Stubbs (2001) prefers – with reference to the latter. Discourse prosody is often derived from semantic prosody (Stubbs, 2001, pp.89–95). However, this connection is not invariably fixed. The former can override the latter such that, say, an otherwise positive situation is assessed negatively; we actually see this in (7–8). Therefore, while they are connected, it is necessary to include both as separate categories in the analysis.

To illustrate the application of the above-mentioned features, consider the following example:

- (9) I got news for you, you little psycho, you can’t go around slaughtering everyone who pisses you off. (COCA 2019 TV The Gifted)

Slaughter appears as the COLLEXEME in the *Ving*-position, and GO appears with the modal verb *can* and the contracted negator *n’t*. Furthermore, the utterance has *deontic* MODALITY and serves as a prohibitive *directive* SPEECH ACT. *Slaughter* here expresses an ACT OF VIOLENCE and an inherently negative scenario. Lastly, (9) appears in the TV REGISTER. See Table 1 for an overview of features assigned to this example.

All instances of the two constructions were subject to this procedure. The reader will undoubtedly have noticed that this study draws on a substantial number of feature categories (no less than eight) and a very extensive number of feature levels. For instance, COLLIGATION subsumes 12 feature levels, SEMANTIC

Table 1. Features assigned to (9)

Category	Feature
COLLEXEME	<i>slaughter</i>
COLLIGATION	<i>modalized VP</i>
DISCOURSE PROSODY	<i>negative</i>
MODALITY	<i>deontic</i>
POLARITY	<i>negation</i>
REGISTER	<i>TV</i>
SEMANTIC PROSODY	<i>VIOLENCE</i>
SPEECH ACT	<i>directive</i>

PROSODY 120 feature levels, and COLLEXEME 214 feature levels. Here, it is important to keep in mind that, while a multivariate analysis will undeniably be able to yield findings that capture distributional characteristics of the two constructions in much more detail, it – like any method – is still limited by the factors included. Thus, we are not claiming we have been successful to cover all variables. For instance, it is conceivable that grammatical person in subjects (where applicable) and degree of idiomaticity might play a role. Still, our analysis covers features from various contextual layers, ranging from ones occurring in schematic slots in the construction (i.e., COLLEXEME and COLLIGATION) over other co-textual elements like POLARITY to more abstract contextual ones like SPEECH ACT and REGISTER, and it covers both more features than any analysis of this construction has ever considered and than nearly all collostructional analyses usually consider. Therefore, we hold that our analysis covers enough aspects to (i) distinguish the functions of the two constructions and (ii) exemplify what we believe is a unique selling point of the analysis.

4.2 Methods

In this study, we present our version of multivariate collexeme analysis that is based on the logic of association rules and here applied in a context reminiscent of simpler distinctive collexeme analysis. In its most basic form, association rules are an algorithm that takes as input a data set of cases of (typically only) categorical variables and returns as output a list of notable co-occurrences of levels of different variables. At least the conceptual logic of association rules is known to anyone who has ever bought anything at, say, Amazon: while one is browsing or searching for a specific item, Amazon’s algorithm might make suggestions of the type ‘people who also looked at/for what you’re looking at/for often also looked

at this' or '... also often bought this.' Someone who might have bought seasons 1–3 of the TV show Friends might also be interested in buying season 4, at least more so than someone who has never even looked at websites selling products related to Friends. In other words, association rules are essentially long sentences with potentially many protases and one apodosis: If many people who bought seasons 1–3 of Friends also bought season 4, a corresponding association rule would be of the form 'If_{protasis₁} season 1 of Friends: *yes* and if_{protasis₂} season 2 of Friends: *yes* and if_{protasis₃} season 3 of Friends: *yes*, then_{apodosis} season 4 of Friends: *yes* (or likely enough to be suggested)' and, in the language of association rules, the three protases together make up the LHS (left-hand side) whereas the apodosis is referred to as the RHS (right-hand side).

That means association rules result from (i) multidimensional cross-tabulation – 4-dimensional in the above Friends example – and (ii) the computation of a variety of statistics that characterize each conditional sentence. Because the number of rules rises exponentially with the numbers of variables involved and with the numbers of levels they have, in analytical practice, one usually takes three main steps to keep the computation manageable and the output interpretable.

First, one delimits the set of rules the algorithm should retain by setting minimal parameter values the rule must achieve (to demonstrate its worth, so to speak); this step is comparable to setting hyperparameters in predictive modeling contexts. One such parameter is called *support*, which is the proportion of cases in the data set that exhibit all the variable levels mentioned in both the LHS and the RHS. In other words, support is the proportion of the data that a rule 'addresses' or 'speaks about'; in the above example, it's the proportion of customers that bought all seasons 1–4 of Friends. In our study we set a very permissive minimum support threshold of 0.01 to maximize recall of rules.

Another such parameter is called *confidence*, which is the proportion of times the LHS is associated with the RHS out of all times the LHS is observed. More technically, it is the support divided by the proportion of the LHS; in the above example, it's the proportion of customers that bought seasons 1–3 that then also bought season 4. In our study, we set a minimum confidence threshold of 0.5, i.e. the rule had to be 'right' at least half the time.

The second and third main steps to keep computations and output manageable and interpretable are to apply filtering and ordering to the set of rules that survives the above two screening parameters. One simple way in which filtering can be done is by deleting all rules that are what is technically called *redundant*. A rule is redundant "if a more general rules [sic] with the same or a higher confidence exists" (documentation for `arules::is.redundant`, see below for package information). In other words, a rule r_s is redundant if there is a more general rule

r_g , i.e. a rule whose LHS is a subset of r_s 's LHS, but which has at least the same confidence; filtering by redundancy would amount to discarding r_s because the more general – i.e. less complex – rule r_g does just as well or better; this filtering can be seen as a simple way of implementing Occam's razor. Another way of filtering is to retain only those rules whose LHS or RHS includes a certain variable or level. In our study, we did eliminate redundant rules from consideration and retained only rules whose RHS featured one of the two constructions to be contrasted; this step implements the “distinctive” component of our multivariate collexeme analysis. That means the distinctive collexeme analysis is built into the overall association rules analysis such that the distinctive collexeme status is determined alongside the distinctiveness of the other features. It is, in other words, not necessary to first carry out a distinctive collexeme analysis and then an association rules analysis.⁴

The third and final step involves ranking the rules by some relevant statistics to get an importance-based ranking. We chose the statistic of *lift* for this, which is the ratio of the observed support to the support expected if LHS and RHS were independent. Thus, a *lift*-value of 1 indicates the independence of the RHS from the LHS; *lift* is therefore like the corpus-based association measure *PMI* but without the logging.

4. At a first superficial glance, the proposed approach may seem less novel than we portray it here. One reviewer in particular stipulated a connection of our approach to predictive modeling approaches as commonly applied in alternation research. However, our approach is different in several regards, including, but not limited to, the fact that (i) association rules can address interactions that at least regression-based approaches would struggle with and (ii) predictive modeling approaches are *supervised* learning approaches that require a response variable (and can only be applied to distinctive collexeme methods), are evaluated using metrics like precision, recall, (chance-corrected) accuracy, and others, and are often approached from a hypothesis-testing perspective and, thus, come with *p*-values. Association rules are different in nearly all those regards: They are not necessarily supervised, can also be used for simple collexeme analyses, involve no evaluations against the gold standard speaker choices, and are exploratory and, thus, not ‘part of’ the null-hypothesis significance testing paradigm. In the same vein, our approach is also very different from Behavioral Profiles (BP, of the kind developed by Gries and Divjak): while both association rules and BP require annotated concordance lines, this is where the similarities end: unlike the present approach, BP's steps 3 and 4 involve computations of sense- or item-based percentage vectors of annotated features that are then usually explored using cluster-analytic methods; see Gries (2010, pp. 326–328) for the complete BP protocol. Finally, note that our claim is *not* that association rules are superior to these approaches – our claims are (i) that association rules are complementary because they approach things in ways that differ from both and (ii) that they are superior to the kind of (monofactorial) collostructional studies that has been predominant for the last 20 years.

For implementation, we used the well-known apriori algorithm in the R package *arules* (Hahsler et al., 2005, see also <https://github.com/mhahsler/arules>) with the support and confidence thresholds discussed above. The first general application generated a set of 12,927 rules, but only 2221 of those featured the variable *CONSTRUCTION* in the RHS, and only 142 of those were non-redundant. Before we turn to the full results, we discuss an example – namely, the rule in Table 2.

Table 2. An association rule

Construction	GO (<i>a</i>)round and V
COLLEXEME	
COLLIGATION	modal
DISCPROS	
MODALITY	deontic
POLARITY	affirm
REGISTER	
SEMPROS	
SPEECHACT	
Frequency	6
Support	0.011
Confidence	0.545
Lift	2.938

This rule is to be understood as follows: There are 11 instances of the *COLLIGATION modalized VP*, *deontic MODALITY*, and *affirmative POLARITY* (${}^6/_{0.545}$), and 6 of those are with the construction *GO (*a*)round and V*. Therefore, *support* is ${}^6/_{544} \approx 0.011$ and *confidence* is ${}^6/_{11} \approx 0.545$. The *lift*-value shows that the 6 instances are nearly three times as many as expected by chance, which makes this one of the stronger rules in the data. Also, it is important to note that the rule is ‘agnostic’ with regard to all other variables, meaning it could be phrased as follows:

If *COLLIGATION* is *modalized VP* and
MODALITY is *deontic* and
POLARITY is *affirmative*,
then *CONSTRUCTION* is *GO (*a*)round and V*
(no matter – the agnostic part – the values of *COLLEXEME*, *DISCPROS*, *REGISTER*, ...).

In the following section, we present the results and discuss them.

5. Findings

Consider Table 3 which lists all association rules for GO (a)round and V.

Table 3. Association rules for GO (a)round and V

COLLEXEME											
COLLIGATION	c_fut		inf		c_fut		inf		modal	s_past	s_past
DISCPROS										neutral	
MODALITY										deontic	
POLARITY	affirm	affirm							affirm		
REGISTER			spoken		spoken				spoken	spoken	
SEMPROS											
SPEECHACT	comm		statem								
Frequency	7	8	6	8	6	6	6	8	12		
Support	0.013	0.015	0.011	0.015	0.011	0.011	0.011	0.015	0.022		
Confidence	1.000	0.727	0.667	0.571	0.545	0.545	0.533	0.522			
Lift	5.386	3.917	3.591	3.078	2.938	2.938	2.873	2.81			

Before we contrast this construction's patterning with the other construction, let us make some general observations. Firstly, there are no preferred COLLEXEMES or SEMANTIC PROSODIES, which is in part due to the low number of instances of the construction. This also suggests that a traditional distinctive collexeme analysis, which *only* considers collexemes and no other features, applied to our data would result in very little, something to which we will return below. Secondly, there are several features where the construction is associated with only a single feature level:

- DISCOURSE PROSODY, only the level *neutral* is attested in a rule;
- MODALITY, only the level *deontic* is attested in a rule;
- REGISTER, only the level *spoken* is attested in a rule;
- POLARITY, only the level *affirmative* is attested in a rule (in fact, the construction only appears in affirmative contexts).

With regard to SPEECH ACT, both *commissives* and *statements* are attested in rules, and we also find several types of COLLIGATION in rules. However, these are what in a modeling context would be main effects – for a truly multivariate perspective (or a multi-dimensional₂ in the sense of Gries 2010, p.340–342), it is more interesting to review the clustering of features, which often necessitates interac-

tive and iterative heuristic sorting of the clusters revealed by the rules to find ‘the best way’ of summarizing the oftentimes multiply overlapping results. Doing so, we find that this construction is attracted by the following features:

- *controlled future tense* but especially when also with *affirmative* POLARITY;
- *affirmative* POLARITY more generally, but especially when the construction is also used with *deontic* MODALITY and *modalized VPs*;
- *spoken* REGISTER, but especially with *infinitives* and *simple past*.

The other construction, GO (*a*)*round Ving*, has a much wider range of rules. We begin with what is closest to a traditional distinctive collexeme analysis by discussing its associations to verbal collexemes. The construction is preferred by the following verbs: *do*, *kill*, *look*, *pretend*, *talk*, and *think* regardless of any other features. This does not mean that they do not co-occur with other features but that the presence of these collexemes is enough to make GO (*a*)*round Ving* the preferred construction. Other verbs prefer this construction as well, but in smaller subsections of the data:

- *say* generally prefers this construction, but especially so with *present participles* or in *fiction*;
- *try* generally prefers this construction, but especially so in the *spoken* REGISTER;
- particularly noteworthy is *tell*, which generally prefers this construction as well, but especially so with *do-insertion*, in *fiction*, and in *commissives*.

Examples (10–11) illustrate *say* in configurations with *present participles*. In (10), *say* occurs in the construction with GO being realized as a *present participle* by virtue of being the main verb in the progressive construction. In (11), GO is realized as a *present participle*, as *going around saying...* serves as an *ing*-clause:

- (10) Hillary Clinton is going around saying that it is going to be her number one job as president. (COCA 2015 SPOK Anderson Cooper)
- (11) There’s nothing more grotesque than somebody going around saying, “I’m a writer. I’m a writer. I’m a writer.” (COCA 2015 MOV The End of the Tour)

In both cases, the present participial realization of GO contributes a stronger sense of unboundedness to the iterative action expressed by the construction. In (12–13), we see an example of *say* in the construction occurring in the *fiction* REGISTER:

- (12) I mean, they couldn’t have the director of the museum going around saying that the single most valuable painting in the place was fake. (COCA 2016 FIC Fantasy/SciFi)

- (13) Jim's been going around saying he should kick your ass.
(COCA 2013 FIC ParisRev)

GO is also realized by *present participles* in (12–13), adding a similar sense of unboundedness. While the construal of DISAPPROVAL is only overt in (11), reflected in the overall negative tone of the utterance with its use of *grotesque* as a descriptor, there is arguably an underlying implied disapproval in all four examples. In (14), the occurrence of *try* in the construction in the *spoken* REGISTER is exemplified:

- (14) Harvey Weinstein went around trying to get nondisclosures too, right?
(COCA 2017 SPOK FOX: Ingram Angle)

Examples (15–17) illustrate *tell* occurring in the construction with *do-insertion*, in *fiction*, and in *commissives*:

- (15) I don't go around telling people that I care about them. I show that I care about them, first.
(COCA 2012 SPOK ABC_ThisWeek)
- (16) Maybe he was a white guy with dark skin and a ponytail going around telling the wasichus he was an Indian.
(COCA 2013 FIC VirginiaQReview)
- (17) I just don't go around telling my life story to every good looking man come along.
(COCA 2019 FIC BlackRenaissanc)

In (16–17), *tell* occurs in the construction in *fiction*, and, in (15) and (17), both the *tell* + *commissives* and the *tell* + *do-insertion* configurations are documented. In all instances of *tell* occurring in the construction with *do-insertion*, the purpose of *do-insertion* is the enablement of negative polarity which we see in both (15) and (17). Regarding *commissives*, in uttering (15) and (17), the speakers not only make statements on their own behaviors, they also commit themselves to not engaging in those behaviors.

Particularly noteworthy is the fact that this construction is preferred by many verbs with 'unpleasant' SEMANTIC PROSODIES involving ACCUSATION, DESTRUCTION, IMMODESTY, THREAT, and VIOLENCE, all of which have general preferences for this construction (i.e. again regardless of other features), but some SEMANTIC PROSODIES are also bundled in configurations with other features. VIOLENCE, the most frequent SEMANTIC PROSODY bundled with this construction, for example, is especially bundled with the REGISTERS of *fiction* and *TV* and COLLIGATIONS of *do-insertion*, *modalized VP* and *simple present*. These configurations are exemplified in the following examples:

- (18) You think I would last long in this profession if I went around braining people with a baseball bat?
(COCA 2017 FIC Bk:RepoMadness)

- (19) You can't go around killing kids!
(COCA 2017 TV It's Always Sunny in Philadelphia)
- (20) Con men don't go around killing people (COCA 2012 SPOK CBS_48 Hours)
- (21) Since when does a pacifist go around breaking other bloke's noses?
(COCA 2013 TV Miss Fisher's Murder...)
- (22) He couldn't go around hacking off English nobles' limbs.
(COCA 2011 FIC Bk:TamedByHighlander)
- (23) Matt, you can't go around hitting people just because you don't like what they say.
(COCA 2010 MOV Breaking the Press)
- (24) First of all, Nancy, nobody goes around killing people simply because their allowance is being cut.
(COCA 2015 SPOK CNN: Nancy Grace)

This preference for 'unpleasant' scenarios corroborates the idea that the construction generally has a negative slant to the point that it is used to express a stance of DISAPPROVAL.

Correspondingly, the rules for this construction also have a strong preference for *negative* DISCOURSE PROSODY: 20 rules involve *negative* DISCOURSE PROSODY (mostly in *fiction*, *movies*, and *TV*, never in *magazines*), but only 4 involve *positive* ones – 16 *neutral* ones are dispersed across REGISTERS; they involve all rules with *magazines* but none with *TV*. *Negative* DISCOURSE PROSODY prefers this construction in registers featuring fictional representations of spoken language. Consequently, the occurrence of the construction in negative contexts may be more of a feature of fictional representations of spoken language than of actual spoken language.

In terms of SPEECH ACTS, *directives* and *commissives* are most attested in rules involving this construction and never with *positive* DISCOURSE PROSODY. We saw examples of the use of the construction in *commissive* SPEECH ACTS in (15) and (17), and there is a further example in (25):

- (25) Well, I don't go around assaulting people (COCA 2017 TV Doubt)

This use of the construction (*do-support* + *negation*) is a very common means of performing self-prohibitive *commissive* SPEECH ACTS in the corpus. It can also be used to perform other-prohibitive *directive* SPEECH ACTS, as seen in (26):

- (26) Well, cap'n, an old Indian once told me that when the Great Spirit gives you a horse, you don't go around looking for another dog instead.
(COCA 2010 FIC Bk:SwordsSaddles)

However, *deontic* MODALITY is a more common way of using the construction to perform other-prohibitive *directive* SPEECH ACTS as we saw in (19) and (22–23). It

makes sense that self-prohibitive *commissive* and other-prohibitive *directive* uses of the construction do not prefer *positive DISCOURSE PROSODY*, because it is more concordant with most axiological systems to discourage negative behavior rather than positive behavior.

In terms of *COLLIGATIONS*, several rules involve *do-insertion*, as seen in multiple examples already, and *modalized VP*, as seen in (19) and (22–23) above and (27–29) below; the latter especially are particular noteworthy in how they never involve *positive DISCOURSE PROSODY* and are preferred in *movies, fiction, and TV*:

(27) You can't go around pointing guns at kids (COCA 2015 MOV Dementia)

(28) Ma, you know you can't go around pulling guns on people like that.
(COCA 2011 FIC Bk:RayHope)

(29) You can not just go around putting socks wherever you want.
(COCA 2018 TV LA to Vegas)

Finally, the only level of *MODALITY* preferring this construction is *deontic*, and again never with *positive DISCOURSE PROSODY* and often with directives, as seen in Examples (19), (22–23) and (27–29). Epistemic modality does occur, but not frequently enough to serve as a distinctive feature or even part of a distinctive configuration. The complete absence of *epistemic MODALITY* from association rules suggests that deonticity is a more important functional element of the two constructions than epistemicity.

Figures 1 and 2 provide visual representations of feature interactions associated with the two constructions.

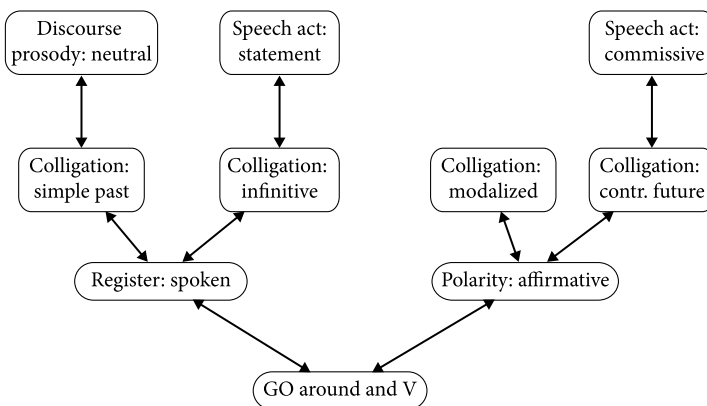


Figure 1. Visual representation of GO (a)round and V

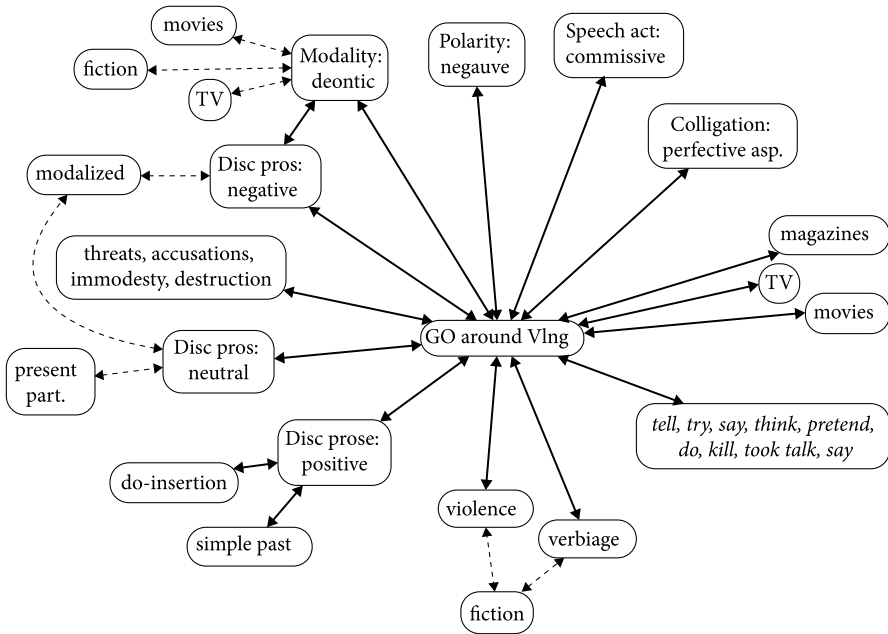


Figure 2. Visual representation of GO (*a*)round Ving

Figure 1 captures all feature interactions in GO (*a*)round and V (seen in Table 3) and shows which configurations prefer the construction, (for instance, *neutral discourse* PROSODY + *simple past* + *spoken* REGISTER or *modalized VP* + *affirmative* POLARITY). Since GO (*a*)round Ving is associated with an abundance of rules, Figure 2 is naturally more complex and summarizes only the main trends among the features preferring this construction (for instance, *deontic* MODALITY tends to occur with *negative* DISCOURSE PROSODY and in the *movies*, *fiction*, and *TV* REGISTERS). Here, a solid line indicates a strong connection (when a set of features occur primarily with one another and not with other features) among features, while a punctuated one indicates a weak connection (when there is more variability in terms of featural cooccurrence).

6. Concluding remarks

The study presented here has a number of theoretical and methodological implications. Our findings suggest that the two constructions are distinguished in terms of several sets of interactions (rather than just in terms of colllexemes), displaying various degrees of complexity, between the eight categories applied in this study: in some cases, a single feature is distinctive in itself, but in others it is a

configuration of features that is distinctive, and these configurations can, but need not, involve collexemes; the ability to be able to discover configurations without collexemes is already the first big advantage over traditional monivariate collexeme analysis. This has implications for not only our understanding of the two constructions and how they relate to one another. It arguably also has some implications for usage-based construction-grammatical theory as such.

Our analysis suggests that the concept of distinctiveness must necessarily be expanded to include a construction's broader context of usage. This is in line with Kemmer and Barlow's (2000) usage-based tenet that context plays a crucial role in the operation of the language system. It should follow from this tenet, then, that, in usage-based construction grammar, constructional features occur along multiple dimensions both within the internal constructional structure and in various recurring contextual layers.

Consequently, construction-grammatical theory might benefit from reintroducing Fillmore's (1988) distinction between internal and external constructional properties (or at least something similar). What is important is that contextual features can be described as constructional properties and part of speakers' operational knowledge of constructions on a par with, but not identical to, structural ones.

Concerning the notion of the construction itself, the multivariate analysis presented here suggests that it is insufficient to view constructions merely as pairings of form and conceptual meaning (which many construction grammarians seem to do in practice despite the inclusion of discourse-pragmatics construction grammar theory) and that it is particularly insufficient if 'form' is operationalized on the basis of just a single constructional slot. In posing contextual features as *bona fide* constructional features, our study underlines the necessity of construction grammarians taking more seriously the tenet that constructional meaning is not only conceptual-semantic but also pragmatic and discourse-functional. Constructional meaning, or better function, must necessarily be viewed as multi-layered such that conceptual meaning is just one layer, while discourse-pragmatics is another, and social meaning yet another and so on and so forth, and, importantly, not all layers need apply to all constructions (this way, a construction may be a construction even without conceptual meaning if it has some other conventionalized function).

All of this leads towards a multidimensional understanding of constructions as entrenched routines emerging in speech communities, to borrow some of Croft's (2005, p. 274) wording, in which the construct-i-con contains information on recurring contextual patterns as *bona fide* constructional knowledge. Consequently, constructions could be described in terms of what Herbst (2018, p. 12) calls collo-profiles – representations of 'collo-items' occurring with constructions. Essentially, our findings in this study constitute a step towards building

what we can call distinctive collo-profiles pertaining to the two constructions. In our analysis, of course, the ‘collo-items’ go beyond lexemes that fill slots in the construction and include also elements in the surrounding co-text as well as even more abstract ones such as register.





To some readers, despite the above theoretical implications, this paper may seem overly descriptive. However, in the evaluation of the results, it is important to compare our results to what a traditional distinctive collexeme analysis (here based on the log-likelihood ratio G^2) would have produced, which is essentially these two lists of distinctive collexemes with no additional contextual, semantic, or discourse features at all:














- GO (*a*)*round* and V: *introduce, check, round* (and really not much more, given that all next verbs are hapaxes in this construction or have low and non-significant G^2 -values even if one does not adjust for multiple post-hoc tests);
- GO (*a*)*round* V-ing: *kill, do, pretend* (and really not much more, given that all next verbs have low and non-significant G^2 -values even if one does not adjust for multiple post-hoc tests).






In fact, Jensen (2024) is an example of the application of traditional distinctive collexeme analysis of these very constructions. While providing insight into the semantic-prosodic and discourse-prosodic characteristics of the constructions, Jensen (2024) would, by his own admission, obviously be a failure had he aimed at providing a richer description of the two constructions’ overall distinctiveness. This is because this more traditional analysis would see any *one feature in isolation*, but would by definition *not* be able to see any configurations, i.e. *combinations of features* or one feature limiting another (like GO (*a*)*round* and V preferring *simple past* but only in *spoken* register, or GO (*a*)*round* V-ing often having an unpleasant or even *violent* semantic prosody, but not in the *spoken* register – speaking/conversation arguably being the most fundamental use of language – but in *fiction, movies, and TV*). We submit that the new approach promoted here is *much* more informative and it is noteworthy to point out that it is more informative in spite of the fact that the data set is fairly small. In other words, we see the new approach is already besting the old one even in cases where one does not have much data, but it will of course beat the old approach even more if more data points are available because then the higher degree of resolution afforded by the much larger number of features being studied. This also means the present study is perfectly compatible with work like Herbst (2018), who has suggested (p.18) to transform Goldberg’s (2006, p.18) credo “it’s constructions all the way down” into “it’s collexemes (or items) all the way down”. In a way, our results are the response to Herbst (2018) calling for the description of constructions to include a ‘collo-profile’.

On a strictly methodological level, the approach can be modified and/or extended. As for modification, Olguín Martínez and Gries (2024) propose a similar approach using the statistical approach of a complete hierarchical configurational frequency analysis, an analysis used in a collostructional context for the first time in Stefanowitsch and Gries (2005), but not used much since then, especially not with more than three dimensions. As for extensions, especially for distinctive cases like the present one, it seems as if the kind of work proposed here begins to enter the domain or predictive modeling of alternation cases. Much like studies of particle placement (e.g. Gries, 2003) or the dative alternation (e.g., Bresnan et al., 2007) use a constructional choice as binary or categorical response variable and many different contextual predictors as predictors, so might a collostructional study like ours ultimately be modeled in a similar way (i.e., with GO (a)round and V vs. GO (a)round Ving as a binary response and all our features as predictors). Hampe and Gries (2018) use such an approach – a multinomial regression – to identify prototypical instantiations of eight complex-sentence configurations they call *meso constructions* and any collexeme study of the distinctive kind can be reframed to such a predictive modeling perspective. Whatever exact way researchers will choose, we do hope that *some* kind of multivariate approach will be able to breathe new life into collostructional studies, given how much more instructive the results of it can be.

References

- Bresnan, J., Cueni, A., Nikitina, T., & Baayen, H. (2007). Predicting the dative alternation. In G. Boume, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.
-  Croft, W.A. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
-  Croft, W.A. (2005). Logical and typological arguments for Radical Construction Grammar. In J.-O. Östman (Ed.), *Construction grammars: Cognitive grounding and theoretical extensions* (pp. 273–314). Amsterdam: John Benjamins.
- Davies, M. (2008–). *The Corpus of Contemporary American English (COCA)*. Retrieved from <https://www.english-corpora.org/coca/>
-  Fillmore, C. J. (1988). The mechanics of “Construction Grammar”. *BLS*, 14, 35–55.
-  Fillmore, C. J., Kay, P., & O’Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language*, 64(3), 501–539.
- Gilquin, G., & Gries, S. Th. (under review). Collostructional analysis. In H. Nesi & P. Milin (Eds.), *International encyclopedia of language and linguistics* (3rd ed.). Amsterdam: Elsevier.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: Chicago University Press.

- Goldberg, A. E. (2006). *Constructions at work*. Oxford: Oxford University Press.
- Gries, S. Th. (2003). *Multifactorial analysis in corpus linguistics: A study of Particle Placement*. London & New York: Continuum Press.
-  Gries, S. Th. (2010). Behavioral Profiles: A fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon*, 5(3), 323–346.
-  Gries, S. Th., & Stefanowitsch, A. (2004a). Extending collocation analysis: A corpus-based perspective on ‘alternations’. *International Journal of Corpus Linguistics*, 9(1), 97–129.
- Gries, S. Th., & Stefanowitsch, A. (2004b). Co-varying collexemes in the *into*-causative. In M. Achard & S. Kemmer (Eds.), *Language, culture, and mind* (pp. 225–236). Stanford, CA: CSLI.
-  Hahsler, M., Gruen, B., & Hornik, K. (2005). arules – A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software*, 14(15), 1–25.
-  Hampe, B., & Gries, S. Th. (2018). Syntax from and for discourse II: More on complex-sentences as meso-constructions. In B. Hampe & S. Flach (Eds.), *Yearbook of the German Cognitive Linguistics Association* (pp. 115–142). Boston: De Gruyter Mouton.
-  Herbst, T. (2018). Is language a collocation? A proposal for looking at collocations, valency, argument structure and other constructions. In P. Cantos-Gómez & M. Almela-Sánchez (Eds.), *Lexical collocation analysis: Advances and applications* (pp. 1–22). Cham: Springer.
- Hilpert, M. (2012). Diachronic collocation analysis: How to use it and how to deal with confounding factors. In K. Allan & J.A. Robinson (Eds.), *Current methods in historical semantics* (pp. 133–160). Berlin: Mouton De Gruyter.
-  Hilpert, M. (2019). *Construction grammar and its application to English* (2nd ed.). Edinburgh: Edinburgh University Press.
-  Hunston, S., & Francis, G. (1999). *Pattern grammar: The principles and practices of corpus-driven grammar*. Amsterdam: John Benjamins.
-  Jackendoff, R. (1997). Twistin’ the night away. *Language*, 73(3), 534–559.
-  Jensen, K. E. (2017). Kan konstruktioner have semantisk prosodi? [Can constructions have semantic prosody?]. *Ny Forskning i Grammatik*, 24, 62–85.
-  Jensen, K. E. (2024). *Well, maybe you shouldn’t go around shaving poodles*: Collocational semantic and discursive prosody in the *go (a)round Ving* and *go (a)round and V* constructions. *Corpus Linguistics and Linguistic Theory*.
- Kemmer, S., & M. Barlow (2000). Introduction: A usage-based conception of language. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. vii–xxviii). Stanford: Stanford University Press.
-  Olguín Martínez, J.F., & Gries, S. Th. (2024). If not for-if it weren’t/wasn’t for counterfactual constructions: A multivariate extension of collocation analysis. *Cognitive Semantics*, 10(2), 158–189.
-  Patten, A. L. (2014). The historical development of the *it*-cleft: A comparison of two different approaches. In N. Gisborne & W.B. Hollmann (Eds.), *Theory and data in cognitive linguistics* (pp. 87–114). Amsterdam: John Benjamins.
-  Schönefeld, D. (2013). *It is ... quite common for theoretical predictions to go untested* (BNC_CMH). A register-specific analysis of the English *go un-V-en* construction. *Journal of Pragmatics*, 52, 17–33.

-  Stefanowitsch, A. (2000). The English GO-(PRT)-AND-VERB construction. *BLS*, 25, 259–270.
-  Stefanowitsch, A., & Flach, S. (2020). “Too big to fail but big enough to pay for their mistakes”: A collostructional analysis of the patterns [too ADJ to V] and [ADJ enough to V]. In G. Corpas & J.P. Colson (Eds.), *Computational and corpus-based phraseology* (pp. 248–272). Amsterdam: John Benjamins.
-  Stefanowitsch, A., & Gries, S. Th. (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243.
-  Stefanowitsch, A., & Gries, S. Th. (2005). Covarying collexemes. *Corpus Linguistics and Linguistic Theory*, 1(1), 1–43.
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.
-  Traugott, E. C., & Trousdale, G. (2014). Contentful constructionalization. *Journal of Historical Linguistics*, 4(2), 256–283.

Address for correspondence

Kim Ebensgaard Jensen
 Department of English, Germanic, and Romance Studies
 University of Copenhagen
 Emil Holms Kanal 6
 DK-2300 København S
 Denmark
 ebensgaard@hum.ku.dk

Biographical notes

Kim Ebensgaard Jensen is an Associate Professor of English Linguistics in the Department of English, Germanic, and Romance Studies at University of Copenhagen in Denmark. His research interests include cognitive linguistics, construction grammar, usage-based linguistics and the discourse-cognition-culture intersection as well as corpus linguistics and, more broadly, the Digital Humanities.

 <https://orcid.org/0000-0001-6952-2366>

Stefan Th. Gries is a Professor of Linguistics at the University of California, Santa Barbara, and a Chair of English Linguistics (25%, Corpus Linguistics with a focus on quantitative methods) at the Department of English at Justus-Liebig-Universität Giessen. His research interests include corpus linguistics and quantitative methods as well as usage-based linguistics.

 <https://orcid.org/0000-0002-6497-3958>

Publication history

Date received: 14 May 2024

Date accepted: 5 February 2025

Published online: 9 April 2025