The similative-pretence alternating pair and filler-slot relations

A revised version of distinctive collexeme analysis

Jesus Olguin Martinez[1] and Stefan Th. Gries[2,3]
[1] Illinois State University | [2] University of California, Santa Barbara | [3] JLU Giessen

The present paper investigates the similative-pretence alternating pair in Mexican Spanish (*como/igual que* and *como si* constructions) based on the analysis of 1362 instances from The Corpus del Español NOW (News on the Web) corpus. We apply a revised version of distinctive collexeme analysis to our concordance data to explore the variety of different verbs that can occur in the first slot of similative and pretence constructions. Specifically, we use (i) a new measure that distinguishes the attraction that a verb lemma exerts on a construction from the attraction that a construction exerts on a verb lemma and (ii) bootstrapping on the level of files to, for the first time, provide proper confidence intervals in the context of collostructional studies. The from-verb lemma-to-construction analysis shows that there is a significant attraction of epistemic lemmas (e.g., *parecer* 'to seem') to occur in *como* 'like', and of mistaken identity lemmas (e.g., *actuar* 'to act') to appear in *como si* 'as if' constructions. The from-construction-to-verb lemma analysis, on the other hand, demonstrates that *igual que* 'like' significantly attracts epistemic perception verbs (e.g., *ver* 'to look'), and *como si* 'as if' constructions significantly attract the verb lemma *sentir* 'to feel.'

Keywords: comparative construction, Spanish, similarity, collexeme analysis, filler-slot relations

1. Introduction: Setting the stage

Among the most basic cognitive processes common to all human beings are those of comparing one thing to another and of making hypotheses on the basis of available evidence (Fortescue 2010: 117). This mental act of comparison often finds

https://doi.org/10.1075/cf.23020.olg | Published online: 28 February 2025
Constructions and Frames ISSN 1876-1933 | E-ISSN 1876-1941

its linguistic encoding in similarity constructions (e.g., *the boy runs like a hare*), which belong to the functional domain of qualitative comparison and "bring together the two terms of the comparison on the basis of similarity or likeness" (Fuchs 2014: 133). Various subtypes of similarity have been taken into consideration, including but not limited to physical resemblance (similarity in shape or other visual property), functional resemblance (acting in the manner of something else or having the same status), and evaluative resemblance (similarity in value) (Fortescue 2010: 117). For the most part, they have been explored from a typological perspective. For instance, Haspelmath & Buchholz (1998) analyze equative (e.g., *he is as tall as Mary*) and similative constructions (e.g., *she jumps like a kangaroo*) based on data from 43 European languages. Olguín Martínez (2021) explores pretence constructions (e.g., *she treats me as if I were a stranger*) in a sample of 61 languages. Despite this wealth of typological research, few studies have explored similarity constructions in individual languages using corpora. This type of corpus study in individual languages is important in that it can lead us to uncover new correlations between linguistic features used in different but related similarity constructions and new diachronic developments.

The present study seeks to contribute to fill this gap with a corpus-based analysis of similarity constructions in Mexican Spanish. Similarity is expressed in four different ways in this variety of Spanish. First, there are constructions used to indicate that two referents have a gradable property to the same degree (see (1)). These constructions are known in the literature as **equative constructions** (see Haspelmath & Buchholz 1998; Henkelmann 2006; Jiménez Juliá 2003).

(1) *ser* + (*tan*) + ADJ + *como* ____ (e.g., *el hombre es tan alto como Roberto* 'the man is as tall as Roberto').

Second, there are constructions as in (2) depicting an action or state identical to that of the main clause. These constructions have been called **relative *como* 'like' constructions** (Jiménez Juliá 2003). Hetterle (2015: 54) mentions that they answer the question *how?* in that they describe the character of a situation comparing it to a real situation (Jiménez Juliá 2003). Relative *como* 'like' constructions can commonly be paraphrased, in many languages, with a relative clause that appears with a generic head noun, such as *way/manner* (Olguín Martínez 2021). Accordingly, (2) could be paraphrased as *Juan trabaja de la misma manera que lo hace su hermano* 'John works the same way in which his brother works.'

(2) ____ *como* ____ (e.g., *Juan trabaja como lo hace su hermano* 'John works like his brother works').

Third, there are constructions including both an identificational relation ('X is Y') and a dative-based reference toward the experiencer of the similarity relation

('X is Y for Z'), as in (3) and (4). These constructions are known in the literature as **similative constructions** (see Haspelmath & Buchholz 1998). In these patterns, the concept of likeness is fully inferential (Trujillo 1990). This indicates that these expressions may be derived metonymically or metaphorically in that "they represent fossilized patterns of cognitive processes conventionalized over time" (Schulze 2017: 36).

(3) _____ *igual que* + NP (e.g., *mi abuelo nada igual que un pez* 'my grandfather swims like a fish').

(4) _____ *como* + NP (e.g., *mi abuelo nada como un pez* 'my grandfather swims like a fish').

Fourth, **pretence constructions** typically portray an imagined ('do X as if it was caused by Y') or counterfactual ('do X as if Y were true') meaning (Jiménez Juliá 2003; Darmon 2017: 372–373). These constructions, like (5), are similar to similatives in that the concept of likeness is fully inferential.[1]

(5) _____ *como si* _____ (e.g., *mi abuelo nada como si fuera un pez* 'my grandfather swims as if he were a fish').

The Spanish similarity constructions discussed above (_____*como/igual que*+NP) and the pretence constructions (_____*como si*_____) are similar to one another. For instance, the construction: *mi abuelo nada como/igual que un pez* 'my grandfather swims like/as a fish' is similar to *mi abuelo nada como si fuera un pez* 'my grandfather swims as if he were a fish.' Such pairs of functionally — meaning semantically or information-structurally — more-or-less equivalent expressions are ubiquitous in human language.

The present paper investigates the similative-pretence alternating pair in Mexican Spanish. To keep the scope of the discussion manageable, we focus on instances in which the element following the *como/igual que* and *como si* lexical items is a Noun Phrase (NP) or a Locative Noun Phrase (LOC NP). This means that relative *como/igual que* constructions and clausal *como si* constructions are excluded from the present study (see Jiménez Juliá 2003: 11 for discussion of these constructions):

(6) *El perro come como lo hacen las gallinas.*
'The dog eats the same way hens eat.'

(7) *El perro come igual que lo hacen las gallinas.*
'The dog eats the same way hens eat.'

1. Pretence constructions can also be encoded with the expression *como que*, as in *pretende como que no tiene dinero* 'he pretends not to have any money' (Jiménez Juliá 2003).

(8) *El hombre grita como si alguien lo estuviera matando.*
'The man screams as if someone were killing him.'

The following *como/igual que* and *como si* constructions with NPs and LOC NPs are taken into account in the present investigation:

(9) a. *como* 'like' construction with NP
_____ *como* +NP (e.g., *se comporta como un doctor* 'he acts like a doctor').
b. *como* 'like' construction with LOC NP
_____ *como* +LOC.NP (e.g., *se siente como en la playa* 'it feels like at the beach').

(10) a. *igual que* 'like' construction with NP
_____ *igual que* +NP (e.g., *se comporta igual que un doctor* 'he acts like a doctor').
b. *igual que* 'like' construction with LOC NP
_____ *igual que* +LOC.NP (e.g., *se siente igual que en la playa* 'it feels as if were on the beach').

(11) a. *como si* 'as if' construction with NP[2]
_____ *como si* +*ser*.SUBJ + NP (e.g., *se comporta como si fuera un doctor* 'he acts as if he were a doctor').
b. *como si* 'as if' construction with LOC NP
_____ *como si* +*estar*.SUBJ + LOC.NP (e.g., *se siente como si estuvieramos en la playa* 'it feels as if we were on the beach').

In analyzing the similative-pretence alternating pair in Mexican Spanish, we are paying particular attention to the range of verbs that can occur in the first slot in the *como/igual que* and *como si* constructions in (9) to (11). While the first slot in these constructions can be filled by a variety of different verbs (e.g., *sentir* 'to feel', *es* 'it is', *ver* 'to look', *comportar* 'to behave', *actuar* 'to act', *sonar* 'to sound', and *parecer* 'to seem'), much less is known about both the specific associations between these constructions and the verbs they take and, more importantly, what such patterns reveal about these constructions' functions, and better descriptions of these patterns can in turn inform later analyses of, say, their acquisitional paths or their processing. The co-occurrence patterning of lexemes and constructions is functionally motivated (Goldberg 1995: 50; Gries & Stefanowitsch 2004: 99), which gives rise to a joint distribution of lexemes in constructions that are known in the literature as **filler-slot relations** (see Goldberg 1995; Fillmore & Kay 1999;

2. As can be seen in the constructions in (11), the lexical item *como si* is not immediately followed by a NP or a LOC NP. Instead, it is followed by the copula *ser* or *estar* 'to be' that must be in the subjunctive.

Hilpert 2013, 2021; Diessel 2019, 2020).[3] In a usage-based framework, such probabilistic associations constitute part of each language user's individual and ever-changing exemplar-based representation of linguistic knowledge (Beckner et al. 2009) and, in that framework, such probabilistic associations have often been studied using methods from the family of collostructional analysis. This family of methods is based on one of the most central working axioms of corpus linguistics, namely, the distributional hypothesis (Harris 1954) that distributional similarity — i.e., patterns of similarity in frequencies of co-occurrence — reflect functional similarity — i.e., commonality of semantic, pragmatic, information-structural, or other characteristics. In other words, frequency of co-occurrence reflects, and is thus a diagnostic of, similarity of meaning: the fact that verbs like *give* and *tell* are highly frequent in the ditransitive and, in terms of collostructional results, highly attracted to the ditransitive reflects the high degree of similarity of the semantics of the ditransitive construction and the semantics of the transfer meaning of *give* and the metaphorical transfer meaning of *tell* (the famous conduit metaphor of communication is transfer).

Thus, since we (i) follow the usage-based tradition with a corpus-based approach and (ii) assume the semantics of the *como/igual que* and *como si* constructions to be different, we consider it reasonable to assume that these constructions will attract verbs with different semantics, which in turn is something that a (distinctive) collexeme should be able to diagnose. Specifically, we assume the former construction means 'to give the same appearance as something/someone' (see (9), (10)) whereas the latter means 'to imitate, pretend, aspire to the behavior of something/someone' (see (11) and Chamoreau 2017). For the *como/igual que* constructions in (9) to (10), we expect they will prefer to occur with epistemic judgment predicates, such as *parecer* 'to look', *sonar* 'to sound', etc., whose meaning harmonizes with the meaning of *como/igual que* constructions given that they require speakers to provide lexical information regarding their judgments about the status of the proposition ('X gives the same appearance as Y'). Put differently, speakers need to indicate the type of evidence they have to say that 'X resembles Y'. The verbs mentioned above belong to a group of expressions that Palmer (2001: 24) calls **deductive epistemic judgements**, but at the same time they indicate that the judgement is based on some sort of available evidence and hence are also evidential in nature.

For the *como si* constructions in (11), on the other hand, we expect that they will prefer to occur with **mistaken identity verbs**, such as *actuar* 'to act' and

3. The fact that the meaning of a construction tends to harmonize with the meanings of the lexical elements that typically occur in it is referred to as the semantic coherence principle (Goldberg 1995: 50).

comportarse 'to behave', which harmonize with the semantics of the *como si* constructions in (11), given that they mean to behave or comport oneself in imitation of something else. In this scenario, something about the behavior of 'X' resembles that of 'Y'.

Table 1 provides a summary of the hypotheses formulated in the previous paragraphs. The lexical and grammatical items refer to the constructional components of *como/igual que* and *como si* 'as if' constructions. The following columns contain information regarding the hypothesized meaning/semantics of each construction and information regarding the lexical items that we hypothesize may occur in the first slot of *como/igual que* and *como si* 'as if' constructions (hypothesized co-occurrence preference).

Table 1. Summary of hypotheses

Construction components		**Hypothesized meaning**	**Hypothesized co-occurrence preference**
Lexical	**Grammatical**		
como	+ NP	'to give the same appearance as something/someone'	epistemic judgment verbs
'like'	+ LOC.NP		
igual que	+ NP		
'like'	+ LOC.NP		
como si	+ NP	'to imitate, pretend, aspire to the behavior of something/someone'	mistaken identity verbs
'as if'	+ LOC.NP		

In the present paper, we explore whether these predictions hold based on the analysis of 1362 instances from The Corpus del Español NOW corpus (News on the Web). We go methodologically beyond previous work by applying a revised version of distinctive collexeme analysis (Gries & Stefanowitsch 2004) to our data, specifically a newly-developed extension of the analysis that adds the equivalent of confidence intervals to the collexeme strengths usually reported (which by definition incorporates the dispersion of the filler-slot types across the corpus).

The plan of the paper is as follows: in the next section, we provide the methods and results of the present study. We first introduce our data (Section 2.1) and then briefly walk the reader through the general operating principles of collostructional analysis and the revised version used here (Section 2.2). Finally, Section 3 offers an interpretation of our results while Section 4 concludes with pointers toward theoretical issues and a number potentially fruitful areas for future research.

2. Methods and results

This section introduces the corpus data used here to analyze the similative-pretence alternating pair in Mexican Spanish and outlines the revised version of distinctive collexeme analysis employed in the present research. The term **collexeme analysis** refers to a family of methods for the study of interrelations between constructions and their lexical items. Since its introduction over twenty years ago, this method has been popular both among construction grammarians and corpus linguists. Collexeme analysis covers three different techniques. First, **simple collexeme analysis** studies one slot in one construction and the words occurring in that slot. Second, **distinctive collexeme analysis** is a variant aimed at uncovering differences in the statistical associations that hold between a particular slot in two (and theoretically more) related constructions. Third, **covarying collexeme analysis** identifies the association strength between pairs of lexical items occurring in two different slots of the same construction (see Stefanowitsch & Gries 2003; Gries & Stefanowitsch 2004 on these various techniques).

2.1 Corpus data, data extraction, and annotation

The Spanish corpus chosen for the investigation of the similative-pretence alternating pair in Mexican Spanish was The Corpus del Español NOW (News on the Web) corpus. It contains about 7.6 billion words from web-based newspapers and magazines in 21 Spanish-speaking countries from 2012 to 2019. This corpus represents a growing internet resource: about four to five million words are added to this corpus each day. The data used here consist for the most part of web-based material (e.g., newspapers), a written genre that has never been explored in the context of the similative-pretence alternating pair.

Regarding data extraction, a near-exhaustive concordance of the similative-pretence alternating pair was performed by searching The Corpus del Español NOW for the forms *como/igual que* and *como si*. This generated a large sample in which the lexical items *como/igual que* and *como si* were followed by NPs, LOC NPs, and clauses, which was then trimmed down to exclude constructions in which *como/igual que* and *como si* were followed by clauses (e.g., *el perro come como lo hacen las gallinas* 'the dog eats the same way hens eat'). Note that, to avoid *como si* constructions followed by a clause (e.g., *el hombre grita como si alguien lo estuviera matando* 'the man screams as if someone were killing him'), we added the Spanish copula in the subjunctive to the *como si* in our search (i.e., *como si fuera(n), como si hubieras sido, como si hubieran sido, como si estuviera(s), como si estuvieran, como si hubiera(s) estado, como si hubieran estado*). Spot checks

show that this procedure is more efficient when searching for *como si* constructions appearing with NPs and LOC NPs. We also disregarded idiomatic *como/igual que* and *como si* constructions that have been identified in the literature (e.g., Batchelor & Pountain 2005) such as *caer como anillo al dedo* 'it's like hand in glove' and *sentirse como pez en el agua* 'to feel like a fish in water'. To determine whether a *como/igual que* and *como si* construction is idiomatic, linguists have not only relied on the non-compositionality criterion, but also on other criteria such as the formal fixedness of a construction, the degree of conventionalization of a construction, and the syntactic flexibility of a construction, i.e., the extent to which a construction licenses syntactic variations (see Wulff 2008: 1 for more detailed discussion of these criteria). The resulting database contains 1362 instances of the *como/igual que* and *como si* constructions including NPs and LOC NPs.

For these data, we then coded the relevant variables for our analysis: (i) verbs that can occur in the first slot in the *como/igual que* and *como si* constructions, (ii) whether the NP following the *como/igual que* and *como si* items was locative or non-locative, and (iii) the construction. These were manually annotated by inspecting each of the 1362 *como/igual que* and *como si* constructions. Table 2 shows the way we organized our data.

Table 2. Organization of the similative-pretence alternating pair data in the present study

Source	Example	Lemma	Locative	Construction
95 19-03-26 MX Milenio.com	*corre como novato que ataca y defiende*	*correr* 'to run'	no	*como* + NP
8 16-07-27 MX Libertad de Expresión Yucatán	*se ríe como si estu-viera en su fiesta de 15 años*	*reir* 'to laugh'	yes	*como si* + *ser* + LOC.NP
32 19-03-04 MX Independiente de Hidalgo	*piensa igual que un no rico*	*pensar* 'to think'	no	*igual que* + NP

We now turn to the revised version of distinctive collexeme analysis (Gries & Stefanowitsch 2004) used here.

2.2 The statistical analysis

The vast majority of collexeme analyses — simple or distinctive — have used a bidirectional association measure (AM) and, more specifically, a bidirectional AM that is ultimately based on a *p*-value/significance test, namely either $p_{\text{Fisher-Yates}}$ or the log-likelihood value G^2. This is suitable for many applied or completely

exploratory purposes. However, it comes with three potentially important problems, which come under the heading of **information loss due to conflation**. First, because the AM used is bidirectional, it cannot by definition distinguish whether

1. a verb lemma attracts a construction strongly while the construction attracts the verb lemma only (much more) weakly;
2. a construction attracts a verb lemma strongly while the verb lemma attracts the construction only (much more) weakly;
3. both construction and the verb lemma attract each other strongly.

Examples from collocational research for these three possibilities are *according to* and *upside down* (where the first word attracts the second, but not vice versa), *of course* and *for instance* (where the second word attracts the first, but not vice versa), and *bona fide* (where both attract each other). It would therefore be nice to be able to distinguish both directions of attraction.

Second, the use of *p*-value AMs has the advantage of being conveniently straightforward for exploratory tasks because such measures conflate both co-occurrence frequency and association strength into one convenient sortable statistic. The downside, however, is that it is also well-known from previous work (Evert 2009; Gries 2022a) that these AMs are also so highly correlated with co-occurrence frequency and, correspondingly, much less with actual association/contingency (as measured by an effect-size AM like the log odds ratio) that their value as a quantificational tool of association proper can be doubtful. Also, it is essential to point out that even if the AMs used in previous work were often based on significance tests, this does not mean they can be interpreted as proper *p*-values. This is not only because researchers have usually not corrected for multiple testing (though see Gries 2005), but also because the *p*-values derived from p_{FYE} or G^2 presuppose that the data points are all independent of each other, which comprehensively collected corpus examples hardly ever are; our approach below will consider this issue better.

Third and relatedly, nearly all work on co-occurrence in corpus linguistics has not provided any indication of the uncertainty that comes with one results, and that in spite of the facts that (i) of course every corpus linguist is well aware of the fact that the corpora we study are just samples that are hopefully representative of the larger population to which we wish to generalize and (ii) most corpus linguists are well aware of the utility of confidence or credibility intervals in many other contexts.

In this study, we follow recent discussions (Gries 2019, 2022a, 2023) and deviate from previous work to improve the analysis in all these regards:

1. We will use a directional AM that allows us to distinguish (i) the attraction that the verb lemma exerts on the construction from (ii) the attraction that the construction exerts on the verb lemma;
2. the specific AM we will use is based on the Kullback-Leibler divergence (KLD, also known as **relative entropy**). This is a measure that is much less correlated with the observed co-occurrence frequencies than the two above-mentioned AMs that have been most widely used in collostructional studies;
3. we will use a bootstrapping approach,[4] which will have three massive advantages:
   a. unlike nearly all other studies, we will actually be able to quantify the variability that comes with our AM results;
   b. since the sampling unit is the corpus files/sources, this approach will also incorporate in at least some way the dispersion of the lemma-construction co-occurrences in the corpus, which will help us distinguish interesting instances of frequent co-occurrence widely dispersed across a corpus from idiosyncractic and thus less interesting frequent co-occurrence due to a single speaker/author;
   c. since we are creating the null hypothesis distribution by respecting the distribution of the verb lemma-construction co-occurrences in the files — unlike p_{FYE} and G^2 — we really can interpret the results as reflecting a significance test and determine which associations are significantly different from 0.

How is the KLD computed? The general formula is represented here (see Cover & Thomas 2006: 9), where P represents a posterior/observed distribution and Q observes a prior distribution.

(12) $KLD\,(P \,||\, Q) = D_{KL}(P \,||\, Q) = \sum_{i=1}^{n} P_i \times log_2 \frac{P_i}{Q_i}$ (multiplier is set to 0 for any $P_i = 0$)

4. Bootstrapping is a statistical method that is often used to quantify the uncertainty of a statistical estimate (such as a confidence interval). We explain the exact steps of this approach for the current study in Section 2.2.1 below but the general logic of bootstrapping involves answering a simple question: what would the result have looked like if the (corpus) sample we have right now would have been (slightly) different? The way bootstrapping provides an answer is by creating, with random resampling from the data one actually has, a high number (e.g., 200 or 500) of (slightly) different data sets and computing the same statistical estimate (a distinctive collexeme strength) for each of these hypothetical data sets. As a result, one has one's actual result and, say, 500 also-possible/likely results, from which one can compute a confidence interval, which is useful to determine what, if any, significantly different effects (e.g., collexeme strengths) one finds in one's data. See Gries (2021: Section 5.2.1.5; 2022b) or Egbert & Plonsky (2020) for more details and examples.

In our context, the prior distributions Q will be either the frequencies of verb lemmas in our data (summed across all constructions) or the frequencies of constructions in our data (summed across all verb lemmas); these are the priors because they reflect the frequencies of the verb lemmas or constructions respectively regardless of the other element. The posterior distributions P, by contrast are the frequencies of the constructions for each verb lemma or the frequencies of the verb lemmas for each construction. Since our data involve what would be a multiple distinctive collexeme analysis — we have more than two verb lemmas and two constructional slots — our interpretation will be based on what one might call the contributions to KLD. The contributions to KLD are simply the individual summands in the above equation in (12) and are the analog to what residuals of, or contributions to, chi-squared are in a chi-squared test; for example, if P is c(0.5, 0.5) and Q is c(0.3, 0.7), then the first contribution to KLD is $0.5 \times \log2(^{0.5}/_{0.3}) = 0.36848$.

As a first step, we loaded the data and did a first descriptive cross-tabulation. From the residuals of an overall chi-squared test, we can see that 42 of the 132 cells feature an observed frequency that is greater than the expected one. Next, we had to compute the two directional associations lemma→construction and construction→lemma, but do so in a way that also provides us with a bootstrapped confidence interval; the next two sections discuss the procedure for each direction and outline the results we obtained.

2.2.1 *From verb lemma to construction*

Using R 4.3.1 (R Core Team 2023), we decided to use a bootstrapping number of 500 iterations, which, given the overall low frequencies in the frequency table of lemmas and constructions should be enough to give us a robust estimate of the uncertainty of the results. Specifically, to collect all results from the bootstrapping, we created a 3-dimensional array with 22 rows (one for each verb lemma), 6 columns (one for each construction), and 500 slices (one for each iteration); each of the 132 for an iteration was to be filled with the contribution to $\text{KLD}_{\text{lemma→construction}}$. Each iteration involved:

1. drawing a (replicably) random sample (with replacement) from all corpus files/sources;
2. retrieving all the corpus examples for these randomly selected corpus files/sources;
3. computing a verb lemma-by-construction matrix for the currently selected sources;

4. computing the contributions to $KLD_{lemma \rightarrow construction}$ by comparing posterior (the distribution of each verb lemma across the constructions) against the prior (the overall frequencies of the constructions in this iteration's sample);
5. saving the results in the relevant slice of our 3-dimensional collector array.

Then, from the collector array, we determine for each verb (i) the median bootstrapped association to each construction (i.e., the median of all 500 contributions to $KLD_{lemma \rightarrow construction}$) and (ii) the 95% CI for these 500 contributions to $KLD_{lemma \rightarrow construction}$, which we defined in the usual way, i.e., as the range from the 2.5% to the 97.5% quantiles of the contributions.

Then we plot the results for all 41 lemma-construction combinations with a positive median bootstrapped contribution to KLD; the restriction to the positive values means we are following the usual practice of collexeme analyses and are focusing our attention on cases where the verb lemma attracts, rather than repels, the construction. In the following plot, the *x*-axis represents the 41 lemma-construction combinations (referenced at the top of the plot and sorted within verb by collexeme strength), the *y*-axis represents the contributions to KLD, and the points represent the medians with the error bar representing the confidence interval (with significant collexeme strengths represented by an asterisk and with bold lines/intervals).

As is clear, while all these combinations involve attraction from the verb lemma to the construction, not all are significantly different from chance because many confidence intervals overlap with 0. The significant co-occurrence pairs are summarized in Table 3 grouped by construction.

Table 3. Significant co-occurrences (verbs attracting constructions)

Construction	Verbs
como + NP	*brincar, caminar, corer, esuchar, mirar, mover, oler, parecer, salir, sentir, ser, sonar, venir, ver*
como + LOC.NP	*sentir*
como si + NP	*actuar, comportar*
como si + LOC.NP	*ser*
igual que + NP	*pensar, ser*

We will discuss the implications of these results below in Section 3.

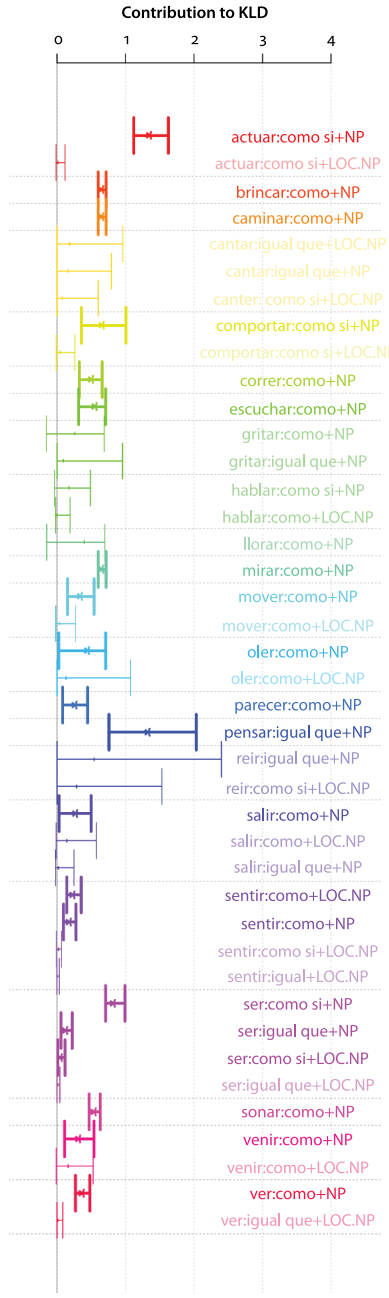

Figure 1. Verb lemmas attracting constructions

2.2.2 *From construction to verb lemma*

The application of our methods in the other direction is largely the same — the only difference is how the posterior *P* and the prior *Q* change for step 4 of the 5-step procedure sketched in Section 2.2.1. In the from-verb lemma-to-construction analysis, the posterior *P* was the distribution of each verb lemma across the constructions and the prior *Q* was the overall frequencies of the constructions in an iteration's sample. In the from-construction-to-verb lemma analysis, the posterior *P* is the distribution of each construction across the verb lemmas and the prior *Q* is the overall frequencies of the verb lemmas in an iteration's sample. Note that all other steps remain the same.

Then we plot the results for all 41 lemma-construction combinations (same highlighting and sorted within construction by collexeme strength) of with a positive median bootstrapped KLD.

Again, all these combinations involve attraction from the construction to the verb lemma, but not all are significantly different from chance; the significant co-occurrence pairs are shown in Table 4.

Table 4. Significant co-occurrences (constructions attracting verbs)

Construction	**Verbs**
como + NP	*cantar, llorar, reir*
como + LOC.NP	*sonar*
como si + NP	*cantar, mover, sentir*
como si + LOC.NP	*brincar, llorar, reir*
igual que + NP	*cantar, gritar, parecer, sonar*
igual que + LOC.NP	*brincar, mover, salir, sentir, venir, ver*

We now turn to the implications of the findings from this and the previous section.

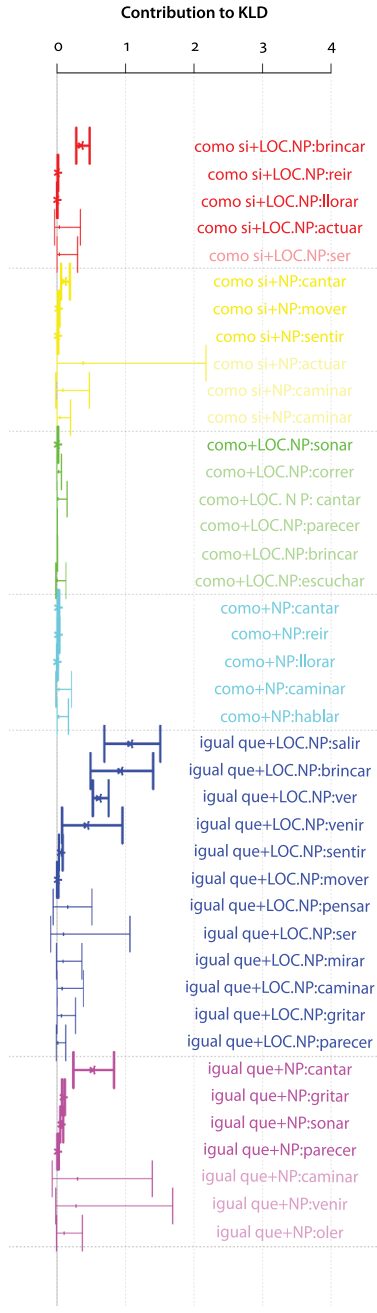

Figure 2. Constructions attracting verb lemmas

3. Discussion

The results presented in Section 2.2 allow for several observations regarding the similative-pretence alternating pair in Mexican Spanish. In this section, we discuss these results and consider their linguistic (Section 3.1) and methodological implications (Section 3.2).

3.1 Linguistic implications

As was discussed in Section 1, the semantics of the *como/igual que* 'like' constructions and *como si* 'as if' constructions are different. While the former means 'to give the same appearance as something/someone', the latter means 'to imitate, pretend, aspire to the behavior of something/someone.' Accordingly, we hypothesized that *como/igual que* 'like' constructions will prefer to occur with epistemic judgment predicates such as *parecer* 'to look.' On the other hand, we hypothesized that *como si* 'as if' constructions will prefer to occur with mistaken-identity verbs such as *actuar* 'to act.' In what follows, we explore whether these predictions hold. The discussion of our results and their linguistic implications is divided into two parts, the from-verb lemma-to-construction analysis in Section 3.1.1 and the from-construction-to-verb lemma analysis in Section 3.1.2.

3.1.1 *The from-verb lemma-to-construction analysis*

The results for this analysis are straightforward and support our initial predictions: epistemic judgment predicates (i.e., *escuchar* 'to sound', *mirar* 'to look', *ver* 'to look', *parecer* 'to seem', *sonar* 'to sound', *sentir* 'to feel', and *oler* 'to smell') significantly attract *como* 'like' constructions:

(13) *Se escucha como el lema de Star Wars*: *May the force be with you!*
'It sounds like the Star Wars' phrase: May the force be with you!"

(14) *Se mira como una misión imposible.*
'It looks like an impossible mission.'

(15) *Se ve como una punta de flecha.*
'It looks like an arrowhead.'

(16) *Parece como un pueblo fantasma.*
'It looks like a ghost town.'[5]

5. As correctly pointed out by one reviewer, the noun phrase in the example in (16) functions as a predicative element. The evidence comes from the fact that *como* 'like' can be elided without affecting the grammaticality of the construction.

Perception verbs used in *como* 'like' constructions show an epistemic function. For instance, *se siente como un buen momento* 'it feels like a good moment' is roughly the same as lit. 'it gives the same appearance as a good moment.' Other examples in which *sentir* 'to feel' has an epistemic function in this construction are the following:

(17) *Se siente como un momento decisivo.*
'It feels like an important moment.'

(18) *Se siente como el principio*
'It feels like the beginning.'

(19) *Se siente como un truco*
'It feels like a trick.'

As for *oler* 'to smell', this verb also has an epistemic function in *como* 'like' constructions in our sample. The example *me huele como a fracaso* 'it smells like a failure' is roughly the same as lit. 'it gives the same appearance as a failure.'

(20) *Me huele como a un nuevo comienzo.*
'It looks like a new beginning.'

(21) *Me huele como a una persona con exito.*
'It looks like a successful person.'

(22) *Me huele como a una nueva Aventura.*
'It looks like a new adventure.'

This use of perception verbs with an epistemic function is not surprising and has been documented in many languages around the world: perception verbs tend to have a polysemous structure, motivated by our experience and understanding of the world and metaphorical mappings. Specifically, their polysemy, as with polysemy in general, usually involves conceptual shifts across domains that are commonly characterized in terms of metaphor (Lakoff & Johnson 1980). Metaphor involves a relationship "between a source domain, the source of the literal meaning of the metaphorical expression, and a target domain, the domain of the experience actually being described by the metaphor" (Croft & Cruse 2004: 55). Put another way, metaphor consists of transposing an existing relationship into a conceptual domain by applying certain qualities from one over the other (e.g., the frequent metaphorical mappings of understanding is seeing; obeying is hearing; conserving is touching; suspecting is smelling; see Lakoff & Johnson 1980; Viberg 1984; Ibarretxe-Antuñano 1999).

Of the epistemic judgment predicates mentioned above, *sentir* 'to feel' is the only verb lemma that significantly attracts *como* 'like' constructions followed by NPs, and LOC NPs, as in the following examples:

(23) a. *Se siente como en un videojuego.*
'It feels like in a videogame.'
b. *Se siente como en la carcel.*
'It feels like in jail.'
c. *Se siente como en un estadio de futbol.*
'It feels like in a soccer stadium.'
d. *Se siente como en la playa.*
'It feels like at the beach.'

In *como* 'like' constructions followed by LOC NPs, the verb lemma *sentir* 'to feel' indicates that the experience of being in a given place is similar to that of being in another one. For instance, the example *se siente como en la playa* 'it feels like at the beach' was uttered in a context in which a speaker was in an aquarium, where he heard a water-splashing noise from the filters and pumps. It is likely that this contextual factor led him to make the following comparison: being in an aquarium is similar to being at a beach because the water-splashing noise from the filters and pumps is similar to a sound also heard in a beach. Accordingly, the physical situation surrounding the speaker may play an important role in this type of construction.

Epistemic perception verbs occurring in *igual que* 'like' constructions are rather rare in the from-verb lemma-to-construction analysis. The only verb lemma that significantly attracts *igual que* 'like' construction is *ser* 'to be':

(24) a. *El mundo es igual que Rusia.*
'The world is like Russia.'
b. *Yoda es igual que Mickey Mouse.*
'Yoda is like Mickey Mouse.'
c. *Tengo un amigo que es igual que yo.*
'I have a friend who is like me.'

The verb lemma *ser* 'to be' is polysemous in Spanish in that it can be used for describing an appearance or a behavior. In the from-verb lemma-to-construction analysis, the verb lemma *ser* 'to be' shows an epistemic function in *igual que* 'like' constructions. Based on the discourse context in which (24a) is attested, this construction should be understood as 'Yoda looks like Mickey Mouse' rather than as 'Yoda is imitating Mickey's behavior.'

Before discussing the results of the *como si* 'as if' construction, let us discuss a number of motion verbs that significantly attract *como* 'like' construction (esp. when followed by NPs):

(25) a. *Brinca como conejo. Brinca muy alto.*
'He jumps like a bunny. He jumps very high.'
b. *Camina como pato. Es muy lento.*
'He walks like a duck. He is very slow.'
c. *Corre como yo. Usa el misto tipo de camiseta.*
'He runs like me. He wears the same t-shirt.'
d. *Sale como la víctima del cotejo. Se ve asustado.*
'He starts the match like the victim. He looks scared.'

These verbs can be characterized as mistaken identity predicates like *actuar* 'to act' or *comportarse* 'to behave.' In (25a), the discourse context makes it clear that the literal sense of this example is: 'the boy is imitating the way in which bunnies jump'; same for (25b), in which a boy is imitating the way in which ducks walk. However, there are cases in which *como* 'like' constructions do not signal the meaning: 'X acts/behaves in the same way as Y', but 'X looks like Y.' In (25c), the point is not that she imitates his way of running, rather it is that she is wearing the same outfit as him. Another example is (25d), the verb *salir* 'to leave' indicates that the player looks like someone who will be defeated by the other team. In a literal sense, this example should be understood as 'the player looks like a victim.'

Second, for the *como si* 'as if' construction, the from-verb lemma-to-construction analysis also supports our initial predictions: the mistaken identity verb lemmas *actuar* 'to act' and *comportarse* 'to behave' significantly attract *como si* 'as if' constructions as in examples (26) through (30), where one or more characteristics of the behavior of 'X' may have led a speaker to say that 'X acts/behaves in the same way as Y.'

(26) *La mujer actua como si estuviera loca.*
'The woman acts as if she were crazy.'

(27) *Los jugadores del equipo actuan como si fueran campeones del mundo.*
'The players of the team act as if they were the champions of the world.'

(28) *Mi jefe se comporta como si fuera mi padre.*
'My boss behaves as if he were my father.'

(29) *Se comporta como si fuese el gobernador.*
'He behaves as if he were the governor.'

(30) *La novia de Federico se comporta como si fuera una reina.*
'Federico's girlfriend behaves as if she were a queen.'

(29) was uttered in a context in which Eladio is constantly giving orders to his colleagues, which may be similar to a governor's stereotypical behavior of constantly giving orders to their administrative staff.

It is worth noting that the verb lemma *ser* 'to be' also significantly attracts the *como si* 'as if' construction. This is the only verb that significantly occurs in two constructions: *igual que* 'like' and *como si* 'as if' constructions. Recall that this verb lemma is polysemous in Spanish in that it can be used for describing an appearance or a behavior. Interestingly, in *como si* 'as if' constructions, this verb may describe a behavior (meaning 'X acts/behaves in the same way as Y', see (31) or (32)) or an appearance (meaning 'it looks as if X were Y', see (33)).

(31) *Es como si fuera un bebé de un año que apenas balbucea.*
'It is as if he were a 1-year-old baby who babbles.'

(32) *Es como si fuera un pintor.*
'It is as if he were a painter.'

(33) *Es como si fuera un campo de arena.*
'It is as if it were a sand field.'

This co-occurrence of the epistemic verb lemma *ser* 'to be' and the construction *como si* 'as if' (33) is interesting because it is not expected given that *como si* 'as if' denotes the idea 'to imitate, pretend, aspire to the behavior of something/someone.'

To sum up, we find the hypothesized strong tendency for epistemic perception predicates to significantly attract *como* 'like' constructions, but, surprisingly, epistemic perception verbs do not significantly attract *igual que* 'like' constructions in the from-verb lemma-to-construction analysis. The only verb lemma significantly attracting the *igual que* 'like' construction is *ser* 'to be'. We also showed that motion verbs in *como* 'like' constructions (esp. when followed by NPs) may not only signal the meaning: 'X acts/behaves in the same way as Y', but also 'X looks like Y.' Finally, we do find the hypothesized strong tendency for mistaken identity verbs, such as *actuar* 'to act' and *comportarse* 'to behave' to significantly attract *como si* 'as if' constructions.

3.1.2 *The from-construction-to-verb lemma analysis*

The results of the from-construction-to-verb lemma analysis differ from those of the from-verb lemma-to-construction analysis in several ways. First, *como* 'like' constructions are not significantly attracted to epistemic perception verbs but to the following verbs:

(34) a. *Canta como los mismos ángeles.*
'She sings like the angels.'

b. *Llora como una mujer.*
'He cries like a woman.'
c. *Se ríe como un hombre.*
'She laughs like a man.'

These examples should be understood as: 'X's singing/crying/laughing sounds like Y's singing/crying/laughing.' Under this interpretation, these verbs harmonize with the semantics of *como* 'like' constructions in that they indicate that the sound produced by X when singing/crying/laughing is similar to Y.

Note that the only epistemic perception verb significantly attracted by the *como* 'like' construction is that of *sonar* 'to sound.' In particular, *como* 'like' constructions followed by a LOC NP appear with *sonar* 'to sound' like in (35). In this communicative scenario, the verb lemma *sonar* 'to sound' indicates that the experience of being in a given place is similar to being in another one.

(35) a. *Suena como en una escuela.*
'It sounds like in school.'
b. *Suena como en la selva.*
'It sounds like in the jungle.'
c. *Suena como en un manicomio.*
'It sounds like in a mental hospital.'

Second, as was shown in Section 3.1.1, *igual que* 'like' constructions are not significantly attracted by epistemic perception verbs. In the from-construction-to-verb lemma analysis, *igual que* 'like' constructions significantly attract epistemic perception verbs, as in (36).

(36) a. *Se siente igual que en aquel commercial.*
'It feels like in that commercial.'
b. *Se ve igual que en la television.*
'It looks like in t.v.'
c. *Parece igual que el LED Cinema Display.*
'It looks like the LED Cinema Display.'
d. *Suena igual que las olas del mar.*
'It sounds like the ocean waves.'

Note that *igual que* 'like' constructions followed by LOC NPs significantly attract the verb lemmas *sentir* 'to feel' and *ver* 'to look.' This suggests that speakers use this pattern for indicating that the experience of being in a given place is similar to being in another one. On the other hand, *igual que* 'like' constructions followed by NPs significantly attract the verb lemmas *parecer* 'to seem' and *sonar* 'to sound.' The meaning of these construction is that of 'X resembles Y based on a deduction from facts' (which may or may not be specified).

Como 'like' constructions and *igual que* 'like' constructions significantly attract the verb lemma *sonar* 'to sound.' However, as was mentioned above, while the lemma *sonar* 'to sound' is significantly attracted by *como* 'like' constructions followed by a LOC NP, it is significantly attracted by *igual que* 'like' constructions followed by an NP.

As was mentioned in Section 3.1.1, motion verbs significantly attracted *como* 'like' constructions followed by NPs. In the from-construction-to-verb lemma analysis, *igual que* 'like' constructions significantly attract motion verbs. In particular, they significantly attract motion verbs when they are followed by LOC NPs as in (37), (38), and (39).

(37) *Brincó igual que en las olimpiadas de Sydney 2000.*
'He jumped like in the Sydney 2000 Olympic Games.'

(38) *El equipo salió igual que en su estadio.*
'The team started the match like in their stadium.'

(39) *Ellos vinieron igual que en la fiesta de Sergio.*
'They came like in Sergio's party.'

It is worth noting that in these cases, motions verbs indicate that X performed an action in a specific place wearing the same clothes as in another place. The meaning of (37) is that of lit. 'he jumped wearing the same outfit he wore in the Sydney 2000 Olympic Games.' In a similar fashion, the meaning of (38) is that of lit. 'the team wore the same uniform they usually wear in their stadium.' Similar for (39): the literal meaning of this construction is that of 'they went to a specific place wearing the same clothes they wore when they went to Sergio's party.'

Third, we showed in Section 3.1.1 that mistaken identity verbs such as *actuar* 'to act' and *comportarse* 'to behave' significantly attract *como si* 'as if' constructions. These verbs are not significantly attracted by this construction in the from-construction-to-verb lemma analysis. Instead, the *como si* 'as if' constructions significantly attract the following motion verbs, as shown in (40) and (41), where the semantics of the motion verb lemmas harmonizes with the semantics of the *como si* 'as if' constructions in that they mean 'to imitate, pretend, aspire to the behavior of something/someone.'

(40) *Brincó con tanta fuerza como si estuviera en los juegos olimpicos.*
'He jumped as if he were in the Olympic Games.'

(41) *Se movió como si fuera una liebre cuando brincó la cerca.*
'He moved as if he were a hare when he jumped the fence.'

We showed in Section 3.1.1 that *sentir* 'to feel' has an epistemic function. The example *se siente como un buen momento* 'it feels like a good moment' is roughly

the same as lit. 'it looks like a good moment.' However, in the from-construction-to-verb lemma analysis, *sentir* 'to feel' does not have an epistemic function in *como si* 'as if' constructions. Consider (42) and (43), where the verb lemma *sentir* 'to feel' serves a function similar to *actuar* 'to act' or *comportarse* 'to behave.' Accordingly, the semantics of *como si* 'as if' and the semantics of the *sentir* 'to feel' harmonize in this communicative scenario.

(42) *El hombre se siente como si fuera el rey de la casa.*
'The man feels as if he were the king of the house.'

(43) *Mi jefe se siente como si fuera el dueño de mi tiempo.*
'My boss feels as if he were the owner of my time.'

To sum up the discussion of this section, we have shown that *como* 'like' constructions do not significantly attract epistemic perception verbs. The only epistemic perception verb significantly attracted by *como* 'like' constructions is that of *sonar* 'to sound.' Unlike *como* 'like' constructions, *igual que* 'like' constructions show the opposite picture in that they significantly attract epistemic perception verbs. Furthermore, *igual que* 'like' constructions significantly attract motion verbs. In this scenario, they indicate that X performed an action in a specific place wearing the same clothes as in another place. As for *como si* 'as if' constructions, it was shown that, these constructions significantly attract the verb lemma *sentir* 'to feel.' In this scenario, this verb lemma does not have an epistemic function. Instead, it serves a function similar to *actuar* 'to act' or *comportarse* 'to behave.'

3.2 Methodological implications

This paper has shown that the collostructional method adopted here is better suited for explanatory purposes than the current standard that is nearly always used by researchers. Specifically, we think this is so in three different respects.

First, the present methodology distinguishes the attraction that a verb lemma exerts on a construction from the attraction that a construction exerts on a verb lemma. This is more than just an academic distinction because (i) we have seen that the results are different if one considers both directions of association separately and (ii) more generally, it means that the AM used is more sensitive to something that we *know* speakers are generally sensitive to: given that the usage-based paradigm holds that most of speakers' linguistic abilities is learned/acquired (rather than innate) and is learned/acquired using domain general (learning) mechanisms, ignoring the direction of learning and forming associations, as adopting a bidirectional measure implies, cannot possibly be a good idea

once one goes beyond initial exploratory studies and is something that usage-based linguists would probably not recommend in any other setting.

The second advantage of the present approach is that it uses an association measure (the KLD) which is much less correlated with the observed co-occurrence frequencies than other AMs (e.g., AMs based on significance values). This is so important because it means it is better at isolating and interpreting association effects separately from mere co-occurrence frequency.

Finally, the present approach uses a bootstrapping approach (based on files/speakers) to assess the uncertainty of the results. This is useful because, if one does not quantify the uncertainty coming with one's results or if one quantifies it using parametric confidence intervals, one is likely to overinterpret one's (ranking) results because one either has no empirical basis for deciding which differences between associations are noteworthy/significant or one has an empirical basis, but one that is anti-conservative (given how parametric confidence intervals will be too narrow and, thus, invite overgeneralizations).

Given these advantages, we hope that the current methodology will be more widely applied in future collostructional studies — using directional measures with bootstrapped confidence intervals cannot make the results any worse than the current default approach of bidirectional point estimate only, but it can make them a lot more instructive and insightful for both anyone's current study and whatever work builds on it later.

4. Concluding remarks

The distinction between lexicon and syntax has played an important role in formal linguistic theory. However, this dichotomy, and its presumed independence, has long been challenged in particular by work in the framework of Construction Grammar. Our work could be seen as another contribution to the idea that there are associations between individual lexemes and particular slots of constructions, and that such connections are semantically motivated. However, the present investigation has gone beyond this traditional goal in two respects.

First, here we have explored the distribution of the filler-slot types of an alternating pair traditionally neglected in the corpus-based literature, the similative-pretence alternating pair. This is an important contribution to the usage-based approach in that the expression of comparison is one of the most basic cognitive processes common to all human beings. It is expected that our synchronic results will be valuable to those interested in analyzing other domains, such as the acquisitional paths of the similative-pretence alternating pair, and the changing interrelations of the similative-pretence alternating pair and their lexical collocates from a historical perspective (i.e., corpus-linguistic approach to historical semantics).

In our study, the from-verb lemma-to-construction analysis supported our hypothesized strong tendency for epistemic perception predicates to appear in similative constructions. In this analysis, there was a significant attraction of epistemic lemmas to occur in *como* 'like' constructions. This analysis also supported our hypothesized strong tendency for mistaken identity verbs, such as *actuar* 'to act' and *comportarse* 'to behave' to appear in *como si* 'as if' constructions.

As for the from-construction-to-verb lemma analysis, it also supported our hypothesized strong tendency for similative constructions to significantly attract epistemic perception predicates. In particular, *igual que* 'like' constructions significantly attracted epistemic perception verbs. As was mentioned above, the mistaken identity verbs *actuar* 'to act' and *comportarse* 'to behave' significantly attracted *como si* 'as if' constructions in the from-verb lemma-to-construction analysis. However, in the from-construction-to-verb lemma analysis, *como si* 'as if' constructions significantly attracted the verb lemma *sentir* 'to feel.' In this scenario, this verb lemma served a function similar to *actuar* 'to act' or *comportarse* 'to behave.' What these results indicate is that this directional AM enables us to uncover co-occurrences that a traditional bidirectional AM method does not, which opens up new avenues of research for the future.

References

Batchelor, R. E. & Pountain, C. J. (2005). *Using Spanish: A guide to contemporary usage*. Cambridge University Press.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*, 1–26.

Chamoreau, C. (2017). Multifunctionality and polysemy of the similative marker =*kan* in Pesh. In Y. Treis & M. Vanhove (Eds.), *Similative and equative constructions: A cross-linguistic perspective* (pp. 321–337). John Benjamins.

Cover, T. & Thomas, J. A. (2006). *Elements of information theory*. 2nd ed. Wiley.

Croft, W. & Cruse, A. (2004). *Cognitive linguistics*. Cambridge University Press.

Darmon, C. (2017). The morpheme (a)ŋa in Xamtanga: Functions and grammaticalization targets. In Y. Treis & M. Vanhove (Eds.), *Similative and equative constructions: A cross-linguistic perspective* (pp. 359–385). John Benjamins.

Diessel, H. (2019). *The grammar network. How linguistic structure is shaped by language use*. Cambridge University Press.

Diessel, H. (2020). A dynamic network approach to the study of syntax. *Frontiers in Psychology*, *11*, Article 604853.

Egbert, J. & Plonsky, L. (2020). Bootstrapping techniques. In M. Paquot & S. T. Gries (Eds.), *Practical handbook of corpus linguistics* (pp. 593–610). Springer.

Evert, S. (2009). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook*, Vol. 2 (pp. 1212–1248). Mouton De Gruyter.

Fillmore, C. J. & Kay, P. (1999). *Construction Grammar*. University of California.

Fortescue, M. (2010). Similitude: A conceptual category. *Acta Linguistica Hafniensia*, *42*, 117–142.

Fuchs, C. (2014). *La comparaison et son expression en francais*. Ophrys.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar approach to argument structure*. The University of Chicago Press.

Gries, S. T. (2005). Null-hypothesis significance testing of word frequencies: A follow-up on Kilgarriff. *Corpus Linguistics and Linguistic Theory 1*, 277–294.

Gries, S. T. (2019). 15 years of collostructions: Some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). *International Journal of Corpus Linguistics*, *24*, 385–412.

Gries, S. T. (2021). *Statistics for linguistics with R*. 3rd ed. Mouton de Gruyter.

Gries, Stefan T. (2022a). What do (some of) our association measures measure (most)? Association? *Journal of Second Language Studies*, *5*, 1–33.

Gries, Stefan T. (2022b). Towards more careful corpus statistics: Uncertainty estimates for frequencies, dispersions, association measures, and more. *Research Methods in Applied Linguistics*, *1*, 100002.

Gries, S. T. (2023). Overhauling collostructional analysis: Towards more descriptive simplicity and more explanatory adequacy. *Cognitive Semantics*. Online Publication Date: 07 Aug 2023.

Gries, S. T. & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspectives on 'alternations'. *International Journal of Corpus Linguistics*, *9*, 97–129.

Harris, Z. (1954). Distributional structure. *Word*, *10*, 146–162.

Haspelmath, M. & Buchholz, O. (1998). Equative and similative constructions in the languages of Europe. In J. Van der Auwera (Ed.), *Adverbial constructions in the languages of Europe* (pp. 277–334). Mouton de Gruyter.

Henkelmann, P. (2006). Constructions of equative comparison. *STUF — Sprachtypologie und Universalienforschung*, *59*, 370–398.

Hetterle, K. (2015). *Adverbial clauses in cross-linguistic perspective*. Mouton de Gruyter.

Hilpert, M. (2013). *Constructional change in English: Developments in allomorphy, word-formation, and syntax*. Cambridge University Press.

Hilpert, M. (2021). *Ten lectures on Diachronic Construction Grammar*. Brill.

Ibarretxe-Antuñano, I. (1999). Vision metaphors for the intellect: Are they really cross-linguistic? *Atlantis*, *30*, 15–33.

Jiménez Juliá, T. (2003). *Como* en español actual. *Verba*, *30*, 117–161.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. The University of Chicago Press.

Olguín Martínez, J. (2021). Hypothetical manner constructions in world-wide perspective. *Linguistic Typology at the Crossroads*, *1*, 2–33.

Palmer, F. R. (2001). *Mood and modality*. 2nd ed. Cambridge University Press.

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Schulze, W. (2017). Toward a cognitive typology of *like*-expressions. In Y. Treis & M. Vanhove (Eds.), *Similative and equative constructions: A cross-linguistic perspective* (pp. 33–78). John Benjamins.

Stefanowitsch, A. & Gries, S. T. (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, *8*, 209–243.

Trujillo, R. (1990). Sobre la explicación de algunas construcciones de *como*. *Verba*, *17*, 249–266.

Viberg, Å. (1984). The verbs of perception: A typological study. In B. Butterworth, B. Comrie, & Ö. Dahl (eds), *Explanations for language universals* (pp. 123–162). Mouton de Gruyter.

Wulff, S. (2008). *Rethinking idiomaticity: A usage-based approach*. Continuum.

Address for correspondence

Jesus Olguin Martinez
Department of English
Illinois State University
Normal, IL 61790-4240
United States of America

jfolgui@ilstu.edu

olguinmartinez@ucsb.edu

Co-author information

Stefan Th. Gries
Department of Linguistics
University of California, Santa Barbara

stgries@linguistics.ucsb.edu

Publication history

Date received: 3 May 2023
Date accepted: 30 October 2023
Published online: 28 February 2025