# Filler-slot relations in language contact

## Lexico-syntactic transference from a usage-based perspective

Jesús Olguín Martínez[1] and Stefan Th. Gries[2,3]
[1] University of Hong Kong | [2] UC Santa Barbara | [3] JLU Giessen

The present study investigates the influence of Mexican Spanish similative (e.g., he swims like a fish) and pretence constructions (e.g., he swims as if he were a fish) on those found in four Mesoamerican languages: Huasteca Nahuatl, Papantla Totonac, San Gabriel Huastec, and Uxpanapa Chinantec. Using predictive modeling, we demonstrate that these indigenous languages have not only borrowed the markers *komo* 'like' and *komo si* 'as if' from Mexican Spanish, but have also adopted the lexical preferences (e.g., verb lemmas) associated with these constructions. However, we also identify a number of syntactic differences in how locative and non-locative NPs are treated within similative and pretence constructions in these languages. These findings suggest that, in language contact scenarios, constructions are rarely replicated intact from one language to another. Furthermore, our analysis reveals that while the similative and pretence markers themselves are outcomes of matter replication, the verb lemmas in these constructions result from pattern replication.

**Keywords:** similative, pretence, language contact, filler-slot relations, predictive modeling.

## 1. Introduction

It is well-known that the co-occurrence patterning of lexemes and constructions is functionally motivated (Goldberg, 1995, p. 50; Gries & Stefanowitsch, 2004, p. 99), which gives rise to a joint distribution of lexemes with constructions that are known in the literature as *filler-slot relations* (see Kay & Fillmore, 1999; Hilpert, 2013; Diessel, 2019, 2020). Such probabilistic associations in synchronic data have often been studied using methods from the family of collostructional analysis. This family of methods is based on the distributional hypothesis: "If we consider

https://doi.org/10.1075/rcl.00250.olg | Published online: 20 January 2026
*Review of Cognitive Linguistics* ISSN 1877-9751 | E-ISSN 1877-976X 

words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C" (Harris, 1970, p. 785). In other words, frequency of co-occurrence reflects, and is thus a diagnostic of, similarity of meaning and/or function.[1] For instance, it has been shown that ditransitive constructions (e.g., she gave John a cake) and prepositional dative constructions (e.g., she gave a cake to John) are semantically and pragmatically related, but they have somewhat different senses or meaning preferences. This has been supported by the fact that ditransitive constructions attract verb lemmas, such as give, tell, show, offer, cost, teach, wish, ask, promise, deny, and prepositional dative constructions attract verbs lemmas, such as bring, play, take, pass, make, sell, do, supply, read, hand, and so forth (Stefanowitsch & Gries, 2003). In diachronic research, filler-slot relations have also been the focus of attention in a number of studies. In particular, linguists have used *Diachronic Collostructional Analysis* (Gries & Hilpert, 2008). This is a method that specifically focuses on how certain words (lexical items) become more or less strongly associated with particular constructions (grammatical patterns) across historical periods. For instance, the most frequent verb lemmas of a given construction in the 18th century will be different from the most frequent ones in the 19th and 20th century (Hilpert, 2006, 2008). In Usage-Based Construction Grammar (Usage-Based CxG), such probabilistic associations constitute part of each language user's individual and ever-changing exemplar-based representation of linguistic knowledge (Beckner et al., 2009).

While filler-slot relations have been explored from both a synchronic and diachronic perspective, the analysis of this domain from a language contact perspective is still in its infancy (but see Wilson, 2013; Béchet, 2020; Bullock et al., 2021; Wiesinger, 2021).

The present study aims at contributing to fill this gap by exploring filler-slot relations in four Mesoamerican languages spoken in the same area: Huasteca Nahuatl (HuaNah), Papantla Totonac (PapTot), San Gabriel Huastec (SanGabHua), and Uxpanapa Chinantec (UxpChin). In particular, special attention is paid to the role of Mexican Spanish (MexSpa) in shaping filler-slot relations in these Mesoamerican languages.

---

1. This family of methods 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).

These languages express similative and pretence meanings with the MexSpa borrowed similative and pretence markers, as in the following examples:

(1) *HuaNah (Uto-Aztecan)*
a. *hual-motlalo-k* *komo* *kuatochi.*
DIR-3SG.SBJ.run-PFV like bunny
'He ran like a bunny.' (The bunny and the turtle story-07/15/2022)
b. *hual-motlalo-k* *komo* *si* *el-s* *kuatochi.*
DIR-3SG.SBJ.run-PFV like if 3SG.SBJ.be-IRR bunny
'He ran as if he were a bunny.' (The bunny and the turtle story-07/15/2022)

(2) *PapTot (Totonacan)*
a. *k-a'ːwan* *komo* *ja'í* *chichí.*
1SG-walk like DEF dog
'I walk like the dog.' (The crazy guy-08/14/2023)
b. *k-a'ːwan* *komo* *si* *wan-níːt* *mistuːn.*
1SG-walk like if be-PFV cat
'I walk as if I were a cat.' (The crazy guy-08/14/2023)

(3) *SanGabHua (Mayan)*
a. *na* *Hwaːn* *ʔaːθ-i-l* *komo* *ʔan* *bitsim.*
DEF Juan 3SG.SBJ.run-INACC-INCOMPL like DEF horse
'Juan is running like the horse.' (Our last vacations-08/15/2017)
b. *t'oh-n-al* *komo* *si* *wenk'-ow-al*
3SG.SBJ.work-MIDDL-INCOMPL like if 3SG.SBJ.become-TRANS-INCOMPL
*ʔoːbeh.*
lazy.guy
'He (my cousin) works as if he were a lazy guy.'
(Things that happened last year-08/19/2017)

(4) *UxpChin (Oto-Manguean)*
a. *ii* *komo* *lafa' i* *nU̱U̱.*
3SG.sound like IRR DEF.ANIM star
'It sounds like the star.' (My grandfather-07/16/2018)
b. *ca-cuíi'=b* *komo* *si* *cofa' coo'* *ji* *cuɨ'ñeá.*
COMPL-3SG.run=EMPH like if IRR INDEF.INAN light thunder
'He (my grandfather) ran as if he were a thunder.'
(My grandfather-07/16/2018)

Accordingly, the question is: have speakers of these indigenous languages also copied the lexical preferences of the first slot (verb lemmas) of MexSpa similative-pretence constructions? Note that similative and pretence markers are orthographically represented as *komo* 'like' and *komo si* 'as if' in HuaNah, PapTot,

SanGabHua, and UxpChin. On the other hand, these markers are orthographically represented as *como* 'like' and *como si* 'as if' in MexSpa.

Previous studies have shown that the first slot of MexSpa similative constructions (ending with non-locative NPs and locative NPs), as in (5a)–(b), prefers to occur with epistemic judgment predicates, such as *parecer* 'to seem', *mirar* 'to look', *ver* 'to look', and *sonar* 'to sound', among others (Olguín Martínez & Gries, 2025a). In these patterns, the concept of likeness is fully inferential (Trujillo, 1990) and may be derived metonymically or metaphorically in that "they represent fossilized patterns of cognitive processes conventionalized over times" (Schulze, 2017, p.36). On the other hand, the first slot of MexSpa pretence constructions (ending with non-locative NPs and locative NPs), as in (6a)–(b)), prefers to appear with mistaken identity verbs, such as *actuar* 'to act' and *comportar* 'to behave' (Olguín Martínez & Gries, 2025a). Pretence constructions indicate 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, p.372). These constructions are similar to similatives in that the concept of likeness is fully inferential.

Table 1 summarizes the results reported by Olguín Martínez and Gries (2025a).

(5) *MexSpa (Indo-European)*
  a. *como* 'like' construction with NP
  *se comporta como un tonto.*
  PRON 3SG.ACT.PRS like INDEF fool
  'He acts like a fool.' (252 16-05-23 MX Economíahoy.mx)
  b. *como* 'like' construction with LOC NP
  *se siente como en su casa.*
  PRON 3SG.feel.PRS like LOC 3SG.POSS house
  'It feels like at his house.' (2100 18-06-11 MX Digital Trends Español)

(6) a. *como si* 'as if' construction with NP
  *se comporta como si fuera un tonto.*
  PRON 3SG.ACT.PRS as if 3SG.be.SUBJ INDEF fool
  'He acts as if he were a fool.'
  (2400 18-02-18 MX El Mercurio de Tamaulipas)
  b. *como si* 'as if' construction with LOC NP
  *se siente como si estuvieramos en su casa.*
  PRON 3SG.feel.PRS as if 1PL.be.SUBJ LOC 3SG.POSS house
  'It feels as if we were at his house.' (3099 17-05-20 MX nnc.m)

**Table 1.** Verb lemma types occurring in the first slot of similative and pretence constructions in MexSpa (summarizing Olguín Martínez & Gries, 2025a, p. 80–88)

| Construction type | Construction | Type of verb lemmas |
|---|---|---|
| Similative 'like' (non-locative) | *como* + NP | Epistemic |
| Similative 'like' (locative) | *como* + LOC.NP | Epistemic |
| Pretence 'as if' (non-locative) | *como si* + NP | Mistaken identity |
| Pretence 'as if' (locative) | *como si* + LOC.NP | Mistaken identity |

Previous research has shown that HuaNah, PapTot, SanGabHua, and UxpChin have not only borrowed grammatical markers from MexSpa, but also other constructional properties in which these markers are attested (Olguín Martínez, 2022, 2023, 2024a, b). Accordingly, it seems reasonable to assume that lexical properties of MexSpa similative and pretence constructions be transferred to these Mesoamerican through language contact. MexSpa has established a strong presence in the area, where HuaNah, PapTot, SanGabHua, and UxpChin and other indigenous languages are spoken (Dexter-Sobkowiak, 2022). Indigenous peoples in the area are often bilingual, speaking both their native languages and MexSpa. As the official language of Mexico, MexSpa is used in compulsory education, government at all levels, health services, media, and many other domains. As a result, virtually everyone is exposed to both spoken and written MexSpa (Dexter-Sobkowiak, 2022, p. 2). The following are the hypotheses of the present study:

- Hypothesis 1: Speakers of HuaNah, PapTot, SanGabHua, and UxpChin have not only borrowed the similative marker from MexSpa, but also the same lexical preferences of the first slot of these constructions, i.e., epistemic verbs.
- Hypothesis 2: Speakers of HuaNah, PapTot, SanGabHua, and UxpChin have not only borrowed the pretence marker from MexSpa, but also the same lexical preferences of the first slot of these constructions, i.e., mistaken identity verbs.

From a theoretical perspective, we adopt a Usage-Based CxG approach to language contact, assuming that language contact phenomena can happen on every level (e.g., Boas & Höder, 2018, p. 10) and that in contact situations, structural elements at various levels can be transferred from one language to another (Clyne, 2003).[2] This perspective challenges the notion of a strict division of language into qualitatively distinct and modular components (e.g., lexicon, syntax, and mor-

**2.** See also Thomason and Kaufman (1988) and Heine and Kuteva (2005).

phology) and instead supports an integrated approach that considers both formal and functional aspects of language, as well as varying degrees of structural schematicity of constructions, in the analysis of language contact. From a methodological perspective, we use predictive modeling to determine which factors influence the choice of similative and pretence markers in the different indigenous languages considered here. We use the term *donor language* to refer to MexSpa in that it served as the source of diffusion of X, and we use the term *recipient language* to refer to HuaNah, PapTot, SanGabHua, and UxpChin in that they borrowed X from a donor language.

The structure of this study is as follows. Section 2 introduces the corpus data and outlines the methodological approach used here to analyze similative-pretence constructions in the languages under investigation. Section 3 presents a detailed discussion of the results. In Section 4, we argue that the findings carry implications for the field of contact linguistics. Finally, Section 5 provides a summary of the results of the present research.

## 2. Methods and results

This section introduces the corpus data, describes the method used to compare similative-pretence constructions in MexSpa, HuaNah, PapTot, SanGabHua, and UxpChin, and presents the results.

### 2.1 Corpus data, data extraction, and annotation

For the investigation of similative-pretence constructions in MexSpa, we used the *Corpus del Español NOW* (News on the Web), which was the most suitable resource available to us since it includes data from 21 Spanish-speaking countries rather than exclusively MexSpa and since it was the only corpus available from which we could obtain data on both similative and pretence constructions.[3] Other corpora we considered, such as the *TEDx Spanish Corpus* (Hernandez-Mena, 2019) and the *Corpus del Español Mexicano Contemporáneo* (Lara et al., 2018), primarily featured similative constructions, but not pretence constructions. While the *Corpus del Español NOW* differs in genre from the indigenous corpora, news

3. By MexSpa, we refer to the Mexican national variety of the Spanish language spoken throughout Mexico. We recognize that speech in different regions of the country may display distinctive local features at various linguistic levels (see Smirnova et al., 2023). Nevertheless, we have chosen to use this term, as the *Corpus del Español NOW* does not provide information on dialectal variation within Mexican Spanish.

texts are widely accepted in corpus linguistics as reasonable proxies for broader usage, especially when alternatives are limited. At present, we have no reason to believe that the genre difference would significantly distort our analysis, let alone do so in a systematic way that unfairly skews the results. In fact, when working with under-resourced languages or historical corpora, genre mismatches are a frequent and often unavoidable aspect of linguistic research.

We conducted an exhaustive retrieval of MexSpa similative and pretence constructions from the *Corpus del Español NOW*. Specifically, we searched for the forms *como* 'like' and *como si* 'as if', which yielded a large data set in which these expressions were followed by NPs, locative NPs, and clauses (e.g., *ella actuó como si no hubiera pasado nada* 'she acted as if nothing had happened'). Given that the corpus data for HuaNah, PapTot, SanGabHua, and UxpChin predominantly feature similative and pretence constructions involving locative and non-locative NPs, we chose to focus our analysis on these patterns. Consequently, the MexSpa data were trimmed down to exclude instances where *como* 'like' and *como si* 'as if' were followed by clauses. The resulting MexSpa data set includes 323 instances of *como* 'like' and 25 instances of *como si* 'as if' constructions involving locative and non-locative NPs, as illustrated in the examples in (5a)–(b)) and (6a)–(b)).

For the 348 examples, we then manually coded each of the constructions for the relevant variables for our analysis: (i) verbs that can occur in the first slot of *como* 'like' and *como si* 'as if' constructions, (ii) whether the NP following *como* 'like' and *como si* 'as if' was locative or non-locative, and (iii) the construction schema.

For the analysis of similative-pretence constructions in HuaNah, PapTot, SanGabHua, and UxpChin, we drew on four corpora based on fieldwork conducted by the first author of this study. The HuaNah corpus is based on fieldwork carried out in the village of Teposteco, located in the municipality of Chicontepec in the state of Veracruz. Teposteco has a population of 363 inhabitants, and MexSpa is the primary language of instruction at all educational levels (Eladio Cruz, pers. comm.). The corpus comprises 32 narratives produced by three adult native speakers: Mrs. Duarte, Mr. Rodriguez, and Mr. Cruz. These texts fall into three main categories: fairy tales, personal narratives, and procedural texts. The fairy tales in the corpus explore everyday human experience (e.g., ambition, poverty, hunger, honesty, companionship, love, faith, anger, revenge, sexuality, and cunning, among others). Human characters typically occupy central roles, though they may transform into spiritual or legendary beings and animals. In some fairy tales, animals and plants are personified, imbued with human traits and moral qualities by the Nahua speakers. The personal narratives consist of short accounts in which the speakers recount both positive and negative life experiences. For example, Mr. Cruz reflects on the loss of loved ones, describing the

deep sadness and hopelessness that followed these events. The third category, procedural texts, includes step-by-step explanations of how to perform specific tasks or make certain objects, as described by the speakers. In total, the HuaNah corpus contains 1,032 sentences. Among these, we identified 139 instances of *komo* 'like' constructions and 144 instances of *komo si* 'as if' constructions.

The PapTot corpus is based on fieldwork conducted in *El Remolino*, a town located in the municipality of Papantla, Veracruz. The community has approximately 1,200 inhabitants and is informally divided into neighborhoods. *El Remolino* is primarily a Totonac community, though it also includes mestizo residents. Most of the elderly population is fluent in both Totonac and MexSpa. Notably, they are among the few who still speak Totonac and wear traditional attire. The corpus includes 39 personal narratives told by two native speakers: Mr. Hernandez and Mr. Quintero. These stories recount a range of positive and negative life experiences. In total, the narratives comprise 1,143 sentences, including 39 instances of *komo* 'like' constructions and 14 instances of *komo si* 'as if' constructions.

The SanGabHua corpus is based on fieldwork conducted in *El Mamey San Gabriel*, a community in the municipality of Tantoyuca, Veracruz. The community has approximately 200 inhabitants, and MexSpa serves as the primary language of instruction across all educational levels. The corpus consists of 45 personal narratives collected from two native speakers: Mr. Andrade and Mr. Guzman. These narratives exclusively recount a variety of positive and negative personal experiences. In total, the data set contains 1,021 sentences, including 28 instances of *komo* 'like' constructions and 22 instances of *komo si* 'as if' constructions.

The UxpChin corpus is based on fieldwork conducted in Uxpanapa, Veracruz. This community has approximately 1,589 inhabitants, and MexSpa serves as the primary language of instruction at all educational levels. The corpus comprises 36 personal narratives collected from two native speakers: Mr. Sierra and Mr. Lopez. These narratives exclusively recount a range of positive and negative personal experiences. In total, the data set includes 1,021 sentences, with 22 instances of *komo* 'like' constructions and 18 instances of *komo si* 'as if' constructions.

For each of the HuaNah, PapTot, SanGabHua, and UxpChin similative-pretence constructions, we then coded the same variables as for the MexSpa data. Table 2 illustrates the organization of our data with the help of an exemplary excerpt.

Mention should be made of the following issue: HuaNah, PapTot, SanGabHua, and UxpChin do not seem to contain native similative-pretence constructions. Accordingly, the absence of explicit ways of expressing similative-pretence meanings in these indigenous languages may have provided a niche for the newly

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

| Language | Source | Example | Lemma | Lemma with translation | Locative | Construction schema |
|---|---|---|---|---|---|---|
| MexSpa | 95 19-03-26 MX Milenio.com | *corre como ratero* | *correr* | *correr* 'to run' | no | *como* |
| MexSpa | 41 19-07-01 MX Pulso Diario de San Lui (1) | *se sintió como en casa* | *sentir* | *sentir* 'to feel' | yes | *como* |
| MexSpa | 700 15-06-11 MX Vanidades | *actúa como si fuera la mejor amiga de Rumer* | *actuar* | *actuar* 'to act' | no | *como si* |
| HuaNah | The bunny and the turtle | *hualmotlalok komo si els kuatochi* | *motlalo* | *motlalo* 'to run' | no | *komo si* |
| PapTot | The crazy guy | *ka'ːwan komo ja'ʼí chichí* | *a'ːwan* | *a'ːwan* 'to walk' | no | *komo* |
| SanGabHua | Things that happened last year | *t'ohnal komo si wenk'owal ʔoːbeh* | *t'oh* | *t'oh* 'to work' | no | *komo si* |
| UxpChin | My grandfather | *ii komo lafa' i nU̱U̱.* | *ii* | *ii* 'to sound' | no | *komo* |

interpreted markers to fill (see Mithun, 1992, p.126 for similar observations in Native American languages).[4]

**4.** SanGabHua contains the native similative marker *hajk'i* 'like'. However, it is almost non-existent in the corpus used in the present study (i.e., it only occurs three times in the data). HuaNah may indicate similative meanings with the native construction in (i), which should be understood as: lit. 'it feels like being in this place reaches the same feeling as that of being in his house.' (Olguín Martinez & Gries, 2025b). In this construction, the verb *temanti* 'to reach' functions in a similar way as *komo* 'like'. This native construction has a low frequency in the corpus used in the present study (i.e., it only occurs three times in the data and only with locative NPs).

(i) *ki-machi-k ki-temanti-s pa i-cha.*
3SG.OBJ-feel-PFV 3SG.OBJ-reach-IRR LOC 3SG.POSS-house
'It feels like (we were) at his house.' (The instruments story-07/15/2022)

## 2.2 The statistical analysis and results

To determine which factors are correlated with the response variable construction schema, or, to use more causal language, which factors influence the choice of similative and pretence markers in the different languages, we use predictive modeling, an approach that has become more and more common in especially cognitive-linguistic or usage-based studies of corpus data. Given the characteristics of our data — relatively few data points, a not-tiny number of predictor levels, repeated measurements, and Zipfian distributions (of especially the lemmas) — the maybe most common approach of generalized linear mixed-effects modeling was out of the question. Instead, we used the method of conditional inference forests (see Strobl et al., 2009; Gries, 2021: Sections 7.2–7.3), which is based on the simple logic of classification (and regression) trees, but extends it in a variety of ways. As the name of the method suggests, conditional inference forests, or the more general class of random forests, do not just use one tree but ntree hundreds or thousands of different trees, but also introduce two layers of randomness to the process:

- on the level of the data set: each tree is fit on a different, sampled with replacement, random sample of the complete data. This means that, on average, each sampled data set includes only approximately 63.2% of the original *n* data points of the original sample, meaning each sample for each tree is a 'slightly different version of the original, actual data.'
- on the level of the predictors involved in the trees: at each split in each tree, not all predictors are available to be chosen for a split. Instead, at each split in each tree, the algorithm is only allowed to choose one of the mtry predictors, where mtry is often *p* (the total number of predictors) divided by 3 or exponentiated to the power of 0.5.

These ways to introduce randomization into the algorithm have attractive consequences. First, the fact that data points are randomized helps alleviate the effects of multicollinearity and repeated measurements. Second, that together with the fact that predictors are randomly suppressed 'gives weaker predictors a say' and decorrelates the resulting trees. Third, all of these things make random/conditional inference forests a method that is applicable in small-*n* large-*p* contexts, i.e., if one has a larger number of predictors that seems high given a smaller number of data points, a kind of scenario that regression approaches struggle with.

We began our analysis by computing the no-information rate/baseline, which was an already high 70.9% (the frequency of similative markers in the data); the null deviance of this response variable was accordingly 937.628. We then fit a conditional inference forest (see Hothorn & Zeileis, 2015) to the data with

the following hyperparameter settings: ntree = 1500, mtry = 2, and sampling with replacement. The forest resulted in good predictive accuracy as measured by both its confusion matrix and some other widely-used performance statistics.

**Table 3.** Confusion matrix of observed vs. predicted *komo si/como si* 'as if' vs. *komo/como* 'like' choices

| **Predicted**<br>**Observed** | ***komo si/como si* 'as if'** | ***komo/como 'like'*** | **Sum** |
|---|---|---|---|
| *komo si/como si* 'as if' | 196 | 30 | 226 |
| *komo/como* 'like' | 34 | 518 | 552 |
| Sum | 230 | 548 | 778 |

The accuracy of the forest amounts to 91.8%, which is significantly better than the baseline ($p_{\text{one-tailed binomial test}}<10^{-45}$) and it comes with an excellent *C*-score of 0.96 and a Cohen's $\kappa$ of 0.8. Variable importance scores indicate that the predictor lemma (1.88) was most important — pointing to a high degree of lexically-determined specificity — followed, by some 'distance', by language (0.73) — indicating that there are notable differences between the languages — but then the effect of locative is too small to be notable (0.27). However, in order to also determine how especially lemma and language are correlated with construction schema, we computed partial dependence scores (on the predicted probability scale, see Greenwell 2017): (i) to see which of their levels prefer *komo/como* 'like' and which prefer *komo si/como si* 'as if' but also (ii) to determine whether the predictors are most impactful as what in a regression context would be called main effects or as interactions.

The equivalent of the main effect of lemma is shown in Figure 1: The *y*-axis represents the partial dependence score (as a predicted probability) for *komo/como* 'like' for each lemma shown in the *x*-axis; the lemmas are sorted in increasing order of preference for *komo/como* 'like' from the left (where the predicted probabilities of *komo/como* 'like' are very low to low) to the right (where the predicted probabilities of *komo/como* 'like' are high to very high) and shown with a bar width proportional to the lemma's frequency in the data. The horizontal dashed line at around $y = 0.709$ represents the baseline proportion of occurrence of *komo/como* 'like', meaning (i) verbs like *actuar* 'act', whose bars end below that line, are predicted less strongly than baseline, (ii) verbs like *oler* 'smell', whose bars end above that line, are predicted less strongly than baseline, and (iii) verbs like *correr* 'run' have no strong preference.

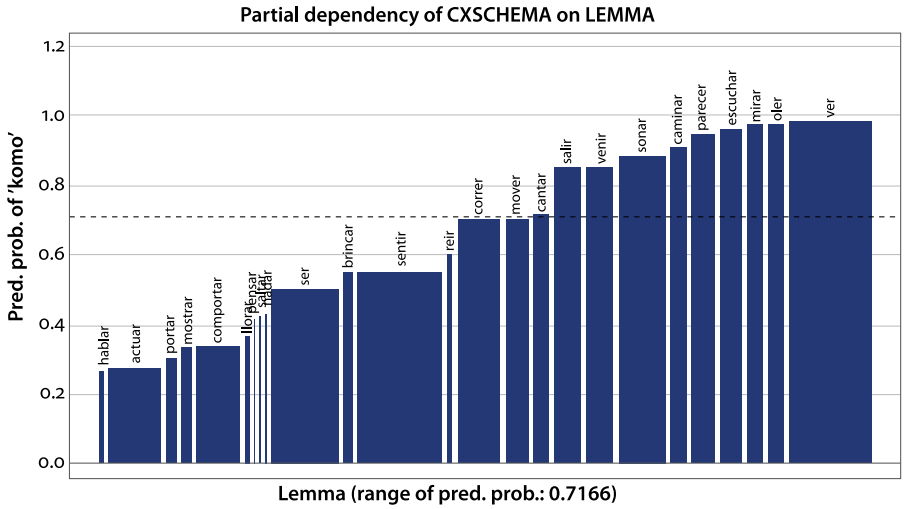


**Figure 1.** Partial dependence scores for lemma

The effect for language does not require visualization because it is very clear and straightforward: MexSpa is the only language that overall prefers *como* 'like', all others prefer *komo si* 'as if'. While many analyses with random forests stop at this point — meaning, with analyses of essentially main effects of predictors — the more interesting aspect of the results especially for an analysis in terms of language contact is of course to see whether the lemma preferences vary across the languages, and the results show they do. Consider Figure 2 for what is essentially the interaction of lemma and language. Given the large number of predicted probabilities — 25 lemmas times 5 languages — we represent the results summarily such that:

- the *y*-axis again represents the predicted probability of *komo/como* 'like' (for each of the 125 combinations);
- for each verb lemma, a line connects 5 points — one for each language in alphabetical order (also repeated in the sub-title abbreviated to four characters) — and the times symbol presents the overall lemma preference.

Several observations are immediately obvious. First, for the majority of lemmas, the second dot, i,e., the one representing MexSpa, has the by far highest score for *como* 'like', which is compatible with the main effect of language reported above. Second, however, there are two classes of verb lemmas: One consists of verbs where the line connecting the five points for the languages is consistently below the baseline and thus preferring *komo si/como si* 'as if' (as with, e.g., *actuar* 'act' and *hablar* 'speak') or consistently above the baseline and thus preferring

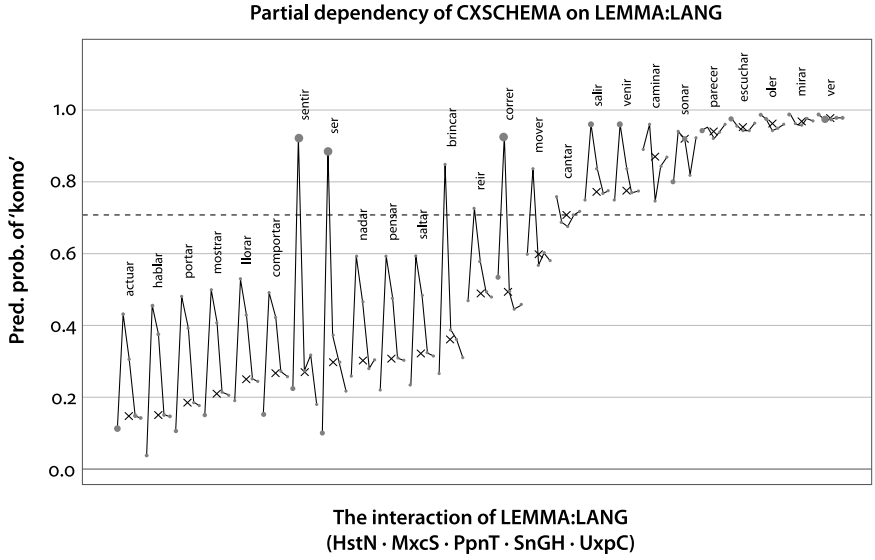


**Figure 2.** Partial dependence scores for lemma: Language

*komo/como* 'like' (as with *salir* 'leave' and *venir* 'come'). As for the former, the verb lemmas *actuar* 'act', *hablar* 'speak', *portar* 'behave', *mostrar* 'show', *llorar* 'cry', *comportar* 'behave', *nadar* 'swim', *pensar* 'think', and *saltar* 'jump' prefer *komo si/como si* 'as if' in all languages, which is interesting because most of them are mistaken identity verbs and align with the semantics of pretence constructions (see Section 3). As for the latter, the verb lemmas *salir* 'leave', *venir* 'come', *caminar* 'walk', *sonar* 'sound', *parecer* 'seem', *escuchar* 'hear', *oler* 'smell', *mirar* 'look', and *ver* 'look' prefer *komo/como* 'like', which in turn is interesting because most of these verb lemmas are epistemic judgment predicates and align with the semantics of similative constructions (see Section 3). Finally, there is a last class of verbs whose constructional preferences differ between the languages; those are *sentir* 'feel', *ser* 'be', *brincar* 'jump', *correr* 'run', *mover* 'move', and *cantar* 'sing'.

A second potentially interesting way to explore the results is to determine the prototypical configurations for each level of the response variable construction schema, i.e., for *komo si/como si* 'as if' and *komo/como* 'like'. We follow Gries (2003a, b) and Bernaisch et al. (2014) and operationalize prototypes on the basis of the configurations of features with the highest predicted probabilities for *komo si/como si* 'as if' and *komo/como* 'like'. This operationalization is based on the definition of prototypes as abstract configurations of features that have the highest cue validity for the categories of interest, here *komo si/como si* 'as if' and *komo/como* 'like', where cue validity in turn is defined such that the cue validity of a feature *f* for a category *c* is high if (i) many, most, or all members of *c* have *f* and (ii)

many, most, or all non-members *c* do not, a definition that perfectly aligns with statistical predictive modeling approaches.

The prototypes of *komo/como* 'like' in MexSpa, HuaNah, PapTot, SanGabHua, and UxpChin all arise only with non-locatives and the epistemic judgment verbs *escuchar* 'hear', *oler* 'smell', *ver* 'look', *mirar* 'look', and *parecer* 'seem'. However, one interesting difference among the languages of the present study should be noted here. Similative constructions with the epistemic judgement verb lemmas *mirar* 'look', *oler* 'smell', *ver* 'look' and with non-locative NPs are prototypes in MexSpa. The same prototypes are also found in HuaNah, SanGabHua, and UxpChin except that they can occur with both non-locative and locative NPs in these Mesoamerican languages. The prototypes of *komo si/como si* 'as if' in MexSpa, HuaNah, PapTot, SanGabHua, and UxpChin all arise only with locative NPs and mistaken identity verbs (e.g., *ser* 'be', *actuar* 'act', *comportar* 'behave', and *portar* 'behave'). Interestingly, the distribution of the mistaken identity verb lemmas is different in MexSpa and the Mesoamerican languages considered here. In MexSpa, the only prototype is that in which pretence constructions occur with the mistaken identity verb lemma *ser* 'be' and are followed by a locative NP. By contrast, in HuaNah, PapTot, SanGabHua, and UxpChin, pretence constructions appear with the mistaken identity verb lemmas *actuar* 'act', *comportar* 'behave', and *portar* 'behave' and are followed by a locative NP.

## 3. Discussion

### 3.1 Partial dependence scores discussion

In this section, we particularly focus on what we consider the most interesting partial dependence scores mentioned in Section 2.2 (Figure 2). First, the verb lemmas *salir* 'leave', *venir* 'come', *caminar* 'walk', *sonar* 'sound', *parecer* 'seem', *escuchar* 'hear', *oler* 'smell', *mirar* 'look', and *ver* 'look' prefer similative constructions in all languages. Second, the verb lemmas *actuar* 'act', *hablar* 'speak', *portar* 'behave', *mostrar* 'show', *llorar* 'cry', *comportar* 'behave', *nadar* 'swim', *pensar* 'think', and *saltar* 'jump' prefer pretence constructions in all languages.

For MexSpa, Olguín Martínez and Gries (2025a) have shown that the first slot of similative constructions with non-locative and locative NPs prefers to occur with epistemic judgment predicates, such as *parecer* 'to seem', *mirar* 'to look', *ver* 'to look', and *sonar* 'to sound', among others, as in the examples in (7a)–(b)).

(7) *MexSpa (Indo-European)*
a. *parece como una buena idea.*
3SG.seem.PRS like INDEF good idea
'It seems like a good idea.' (16-06-05 MX TelevisaDeportes.com)
b. *suena como un buen plan.*
3SG.sound.PRS like INDEF good plan
'It sounds like a good plan.' (17-06-20 MX Aristeguinoticias)

They propose that the meaning of similative constructions is that of 'to give the same appearance as something/someone'. Accordingly, the meaning of epistemic verb lemmas harmonizes with the meaning of similative 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'). As Olguín Martínez and Gries (2025a) put it, speakers need to indicate the type of evidence they have to say that 'X resembles Y'. The MexSpa corpus data of the present study align with these results. In particular, perception verbs used in similative constructions are common in the present study and show an epistemic function. For instance, the example in (7a): *parece como una buena idea* 'it seems like a good idea' is roughly the same as lit. 'it gives the same appearance as a good idea.' This use of perception verbs with an epistemic function has been documented in different 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 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 Ibarretxe-Antuñano, 1999; Croft & Cruse, 2004, p. 55).

Interestingly, HuaNah, PapTot, SanGabHua, and UxpChin also contain similative constructions with similar lexical preferences (i.e., perception verbs used epistemically), as in the examples in (8)–(11). What this seems to indicate is that these Mesoamerican languages have not only borrowed the similative marker from MexSpa, but also similar lexical preferences of the first slot of these constructions, i.e., epistemic verbs. This aligns with Hypothesis 1 (see Section 1).

(8) *HuaNah (Uto-Aztecan)*
*k-huelita komo animas.*
3SG.OBJ-3SG.SBJ.look like dead
'He looks like a dead (guy).' (The drunk boy story-07/17/2022)

(9) *PapTot (Totonacan)*
*Carlos ta-si':ː komo quːlu.*
Carlos INGR-3SG.SBJ.look like old.man
'Carlos looks like an old man.' (My husband-08/29/2023)

(10) *SanGabHua (Mayan)*
*Hwaːn hel komo teʔ.*
Juan 3SG.SBJ.look like tree
'Juan looks like a tree.' (My brother-08/22/2017)

(11) *UxpChin (Oto-Manguean)*
*ca-jnéng komo jaang angel.*
COMPL-3SG.SBJ.look like INDEF angel
'He looked like an angel.' (My grandfather-07/16/2018)

There are other lexical preferences in the five languages in the present study that cannot be characterized as perception verbs used epistemically (i.e., *salir* 'leave', *venir* 'come', *caminar* 'walk') and that deserve some discussion here. These verbs can be characterized as mistaken identity predicates. In (12a), the discourse context makes it clear that the literal sense of this example is 'the boy is imitating the way in which bunnies jump'. 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 (12b), the point is not that he imitates his way of running, rather it is that he is wearing the same outfit as him.

(12) *MexSpa (Indo-European)*
a. *brinca como conejo. Brinca muy alto.*
3SG.jump.PRS like bunny 3SG.jump.PRS very high
'He jumps like a bunny. He jumps very high.'
(18-03-14 MX Reporte Indigo)
b. *corre como yo. Usa el mismo tipo de camiseta.*
3SG.run.PRS like 3SG 3SG.wear.PRS DEF same type of t-shirt
'He runs like me. He wears the same t-shirt.'
(12-12-29 MX Zócalo de Monclova)

A similar function is attested in similative constructions in HuaNah, PapTot, SanGabHua, and UxpChin. For example, in HuaNah, similative constructions with motion verbs can be used in ways comparable to MexSpa. In (13a), the discourse context clearly indicates that the intended meaning is 'the man is imitating the way bunnies jump.' However, there are also cases in which similative constructions do not convey the meaning 'X acts/behaves like Y,' but rather 'X resembles Y.' In (13b), the point is not that he is imitating the way old men leave, but that he is dressed like them.

(13) *HuaNah (Uto-Aztecan)*

a. *hual-motlalo-k* *komo kuatochi.*
DIR-3SG.SBJ.run-PFV like bunny
'He ran like a bunny.'
*mo-linia-yaya* *mo-ihkos-teh-yaya* *kemah*
REFL-3SG.SBJ.move-IPFV REFL-3SG.SBJ.separe-legs-IPFV when
*hual-motlalo-yaya.*
DIR-3SG.SBJ.run-IPFV
'He moved his hind legs while he was running (from one place to another).' (The bunny and the turtle story-07/15/2022)

b. *nopa okichpi-tl kis-k* *komo huehue-tsi.*
DEF boy-ABS 3SG.SBJ.leave-PFV like old-DIM
'The boy left (the house) like an old man.'
*mo-kenti-yaya* *tle* *nopa chichiltik komo nochi huehue-tsi.*
REFL-3SG.SBJ.dress-IPFV DEF DEF red like all old-DIM
'He was wearing a red (cap) like most old people (in our community).'
(The spring-07/18/2022)

The discussion now turns to *pretence constructions*. In the case of MexSpa, Olguín Martínez & Gries (2025a) have demonstrated that the first slot of these constructions — whether or not followed by a locative NP — tends to co-occur with mistaken identity verbs such as *actuar* 'to act' and *comportar* 'to behave', as in (14a)–(b)). These constructions convey meanings related to imitation, pretense, or aspirational behavior (see also Olguín Martínez, 2021; Royo Viñuales & Van Linden, 2025). The semantic compatibility between the construction and mistaken identity verbs lies in their shared focus on enacting behavior that resembles that of someone or something else, that is, 'X behaves in a way reminiscent of Y'. The MexSpa corpus data analyzed in this study corroborates these findings: mistaken identity verb lemmas, such as *actuar* 'act', *portar* 'behave', *mostrar* 'show', and *comportar* 'behave' prefer to occur in the first slot of MexSpa pretence constructions.

(14) *MexSpa (Indo-European)*

a. *se* *comporta* *como* *si* *fuera* *un* *doctor.*
PRON 3SG.act.PRS as if 3SG.be.SUBJ INDEF doctor
'He acts as if he were a doctor.' (250 18-10-05 MX 20 minutos.com.mx)

b. *actúa* *como* *si* *fuera* *el* *rey.*
3SG.act.PRS as if 3SG.be.SUBJ DEF king
'He acts as if he were the king.' (49 16-11-20 MX LEVELUP)

A closer examination of the results reveals that HuaNah, PapTot, SanGabHua, and UxpChin also exhibit pretence constructions with similar lexical preferences (i.e.,

mistaken identity verbs), as illustrated in the examples in (15)–(18). This suggests that these Mesoamerican languages have not only borrowed the pretence marker from MexSpa, but have also developed similar lexical preferences in the first slot of these constructions. These findings support Hypothesis 2 (see Section 1).

(15) *HuaNah (Uto-Aztecan)*
*ixehua-k* *komo si el-s* *se* *tsopilo-tl*
3SG.SBJ.behave-PFV as if 3SG.SBJ.be-IRR INDEF vulture-ABS
'He behaved as if he were a vulture.' (The storm story-07/16/2022)

(16) *PapTot (Totonacan)*
*maqi'ːqlhlá-lh* *komo si wan-níːt* *ja'i* *chi'xkú xa-níː-n*
3SG.SBJ.act-COMPL like if 3SG.SBJ.be-PFV INDEF man DET-3SG.SBJ.die-NMLZ
'He acted as if he were a dead man.' (The fool guy-08/14/2023)

(17) SanGabHua (Mayan)
*toːk'oj-uw-ø* *komo si wenk'-ow-al* *puːkel.*
3SG.SBJ.act-TRANS-COMPL as if 3SG.SBJ.become-TRANS-INCOMPL big
'He acted as if he were big.' (My ranch-08/22/2017)

(18) UxpChin (Oto-Manguean)
*ca-jmée=b* *komo si cofa' dsée=b=re*
COMPL-3SG.behave=EMPH like if IRR be.sick=EMPH=3SG
'He behaved as if he were sick.' (The new teacher-07/14/2018)

### **3.2** Prototype results discussion

It is likely that the lexical preferences of similative and pretence constructions in HuaNah, PapTot, SanGabHua, and UxpChin have been shaped by intensive contact with MexSpa. However, what remains unclear is whether the syntax of these constructions in the respective Mesoamerican languages has also been influenced by MexSpa. This section addresses that question directly. In particular, it focuses on locative and non-locative NPs and their interaction with verb lemmas in similative and pretence constructions. As discussed in Section 1, such constructions may include a NP that can be characterized as either locative or non-locative, as illustrated in HuaNah examples in (19). To investigate this issue, we analyze prototypical similative and pretence constructions in each language included in this study with an emphasis on syntactic differences across the languages.

(19) *HuaNah (Uto-Aztecan)*
a. *hual-motlalo-k* *komo si el-s* *kuatochi.*
DIR-3SG.SBJ.run-PFV like if 3SG.SBJ.be-IRR bunny
'He ran as if he were a bunny.' (The bunny and the turtle story-07/15/2022)

b. *yelia-k* *komo si el-s* *pa* *parke*.
3SG.SBJ.behave-PFV as if 3SG.SBJ.be-IRR LOC park
'He behaved as if he were at a park.' (The butcher story-07/15/2022)

As noted in Section 2.2, prototypical similative constructions in MexSpa, HuaNah, PapTot, SanGabHua, and UxpChin involve the epistemic verb lemmas *mirar* 'look', *oler* 'smell', and *ver* 'look' and non-locative NPs. Interestingly, while these prototypes occur only with non-locative NPs in MexSpa, they arise with both non-locative and locative NPs in HuaNah, PapTot, SanGabHua, and UxpChin. What this seems to indicate is that the Mesoamerican languages in the present study have developed prototypical similative constructions with more complex syntactic patterns than those found in MexSpa. In contrast, prototypical pretence constructions across these languages involve mistaken identity verbs and locative NPs. While the syntax of these constructions is broadly similar (they only occur with locative NPs), there are notable differences in verb selection. In MexSpa, prototypes exclusively involve the mistaken identity verb lemma *ser* 'be'. Meanwhile, the Mesoamerican languages include a wider range of mistaken identity verb lemmas such as *actuar* 'act', *comportar* 'behave', and *portar* 'behave'.

The fact that HuaNah, PapTot, SanGabHua, and UxpChin exhibit a larger number of prototypes than MexSpa is surprising and challenges the widely held claim that language contact leads to grammatical simplification in a recipient language (e.g., Kusters, 2008). As Trudgill (2009, p. 99) argues, such simplification is often attributed to "the relative inability of adult humans to learn new languages perfectly." In learning a new language, adult speakers may reduce grammatical complexity. Against this backdrop, the presence of more diverse prototypes for similative and pretence constructions in the Mesoamerican languages considered here than in MexSpa is unexpected.

One plausible explanation is that MexSpa has alternative ways of expressing similative and pretence meanings, such as *igual que si* 'as if' (e.g., *llueve igual que si fuera invierno* 'it's raining as if it were winter'), *tal como si* 'as if' (e.g., *actuó tal como si no me conociera* 'he acted just as if he didn't know me'), and *igual que* 'like' (e.g., *corre igual que mi hermano* 'he runs like my brother'), among others. Accordingly, MexSpa may have developed distinct prototypes in these constructions that are not attested in *como* 'like' and *como si* 'as if' constructions. Since HuaNah, PapTot, SanGabHua, and UxpChin only contain *komo* 'like' and *komo si* 'as if' for expressing similative and pretence meanings (see Section 2.1), this restriction may account for their greater proliferation of prototypes compared to MexSpa.

## 4. Implications

The findings of this study align with previous research showing that, in language contact situations, a construction is rarely replicated intact from one language to another (Johanson, 2008, p. 67; Matras, 2009, p. 148; Mithun, 2025). As was shown in Section 3, similative and pretence constructions in HuaNah, PapTot, SanGabHua, and UxpChin share similar lexical preferences with their MexSpa counterparts. However, there are a number of syntactic differences in the treatment of locative and non-locative NPs in similative and pretence constructions. This pattern reflects what Johanson (2008, p. 67) terms *selective grammatical copying*, a process in which the diffusion of a construction from one language to another may affect some dimensions (e.g., phonological, semantic, morphological, syntactic, morpho-syntactic) but not others (see also Matras, 2009, p. 148). For example, many Mesoamerican languages have borrowed connectives from MexSpa along with expletive negative markers, as in (20). Expletive negation refers to a negative element that lexically encodes negation but does not alter the truth value of the proposition in which it appears (Espinal, 1992, p. 49). In other words, it is a negative marker without negative meaning. In MexSpa, 'until' clauses may contain the negative marker *no* (21), which is expletive and can be omitted without changing the temporal relation between clauses. Strikingly, while expletive negation in Mesoamerican languages emerged under the influence of MexSpa, it has developed discourse functions absent in the source language. Specifically, when expletive negation appears in the 'until' clause, the proposition is interpreted as conveying surprise; when it is absent, no such evaluative stance is implied (Olguín Martínez, 2024b).

(20) *HuaNah (Uto-Aztecan)*
*nopa diablo ach-tla-tsotsona biolin*,
DEF devil NEG-INDEF.OBJ-3SG.SBJ.play violin
'The devil did not play the violin,
*asta ke amo tlahuelchihua-k-e.*
until that NEG 3PL.SBJ.get.angry-PFV-PL
until they (men) got angry.' (Olguín Martínez, 2024, p. 755)

(21) *MexSpa (Indo-European)*
*el hombre no dormirá*,
DEF man NEG sleep.FUT.3SG
'The man will not sleep,
***hasta** que la fiesta **no** comience.*
until that DEF party NEG start.PRS.SUBJ.3SG
until the party starts.' (Olguín Martínez, 2024, p. 755)

The present study has also demonstrated that in language contact situations, constructions can emerge through both matter and pattern replication. In some cases, speakers of recipient languages borrow grammatical markers from a donor language with their exact forms, although minor differences in substance may occur as these sounds are adapted into the recipient language's native phonological system. This process is referred to as *matter replication* (Sakel, 2007). Conversely, speakers may replicate grammatical patterns from the donor language using native linguistic material, a process known as *pattern replication* (Sakel, 2007). Here, only the structural patterns of the donor language are replicated, without borrowing the phonetic substance. While previous research has shown that recipient languages may exhibit either matter or pattern replication, the present study reveals that both can co-exist within the same construction during lexico-syntactic transfer. Specifically, in HuaNah, PapTot, SanGabHua, and UxpChin, the verb lemmas found in similative and pretence constructions result from pattern replication, whereas the similative and pretence markers themselves are outcomes of matter replication.

## 5. Final remarks

Using predictive modeling, we explored the ranges of factors influencing the choice of similative and pretence markers in the languages of the present study. Based on two evaluation steps: (i) partial dependence scores and (ii) the identification of constructional prototypes separately for MexSpa, HuaNah, PapTot, SanGabHua, and UxpChin, we were able to provide a novel way to explore how constructional templates and their lexical preferences diffuse in language contact situations.

The present study has shown that HuaNah, PapTot, SanGabHua, and UxpChin have not only borrowed similative and pretence markers from MexSpa, but also lexical preferences of the first slot of these constructions. HuaNah, PapTot, SanGabHua, and UxpChin contain similative constructions with similar lexical preferences (i.e., perception verbs used epistemically) as in MexSpa. Likewise, these Mesoamerican languages contain pretence constructions with similar lexical preferences (i.e., mistaken identity verbs) as in MexSpa. However, there are a number of syntactic differences in the treatment of locative and non-locative NPs in similative and pretence constructions. For instance, prototypical similative constructions in MexSpa, HuaNah, PapTot, SanGabHua, and UxpChin involve the epistemic verb lemmas *mirar* 'look', *oler* 'smell', and *ver* 'look' and non-locative NPs. Interestingly, while in MexSpa these prototypes occur only with non-locative NPs, they arise with both non-locative and locative NPs in HuaNah, PapTot,

SanGabHua, and UxpChin. On the other hand, prototypical pretence constructions across these languages involve mistaken identity verbs and locative NPs.

There are other areas of the world like the Huasteca area in Veracruz (e.g., many indigenous languages spoken within the former Soviet Union have incorporated connectives from Arabic, Persian, and Russian; Stolz & Levkovych, 2022). Accordingly, the findings of this study may hold broader relevance for linguists investigating language contact phenomena worldwide, promoting cross-regional comparison. It is our hope here that the proposed method will be valuable to other linguists to explore language contact situations and areal clusters from an integrative, non-modular perspective.

As a sobering note, mention should be made of the following issue. We have proposed that the lexical preferences of similative and pretence constructions in HuaNah, PapTot, SanGabHua, and UxpChin appear to have been shaped by intensive contact with MexSpa. This raises broader questions: do languages spoken in other areas of the world tend to prefer epistemic judgment predicates in similative constructions and mistaken identity predicates in pretence constructions? Put differently, are these preferences not only present in the languages examined here, but also widespread cross-linguistic tendencies? How, then, can we distinguish similarities that result from language contact from those that reflect universal patterns? (See Schapper & Koptjevskaja-Tamm, 2021 for related discussion.) As our knowledge of individual languages and the typology of similative–pretence constructions grows, we should become increasingly adept at discerning parallels due to contact from those rooted in universal patterns.[5] Future research into a wider range of languages should help to verify, extend and, if necessary, amend the picture presented here.

## Acknowledgments

We would like to thank all HuaNah, PapTot, SanGabHua, and UxpChin speakers who have kindly shared with us their data. Many thanks also to two anonymous referees for useful comments on previous drafts of the paper. Needless to say, all remaining shortcomings are our own.

**5.** As correctly pointed it out by one anonymous reviewer: "Usage-Based CxG, by definition, distances itself from the notion of linguistics universals (while accepting universal cognitive principles such as entrenchment, categorization, and processing)". We acknowledge that relying heavily on the concept of linguistic universals is not ideal, since, as Croft (2001, p. 183) argues, "constructions are language-specific, and there is an extraordinary range of structural diversity in constructions encoding similar functions across languages." However, as our survey of languages around the world has expanded, we have observed striking cross-linguistic parallels in the ways similative and pretence meanings are expressed. For this reason, we have chosen to retain the discussion in its current form.

## Abbreviations

| | |
|---|---|
| 1 | first person |
| 2 | second person |
| 3 | third person |
| ABS | absolutive |
| ANIM | animate |
| ART | ARTICLE |
| COMPL | completive |
| DEF | definite |
| DEM | demonstrative |
| DET | determiner |
| DIM | diminutive |
| DIR | directional |
| EMPH | emphasis |
| INACC | inaccusative |
| INAN | inanimate |
| INCOMPL | incompletive |
| INDEF | indefinite |
| INGR | ingressive |
| IPFV | imperfective |
| IRR | irrealis |
| LOC | locative |
| MIDDL | middle |
| NEG | negative |
| NMLZ | nominalizer |
| PFV | perfective |
| PL | plural |
| POSS | possessive |
| PRON | pronominal |
| PRS | present |
| REFL | reflexive |
| SBJ | subject |
| SG | singular |
| SUBJ | subjunctive |
| TRANS | transitive. |

## References


Béchet, Christophe. (2020). An empirical perspective on the contact between English and French: A case study on substitutive complex prepositions. *Linguistics Vanguard*, *6*.

Beckner, Clay, Blythe, Richard, Bybee, Johan, Christiansen, Morten. H., Croft, William, Ellis, Nick C., Holland, John, Ke, Jinyun, Larsen-Freeman, Diane, & Schoenemann, Tom. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*, 1–26.

Bernaisch, Tobias J., Gries, Stefan Th., & Mukherjee, Joybrato. (2014). The dative alternation in South Asian English(es): Modelling predictors and predicting prototypes. *English World-Wide*, *35*, 7–31.

Boas, Hans. C., & Höder, Steffen. (Eds.) (2018). *Constructions in contact: Constructional perspectives on contact phenomena in Germanic languages*. John Benjamins.

Bullock, Barbara E., Serigos Jacqueline, & Toribio, Almeida J. (2021). Exploring a loan translation and its consequences in an oral bilingual corpus. *Journal of Language Contact*, *13*, 612–635.

Clyne, Michael. (2003). *Dynamics of language contact. English and immigrant languages*. Cambridge University Press.

Croft, William. (2001). *Radical construction grammar. Syntactic theory in typological perspective*. Oxford University Press.

Croft, William, & Cruse, Alan. (2004). *Cognitive linguistics*. Cambridge University Press.

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

Dexter-Sobkowiak, Elwira. (2022). Language contact in the Huasteca: The impact of Spanish on Nahuatl and Tének. [Doctoral dissertation, University of Warsaw].

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

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

Espinal, Maria Teresa. (1992). Expletive negation and logical absorption. *The Linguistic Review*, *9*, 333–358.

Kay, Paul, & Fillmore, Charles P. (1999). Grammatical constructions and linguistic generalization: The what's X doing Y? construction. *Language*, *75*, 1–33.

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

Greenwell, Brandon M. (2017). pdp: An R package for constructing partial dependence plots. *The R Journal*, *9*, 421–436.

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

Gries, Stefan Th. (2003b). Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics*, *1*, 1–27.

Gries, Stefan Th. (2021). *Statistics for linguistics with R*. 3rd rev. De Gruyter Mouton.

Gries, Stefan Th., & Stefanowitsch, Anatol. (2004). Extending collostructional analysis: A corpus-based perspectives on 'alternations'. *International Journal of Corpus Linguistics*, *9*, 97–129.

Gries, Stefan Th., & Hilpert, Martin. (2008). The identification of stages in diachronic data: Variability-based neighbor clustering. *Corpora*, *3*, 59–81.

Heine, Bernd, & Kuteva, Tania. (2005). *Language contact and grammatical change*. Cambridge University Press.

Harris, Zellig S. (1970). *Papers in structural and transformational linguistics*. Reidel.

Hernandez-Mena, Carlos. (2019). *TEDx Spanish Corpus. Audio and transcripts in Spanish taken from the TEDx Talks*. Universidad Nacional Autónoma de Mexico.

Hilpert, Martin. (2006). Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory*, *2*, 243–257.

Hilpert, Martin. (2008). *Germanic future constructions: A usage-based approach to language change*. John Benjamins.

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

Hothorn, Torsten, & Zeileis, Achim. (2015). partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research*, *16*, 3905–3909.

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

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

Johanson, Lars. (2008). Remodeling grammar: Copying, conventionalization, grammaticalization. In Siemund, Peter & Noemi Kintana (Eds.), *Hamburg Studies on Multilingualism* (pp. 61–79). John Benjamins.

Kusters, Wouter. (2008). Complexity in linguistic theory, language learning, and language change. In Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson (Eds.), *Language complexity: Typology, contact, and change* (pp. 3–22). John Benjamins.

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

Lara, Luis F., Medina Urrea, Alfonso, Rosales Martínez, Alejandro, Diez Sánchez, Carlos, F., & Serralde Galicia, Juan L. (2018). *El Corpus del español mexicano contemporáneo*. https://cemcii.colmex.mx/

Matras, Yaron. (2009). *Language contact*. Cambridge University Press.

Mithun, Marianne. (1992). External triggers and internal guidance in syntactic development: Coordinating conjunction. In Gerritsen, Marinel & Dieter Stein (Eds.), *Internal and external factors in syntactic change. Trends in linguistics* (pp. 89–129). Mouton De Gruyter.

Mithun, Marianne. (2025). Constructions and language contact. In Fried, Mirjam & Kiki Nikiforidou (Eds.), *The Cambridge handbook of Construction Grammar* (pp. 469–496). Cambridge University Press.

Olguín Martínez, Jesus. (2021). Hypothetical manner constructions in world-wide perspective. *Journal of Linguistic typology at the crossroads*, *1*, 2–33.

Olguín Martínez, Jesus. (2022). Contact-induced language change: The case of Mixtec adverbial clauses. *Journal of Language Contact*, *15*, 1–70.

Olguín Martínez, Jesus. (2023). Areality of clause-linkage: The consecutive construction in Mesoamerican languages. *Voprosy Jazykoznanija*, *3*, 122–142.

Olguín Martínez, Jesus. (2024a). Semantically negative clause-linkage: 'Let alone' constructions, expletive negation, and theoretical implications. *Linguistic Typology*, *28*, 1–52.

Olguín Martínez, Jesus. (2024b). 'Until' clauses and expletive negation in Huasteca Nahuatl. *Studies in Language*, *48*, 753–780.

Olguín Martínez, Jesus, & Gries, Stefan Th. (2025). The similative-pretence alternating pair and filler-slot relations. A revised version of distinctive collexeme analysis. *Constructions and Frames*, *17*, 65–91.

Olguín Martínez, Jesus, & Gries, Stefan Th. Gries. (2025). Similative-pretence constructions in language contact situations: A Usage-Based Construction Grammar perspective. *Cognitive Linguistic Studies*, *12*, 292–321.

Royo Viñuales, Victor, & Van linden, An. (2025). Beyond hypothetical manner: A functional typology of insubordinate *como si*-clauses. *Folia Linguistica*, *59*, 759–784.

Sakel, Jeanette. (2007). Types of loan: Matter and pattern. In Matras, Yaron & Jeanette Sakel (Eds.), *Grammatical borrowing in cross-linguistic perspective* (pp. 15–30). Mouton de Gruyter.

Schapper, Antoinette & Koptjevskaja-Tamm, Maria. (2021). Introduction to special issue on areal typology of lexico-semantics. *Linguistic Typology*, *26*, 199–209.

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

Smirnova, Irina, Vetrinskaya, Victoria, & Clemente–Smirnova, Svetlana. (2023). Grammatical features of Spanish in the Mexican state of Oaxaca. *E3S Web of Conferences 371*, 05034.

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

Stolz, Thomas, & Levkovych, Nataliya. (2022). On loan conjunctions: A comparative study with special focus on the languages of the former Soviet Union. In Nataliya Levkovych (Ed.), *Susceptibility vs. resistance: Case studies on different structural categories in language contact situations* (pp. 259–392). De Gruyter Mouton.

Strobl, Carolin, Malley, James, & Tutz, Gerhard. (2009). An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychological Methods*, *14*, 323–348.

Thomason, Sarah G., & Kaufman, Terrence. (1988). *Language contact, creolization, and genetic linguistics*. University of California Press.

Trudgill, Peter. (2009). Sociolinguistic typology and complexification. In Sampson, Geoffrey, Gil, David & Peter Trudgill (Eds.), *Language complexity as an evolving variable* (pp. 98–109). Oxford University Press.

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

Wiesinger, Evelyn. (2021). The Spanish verb-particle construction [V para atrás]. In Boas, Hans C. & Steffen Höder (Eds.), *Constructions in Contact 2 Language change, multilingual practices, and additional language acquisition* (pp. 139–187). John Benjamins.

Wilson, Damián V. (2013). One construction, two source languages: Hacer with an English infinitive in bilingual discourse. In Carvalho, Ana & Sara Beadrie (Eds.), *Proceedings from the 6th International Workshop on Spanish sociolinguistics* (pp. 123–134). Cascadilla Proceedings Project.

## Address for correspondence

Jesús Olguín Martínez
Department of Linguistics
University of Hong Kong
Run Run Shaw Tower, Centennial Campus
Pokfulam Road
Hong Kong
jfolguinmartinez@gmail.com

## Biographical notes

**Jesús Olguín Martínez** is an Assistant Professor in Linguistics at the University of Hong Kong. He is a typologist at the intersection of corpus linguistics, usage-based linguistics, language documentation, Construction Grammar (CxG), cognitive linguistics, and quantitative linguistics, who uses a variety of different methodologies to explore natural discourse data across Indo-European and non-Indo-European languages. His research spans three closely connected areas: cross-constructional analysis, the interaction between syntax and other domains of language use, and language contact in Mesoamerica.

https://orcid.org/0000-0002-4555-4213

**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: 5 September 2025
Date accepted: 26 November 2025
Published online: 20 January 2026