Similative-pretence constructions in language contact situations

A Usage-Based Construction Grammar perspective

Jesús Olguín Martínez and Stefan Th. Gries
Illinois State University | Justus-Liebig University Giessen

The present study introduces a method that can be used to explore in a quantitatively rigorous yet less demanding way (both in terms of data and statistical requirements) how constructional templates and their lexical preferences (lexico-syntactic transference) diffuse in language contact situations. The study investigates the influence of Mexican Spanish similative-pretence constructions on Huasteca Nahuatl similative-pretence constructions as a proof-of-concept kind of application for our method. Speakers of Huasteca Nahuatl have borrowed the markers *komo* 'like' and *komo si* 'as if' from Mexican Spanish to express similative (e.g., *she swims like a fish*) and pretence meanings (e.g., *she swims as if she were a fish*), respectively. Using a conditional inference forest, the paper demonstrates that speakers of Huasteca Nahuatl have not only borrowed these markers from Mexican Spanish, but also lexical preferences (e.g., verb lemmas) of the constructions in which these markers occur. These findings show that the rigid partition of structural levels that has been adopted by traditional models of language contact proves inadequate for describing complex language situations. The method introduced here provides an integrative, non-modular way to explore language contact from a Usage-Based Construction Grammar perspective.

Keywords: language contact, usage-based, similatives, pretence, Huasteca Nahuatl, Spanish

1. Introduction

Usage-Based Construction Grammar (CxG) linguists have been largely interested in the analysis of the historical development of constructions (e.g., Elizabeth

https://doi.org/10.1075/cogls.25015.olg
Cognitive Linguistic Studies 12:2 (2025), pp. 292–321. ISSN 2213-8722 | E-ISSN 2213-8730

Closs Traugott & Graeme Trousdale 2013: 39; Lotte Sommerer & Elena Smirnova 2020). In particular, special attention has been paid to how constructions may give rise to new constructions, i.e., new nodes in the construction, which is called constructionalization (Traugott 2003), and how the interrelations of grammatical constructions and their lexical collocates may change over time in individual languages (Diachronic Collostructional Analysis; Hilpert 2006; Gries & Martin Hilpert 2008).

Despite Usage-Based CxG linguists' concern with processes of emergence and change, overt links to contact linguistics are still in their infancy (Nikolay Hakimov & Ad Backus 2021). Recent years "have seen a slow but steadily increasing interest in CxG and multilingualism, resulting in a small, but growing body of literature" (Hans C. Boas & Steffen Höder 2018: 6). For instance, Höder (2012, 2014) introduces a framework called Diasystematic Construction Grammar, a Usage-Based CxG approach to language contact phenomena, ranging from contact-related language change to multilingual practices. Besides this effort, other attempts have been made at analyzing language contact situations from a CxG perspective, such as Hilpert and Jan-Ola Östman (2016), Boas and Höder (2018, 2021), and Eline Zenner et al. (2019).

One of the main goals of Usage-Based CxG studies dedicated to the analysis of language contact has been the following. It is well-known that the structural outcome of language contact is rarely restricted to only one level of linguistic structure, "but usually involves what is traditionally thought of as belonging to different parts of the language system" (Boas & Höder 2018: 25), such as lexico-syntactic transference, i.e., the diffusion of lexical items and corresponding syntactic constructions (Michael Clyne 2003). A number of studies have made inroads into this territory. First, Damián Vergara Wilson (2013) explores the compound verb construction *hacer*+V. This is a multi-word construction in which the verb *hacer* 'to make' contributes tense-aspect-mood while the open verbal slot takes a wide variety of bare English infinitives contributing the lexical information. In his qualitative investigation based on data from the New Mexico Spanish/English Bilingual corpus, he finds out that the light verb *hacer* 'to make' was followed by verb lemmas, such as *apply, audit, cater, clock out, collect, cook, cremate, dance, decide, feed, impress, pay attention, slide*, and *walk*, among others. These verb lemmas do not match the verb lemmas of the Spanish *hacer*+V construction. Accordingly, Wilson argues that this provides evidence that bilingual discourse practices can nurture the emergence of new constructions.

Second, Evelyn Wiesinger (2021) investigates how lexical preferences (i.e., verb lemmas) of constructions may diffuse through language contact. In her study, she provides a corpus investigation on the role of V *para atrás* constructions in

European, Mexican, and US Spanish and shows how the verb lemmas of this construction may diffuse through language contact.

Third, Barbara E. Bullock et al. (2021) examines a partially specific Spanish construction: *agarrar* NP construction, which means *to get something* in Spanish varieties of Texas. In this construction, the subject is the recipient or beneficiary of an abstract attainment. As the core sense of *agarrar* as a lexical verb is that of physical grasping, the verb in these constructions is clearly not used in its literal sense. Instead, *agarrar* appears to take on the more generalized sense of its English equivalent *get*, which is normally expressed in Spanish via verbs such as *obtener, recibir, conseguir, sacar*, and *ganar*. By using the bilingual Spanish in Texas corpus as well as in monolingual Spanish corpora, they demonstrate that the verb *agarrar* 'grab, grasp' in this construction in Texas is modelled on the English support verb construction *get* NP (e.g., *get help*) in that it is followed by similar NPS (e.g., achievement or benefit). The lexical preferences of the Spanish Texas *agarrar* NP constructions may thus be regarded as a bona fide example of contact-induced conceptual transfer.

Fourth, Christophe Béchet (2020) explores the potential contact effects of French on English with regard to the use of substitutive complex prepositions using a highly statistical method, multifactorial deviation analysis with regression/random forests [MuPDAR(F)]. This method has been applied successfully in learner corpus research to determine how and why choices made by non-native speakers differ from those made by native speakers (Gries & Allison S. Adelman 2014; Sandra C. Deshors & Gries 2016). Correspondingly, Béchet's goal was to show in what ways and to what extent English conforms to French in the use of *in lieu of* and *in place of*, but also the extent to which it deviates from French. Béchet first modeled the English and French speakers' choice between *in lieu of* and *in place of* using random forests. Then, he applied this French native-speaker model to the English learners' data, assuming constructional correspondence between *au lieu de* and *in lieu of* as well as between *à la place de* and *in place of*. After this, he computed a deviation score (DEV) on the basis of a reference construction, so that P(*à la place de*)–P(*in place of*) = DEV and then explored the deviations visually and qualitatively. He demonstrates that English speakers borrowed the forms of substitutive linkers from French but did not internalize all their foreign distributional properties (e.g., the use of the same verb lemmas).

While the studies mentioned above have made inroads into the study of lexico-syntactic transference from a Usage-Based CxG perspective, the paper argues that they are qualitative and not collostructional (e.g., Wiesinger 2021), have not employed rigorous statistical methods to explore this domain (e.g., Wilson 2013) and have used quantitative methods that make too high demands in terms of (i) what kinds and how much data are needed and (ii) technicality (e.g.,

MuPDAR). The present study introduces a new predictive-modeling approach that can be used for exploring in a quantitatively rigorous yet less demanding way (both in terms of data and statistical requirements) how constructional templates and their lexical preferences diffuse in language contact situations. Drawing inspiration from the very recent study of Olguín Martínez and Gries (2025), the paper follow their lead and investigate the influence of Mexican Spanish similative-pretence constructions in Huasteca Nahuatl similative-pretence constructions as a proof-of-concept kind of application for our method.

Speakers of Mexican Spanish express similative meanings by the constructions in (1) and pretence meanings by the constructions in (2). Note that Noun Phrases (NP) of both constructions may be non-locative or locative.

(1) a. *como* 'like' construction with NP
*se comporta **como** un doctor.*
PRON 3SG.act.PRS like a doctor
'He acts like a doctor.' (4000 16-06-12 MX Mundo TKM México)
b. *como* 'like' construction with LOC NP
*se siente **como** en la playa.*
PRON 3SG.feel.PRS like LOC DEF beach
'It feels like at the beach.' (38 19-03-25 MX México Desconocido)

(2) a. *como si* 'as if' construction with NP
*se comporta **como si** fuera un doctor.*
PRON 3SG.act.PRS as if 3SG.be.SUBJ a doctor
'He acts as if he were a doctor.' (250 18-10-05 MX 20 minutos.com.mx)
b. *como si* 'as if' construction with LOC NP
*se siente **como si** estuvieramos en la playa.*
PRON 3SG.feel.PRS as if 1PL.be.SUBJ LOC DEF beach
'It feels as if we were at the beach.' (39 19-04-17 MX Diario La Verdad)

Speakers of Huasteca Nahuatl express similative and pretence meanings in similar ways, as is shown in examples in (3) and (4):

(3) a. *komo* 'like' construction with NP
*hual-motlalo-k **komo** kuatochi.*
DIR-run-PFV like bunny
'He ran like a bunny.' (The bunny and the turtle story-07/15/2022)
b. *komo* 'like' construction with LOC NP
*k-huelita **komo** pa mo-cha.*
3SG.OBJ-look like LOC 2SG.POSS-house
'It looks like (we were) at your house.' (The poor man story-07/17/2022)

(4) a. *komo si* 'as if' construction with NP
hual-motlalo-k ***komo si*** *el-s* *kuatochi.*
DIR-run-PFV like if be-IRR bunny
'He ran as if he were a bunny.' (The bunny and the turtle story-07/15/2022)

b. *komo si* 'as if' construction with LOC NP
hual-motlalo-k ***komo si*** *el-s* *pa* *tikita.*
DIR-run-PFV like if be-IRR LOC work
'He ran as if he were in a competition.'
(The bunny and the turtle story-07/15/2022)

As shown in these examples, speakers of Huasteca Nahuatl have borrowed locative and non-locative *komo* 'like' and *komo si* 'as if' constructions from Mexican Spanish. Accordingly, the question is, have speakers of Huasteca Nahuatl also copied the lexical preferences of the first slot (verb lemmas) of Mexican Spanish similative-pretence constructions?

For Mexican Spanish, Olguín Martínez and Gries (2025) have shown that the first slot of similative *como* 'like' constructions (with and without locative NPs), as in (1), prefers to occur with epistemic judgment predicates, such as *parecer* 'to seem,' *mirar* 'to look,' *ver* 'to look,' and *sonar* 'to sound,' among others. The meaning of similative *como* 'like' 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 *como* 'like' 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 (2025) put it, speakers need to indicate the type of evidence they have to say that 'X' resembles 'Y.' On the other hand, they also find that the first slot of pretence *como si* 'as if' constructions (with and without locative NPs), as in (2), prefers to appear with mistaken identity verbs, such as *actuar* 'to act' and *comportar* 'to behave.' The meaning of pretence *como si* 'as if' constructions is that of 'to imitate, pretend, aspire to the behavior of something/someone' (see also Olguín Martínez 2021; Víctor Royo Viñuales & An Van linden 2024). This indicates that the meaning of mistaken identity verb lemmas harmonizes with the semantics of this construction 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 summarizes their results.

Bringing together these findings for Mexican Spanish with previous work by Olguín Martínez (2024a, 2024b) and others demonstrating that Huasteca Nahuatl grammar has been heavily influenced by Mexican Spanish in different ways leads to the following hypotheses, which is what we are exploring here in this paper:

Table 1. Verb lemma types occurring in the first slot of similative and pretence constructions in Mexican Spanish (summarizing Olguín Martínez & Gries 2025)

Construction type	Construction	Tye 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

- **Hypothesis 1**: Speakers of Huasteca Nahuatl have not only borrowed the pretence marker *como si* 'as if' from Mexican Spanish, but also the same lexical preferences of the first slot of these constructions (with and without locative NPs), i.e., mistaken identity verbs.
- **Hypothesis 2**: Speakers of Huasteca Nahuatl have not only borrowed the similative marker *como* 'like' from Mexican Spanish, but also the same lexical preferences of the first slot of these constructions (with and without locative NPs), i.e., epistemic verbs.

Previous research has shown that Huasteca Nahuatl not only borrowed clause-linking devices from Mexican Spanish, but also other constructional properties in which these conjunctions are attested: expletive negation (Olguín Martínez 2024a, 2024b) and matrix verbs of different types of complement clauses (Fredric W. Field 2002: 142–143). Justyna Olko (2020: 36) shows that the borrowing of Mexican Spanish conjunctions not only affects lexicon of Nahuatl, but also transforms syntactic patterns at the interclausal level. This typological shift has made Nahuatl and Mexican Spanish even more compatible and greatly reduced the constraints on the borrowability of additional foreign elements in Nahuatl, such as verbs and nouns (Olko et al. 2018). Accordingly, it seems reasonable to assume that lexical preferences of Mexican Spanish similative and pretence constructions will also be transferred to Huasteca Nahuatl similative and pretence constructions due to intense language contact.

Theoretically, the study adopts a Usage-Based CxG approach to language contact and assume that, in language contact situations, different structural levels are transferred from one language to another; put differently, we assume an integrative, non-modular approach to language contact (e.g., Boas & Höder 2018: 10) and reject a rigid differentiation and partition of language into qualitatively completely different and modular structural levels (lexicon, syntax, morphology, etc.) The study uses the term donor language to refer to Mexican Spanish in that it served as the source of diffusion of X. Moreover, we use the term recipient to

refer to Huasteca Nahuatl in that it is the language that borrowed X from a donor language.

Methodologically, the study adopts a predictive-modeling perspective and fitted a Conditional Inference Forest (CIF) to explore this domain. This method belongs to the family of recursive partitioning methods, which have gained popularity in corpus linguistics (Gries 2021: Chapter 7), and will be explained in detail below. The investigation of language contact situations involves the analysis of complex cognitive/psycholinguistic phenomena and multifactorial relationships (e.g., why speakers make certain choices and how those might change). Accordingly, a predictive-modeling technique such as CIF s provides a natural fit to the analysis of language contact, in particular because such forests strike a good balance between predictive power and interpretability (a balance that regression models sometimes do not attain).

The remainder of the paper is structured as follows. Section 2, presents corpus data and then walks the reader through the method used here to analyze similative-pretence constructions in Mexican Spanish and Huasteca Nahuatl. Section 3 discusses in detail the results of Section 2 and argues that they have important methodological and theoretical implications to contact linguistics in general and to contact-induced change structural replication models (matter and patter replication). Section 4 concludes with pointers toward theoretical issues and a number of potentially fruitful areas for future research.

2. Methods and results

This section introduces the corpus data, outlines how the study fitted the CIF to compare Mexican Spanish and Huasteca Nahuatl similative-pretence constructions, and the results.

2.1 Corpus data, data extraction, and annotation

To investigate the influence of Mexican Spanish similative-pretence constructions in Huasteca Nahuatl similative-pretence constructions, the study used two corpora: *Corpus del Español NOW* (News on the Web) and a Huasteca Nahuatl corpus drawn from one field work period in the village of Teposteco in the Huasteca region.

The Spanish corpus chosen for the investigation of similative-pretence constructions in Mexican Spanish was the *Corpus del Español NOW* (News on the Web). It contains about 7.6 billion words and is a dynamic corpus: About four to five million words are added to this corpus each day. While this corpus is

not ideal — it is not exclusively based on Mexican Spanish data (i.e., it contains data on 21 Spanish-speaking countries) — it was the best resource the study could access. Other corpora we tried to work with, such as the *TEDx Spanish Corpus* (Carlos Hernandez-Mena 2019) and the *Corpus del Español Mexicano Contemporáneo* (L. F. Lara et al. 2018), mostly only featured the constructions of interest with clauses rather than with NPs and locative NPs (as is discussed below, this is not ideal for the present study given that we only consider similative and pretence constructions with NPs and locative NPs); the *Corpus del Español NOW* was the only corpus available to us from which the study could obtain similative and pretence constructions realized with NPs and locative NPs.

The procedure for data retrieval was the following. An exhaustive concordance of similative and pretence constructions was performed by searching the *Corpus del Español NOW* for the forms *como* 'like' and *como si* 'as if.' This generated a large sample in which *como* 'like' and *como si* 'as if' were followed by NPs, LOC NPs, and clauses, which was then trimmed down to exclude constructions in which *como* 'like' and *como si* 'as if' were followed by clauses (e.g., *el hombre grita como si alguien lo estuviera matando* 'the man screams as if someone were killing him').[1] The study also did not consider idiomatic *como* 'like' and *como si* 'as if' constructions (e.g., Ron Batchelor & Christopher 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* 'like' and *como si* 'as if' 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 Stefanie Wulff 2008: 1 for more detailed discussion of these criteria). The resulting database for Mexican Spanish contains 323 instances of *como* 'like' and 25 instances of *como si* 'as if' constructions including NPs and LOC NPs.

For these data, the study then coded the relevant variables for the 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. These were manually annotated by inspecting each of the *como* 'like' and *como si* 'as if' constructions. Table 2 shows the way the study organized the data. Note that the response variable is CXSCHEMA with

1. Similative 'like' markers followed by clauses (e.g., *he loves you like I do*) are known in the literature as real manner clauses. They depict an action or state identical to that of the main clause (Katja Hetterle 2015: 54). Put another way, they describe the character of a situation comparing it to a real situation. Given that similative and real manner constructions have different semantic functions (Olguín Martínez 2021), the study has decided to disregard real manner constructions from the present study.

its only two levels (i.e., *como si* 'as if' and *como* 'like') because the study has the predictor locative with its two levels (*no* and *yes*).

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

Source	Example	LEMMA	LEMMAwTRANSL	Locative	CXSCHEMA
95 19-03-26 MX Milenio.com	*corre como novato que ataca y defiende*	*corer*	*corer* 'to run'	no	*como*
30 12-09-15 MX Vanguardia.com.mx	*habla como en sueños*	*hablar*	*hablar 'to speak'*	yes	*como*
687 15-10-25 MX La Razon	*se siente como si fuera Navidad*	*sentir*	*sentir 'to feel'*	no	*como si*
8 16-07-27 MX Libertad de Expresión Yucatán	*se ríe como si estu-viera en su fiesta de 15 años*	*reir*	*reir* 'to laugh'	yes	*como si*

Huasteca Nahuatl is a Southern Uto-Aztecan language that belongs to the Uto-Aztecan language family.[2] The data used in this work are drawn from one field work period in the village of Teposteco in the Huasteca region. This is situated in the Municipality of Chicontepec, in the state of Veracruz. This village has 363 inhabitants, and Spanish is used as the main means of instruction in all the different educational levels (Eladio Cruz, pers. comm.). Following the Spanish conquest of Mexico, Spanish was introduced into the linguistic landscape of the Huasteca region, gradually gained more speakers, and is now the dominant and most prestigious language of the region where Huasteca Nahuatl is spoken. This is a clear scenario in which a more prestigious language influences the one with less prestige (Raymond Hickey 2010:7). A number of studies have shown that different domains of Huasteca Nahuatl grammar have been influenced by Spanish. However, it has been demonstrated that one of the common ways in which Spanish has influenced Nahuatl grammar is concerned with clause-linking devices. One of the first 20th century accounts of Spanish influence on Nahuatl comes from Benjamin Lee Whorf (1946:379), who shows that Nahuatl borrowed

2. Because of the geographical distances among Nahuatl speakers, many spoken Nahuatl varieties have arisen. José Antonio Flores Farfán (2010:38) mentions that modern Nahuatl is a set of about 12 varieties with different degrees of mutual intelligibility.

many conjunctions from Spanish: *para* 'in order to,' *porque* 'because,' and *mientras* 'while.' Field (2002: 142–143) notes that the convergence of Huasteca Nahuatl and Spanish is evident in the high frequency of borrowed Spanish function words, such as coordinating conjunction (*o* 'or,' *pero* 'but'), subordinating conjunctions (*hasta* 'until,' *porque* 'because,' *como* 'as, since, how, like'), and complementizers (*que* 'that'). Jane H. Hill and Kenneth C. Hill (1986: 271–272) point out that speakers of Huasteca Nahuatl have not only borrowed adverbial conjunctions from Spanish, but also Spanish relativizing elements including *que* 'that,' de 'that,' *lo que* 'that,' and *donde* 'where,' which have gradually replaced native relativizing strategies. Besides adverbial and coordinating conjunctions, Lyle Campbell (1987) mentions that Nahuatl also borrowed comparative markers from Spanish: *más* 'more'…*que* 'than.'

For the analysis of Huasteca Nahuatl, clearly an underresourced language compared to Mexican Spanish, the corpus consists of 32 stories told by 3 native adult speakers: Mrs. Duarte, Mr. Rodriguez, and Mr. Cruz. The narratives can be divided into different types: fairy tales, personal narratives, and procedural texts.[3] As for the fairy tales, each of them recreates basic aspects of daily life, such as ambition, poverty, hunger, honesty, companionship, love, faith, anger, revenge, sexuality, cunning, among many others. Human beings are the main characters in the fairy tales and can become spiritual beings, legendary beings, and animals. Some of these primary or secondary roles in the fairy tales are also performed by animals or plants personified by means of human qualities and virtues conferred on them by the Nahua speakers. Regarding personal narratives, Mrs. Duarte, Mr. Rodriguez, and Mr. Cruz provided short stories in which they recounted positive or negative experiences. For instance, in one narrative, Mr. Cruz talked about the loss of valued persons. He mentioned how these undesirable events have caused him intense feelings of sadness and hopelessness. The third type of narrative is that of procedural texts, in which speakers provided instructions regarding how to do or make something through a series of steps. In total, the narratives contain 1,032 sentences. Some comments on how the study extracted and annotated the Huasteca Nahuatl data follow here.

The procedure for data retrieval was similar to the one the paper followed for Mexican Spanish similative-pretence constructions. First, the study performed a

3. The study is aware of the fact that the *Corpus del Español NOW* is also a different genre than the Huasteca Nahuatl data (see also below Section 4). However, there are worse genres to compare the Huasteca Nahuatl data to, news data are often used as a stand-in for more varied data in corpus linguistics, and the study has no reason right now to suspect that the genre difference would be not just damaging to our analysis, but also systematically damaging (in the sense of 'unfairly skewing the results in a certain direction'); with under-resourced languages or historical data, differences in genre are omnipresent in corpus linguistics and future studies will show whether our results can be confirmed

search of similative and pretence constructions by searching our corpus for the forms *como* 'like' and *como si* 'as if' (orthographically represented in Huasteca Nahuatl as *komo* 'like' and *komo si* 'as if'). Second, the study excluded *komo* 'like' and *komo si* 'as if' followed by clauses and the study only considered constructions in which these lexical items were followed by NPs and LOC NPs. Third, the study then excluded idiomatic similative and pretence constructions by following the same criteria mentioned above: non-compositionality criterion, formal, degree of conventionalization of a construction, and syntactic flexibility of a construction. By following this process, the study was able to collect data on 283 similative-pretence constructions, 139 instances of *komo* 'like,' and 144 instances of *komo si* 'as if' constructions.

By inspecting each of these constructions, the study then coded the relevant variables for our analysis: (i) verbs that can occur in the first slot of *komo* 'like' and *komo si* 'as if' constructions, (ii) whether the NP following *komo* 'like' and *komo si* 'as if' was locative or non-locative, and (iii) the construction. As can be seen in Table 3, the Huasteca Nahuatl was organized in the same way as the Mexican Spanish data.[4]

Table 3. Organization of the Huasteca Nahuatl similative-pretence data in the present study

Source	Example	LEMMA	LEMMAwTRANSL	Locative	CXSCHEMA
The bunny and the turtle	*hualmotlalok komo si els kuatochi*	*motlalo*	*motlalo* 'to run'	no	*komo si*
The drunk monkeys	*khuelita komo pa mocha*	*huelita*	*huelita* 'to look'	yes	*komo*

4. Note that similative meanings can also be expressed with *igual que* 'like' constructions in Mexican Spanish (e.g., *corre igual que un perro* 'he runs like a dog'). In these patterns, the concept of likeness is fully inferential (Ramón Trujillo 1990). This indicates that these expressions may be derived metonymically or metaphorically in that "they represent fossilized patterns of cognitive processes conventionalized over times" (Wolfgang Schulze 2017: 36). These constructions are not taken into account here given that Huasteca Nahuatl has not borrowed *igual que* 'like' constructions from Mexican Spanish.

2.2 The statistical analysis and results

The study began by computing the no-information baseline that any predictive modeling approach would need to beat, which is the proportion of the more frequent level of the response variable CXSCHEMA; that baseline turned out to be $^{462}/_{(169+462)}$=73.22% (the proportion of *como* 'like' in our data).

The predictive modeling method of CIFs is an extension of the method of decision trees. While the study cannot give a full-fledged introduction of all the complexities of this method in this section, here is at least a bit of an introduction. Decision trees are a recursive partitioning method that aims at identifying structure in the data that is predictive of a response variable (here, CXSCHEMA with its two levels *como si* 'as if' and *como* 'like') and it identifies that structure by iteratively/successively bifurcating (i.e., splitting into two groups) the data into smaller and smaller subsets that have an increasingly imbalanced, and therefore predictive, distribution of the response variable (see Gries 2021: Section 7.1 for details). While trees are often fairly easy to run on data and interpret, they are often not generalizing well to unseen data so scholars often turn to forests of trees to improve predictive performance. Forests in general are an extension of trees that consist of fitting *ntree* [hundreds or even (tens of) thousands of] individual trees to the data, but with two randomization twists:

- every one of the *ntree* trees (here, *ntree* is set to 1500) is fit on a different randomly sampled (with replacement) subset of the original data; this makes the algorithm see many different versions of the actual data and, thus, makes it more likely that the resulting forest will generalize better to new data (rather than overfit on the basis of some idiosyncrasies of the full data set);
- at every split in every tree, the forest algorithm is only offered a randomly-chosen subset of *mtry* predictors (here, the study sets *mtry* to 2); this makes it harder for very strong predictors to always dominate the outcome and, in a sense, also lets weaker predictors have a chance at indicating how they correlate with the response variable.

Forests then generate predictions by amalgamating the predictions of all *ntree* trees in the data, meaning, just like every other predictive modeling method, they return for every case in the original data set (here, the 169 + 462 = 631 cases) a predicted probability of each of the two levels of the response variable and the level with the highest predicted probability is then predicted. The analyst can then compare the observed levels of the response variable with the predictions by the forest to see if the forest was able to identify a good deal of predictive structure in the data, i.e., whether the predictors (here, LANGUAGE, LOCATIVE, and LEMMA) have predictive power with regard to the response.

If a forest has a good degree of predictive power, analysts should explore two more things: (i) variable importance measures (similar to effect sizes) and (ii) some measure of effect directions. The former do exactly what one might think they do: They answer the question 'how important is each variable for the predictive success of the forest?,' i.e., is LANGUAGE more or less important than LOCATIVE? The latter answer the question 'how does a specific variable (level) affect the predictions?,' i.e., does LOCATIVE: no boost or lower the probability of CXSCHEMA: *como si* 'as if'? For the former, the study will use *conditional variable importance scores* from the R package partykit (Torsten Hothorn & Achim Zeileis 2015), for the latter, the study will compute what are called *individual conditional expectations*, i.e., statistics that quantify how much predictions change if, for each case, one predictor is changed from one level to another (see Carolin Strobl et al. 2024).

For the CIF, the study used the R package::function combination partykit::cforest (see Hothorn & Zeileis 2015, version 1.2–23) with the response variable of CXSCHEMA, the predictors of LANGUAGE (*Huasteca Nahuatl* vs. *Mexican Spanish*), LOCATIVE (no vs. yes), and LEMMA (24 different lemmas), and the above values of *ntree* and *mtry*. To assess the quality of the CIF, the study computed the CIF's predictions — both predicted probabilities of each level of CXSCHEMA and categorical constructional choices — and its confusion matrix (see Table 4) together with a variety of metrics that serve to evaluate the CIF's discriminatory power.

Table 4. The confusion matrix of our CIF

	Predicted: *como si* 'as if'	**Predicted: *como* 'like'**	**Sum**
Observed: *como si* 'as if'	**144**	25	169
Observed: *como* 'like'	21	**441**	462
Sum	165	466	631

The CIF comes with a prediction accuracy of $^{144+441}/_{631}$=92.71%, which is highly significantly better than the baseline ($p_{\text{exact binomial test}} < 10^{-35}$); the CIF's *C*-score is 0.96, Cohen's *k* is 0.813 and McFadden's R^2 is 0.635, which all indicate a very good performance.

Both unconditional and conditional variable importance score indicate that LEMMA is the strongest predictor ($importance_{\text{cond}} = 1.136$), followed by LANGUAGE ($importance_{\text{cond}} = 0.386$) and LOCATIVE ($importance_{\text{cond}} = 0.274$).

To also be able to interpret the results from the CIF with regard to our hypothesis (speakers of Huasteca Nahuatl have not only borrowed similative and pretence constructions from Mexican Spanish, but also the same lexical preferences of the first slot of these constructions), the study added two additional

evaluation steps: (i) the above-mentioned assessment of individual conditional expectations and (ii) the identification of constructional prototypes separately for each level of LANGUAGE. Both these methods required a first step of creating a data frame with all theoretically possible 2 ×2×24=96 combinations of LANGUAGE, LOCATIVE, and LEMMA and computing predictions for each of them.

2.2.1 Individual conditional expectations

The method of Individual Conditional Expectations (ICE) is based on the idea of determining for one's data how predictions change if, for each case, one predictor is changed from one level to another; the more widely-known method of partial dependence scores (see Gries 2021, Section 7.2.1) is actually based on averaging ICE scores. The current study specifically computed how much the predicted probability of *como* 'like' changed for each combination of LOCATIVE and LEMMA when the predictor changed from LANGUAGE: *Huasteca Nahuatl*; consider Figure 1, which

- has all verb lemmas on the *x*-axis;
- distinguishes for each lemma the locative use (in red) from the non-locative use (in blue);
- shows for each how much the predicted probability of *como* 'like' changes when LANGUAGE changes from *Huasteca Nahuatl* to *Mexican Spanish*.

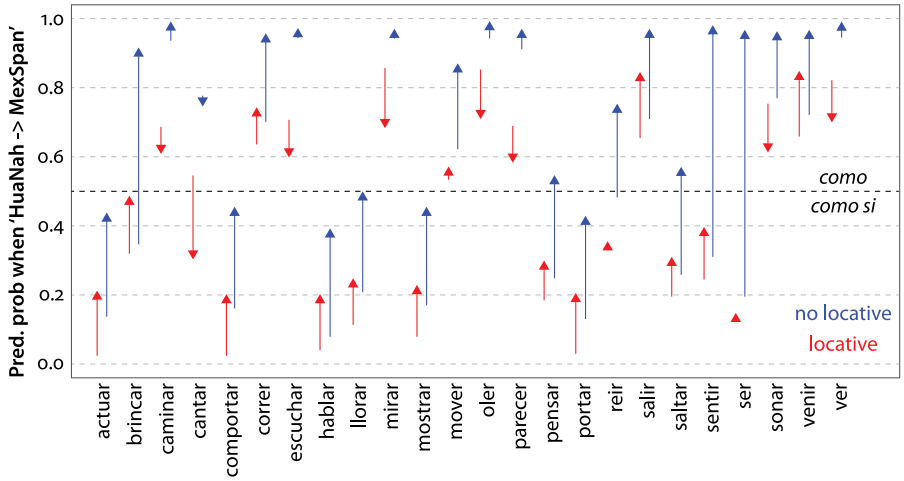

Figure 1. ICE prediction changes for all combinations of LANGUAGE, LOCATIVE, and LEMMA[5]

5. Note that Figure 1 portrays what in a regression modeling context would be the three-way interaction LANGUAGE:LOCATIVE:LEMMA. This is because it represents the predicted probability

As an example, the leftmost (red) arrow represents the facts that, when the verb lemma is *actuar* 'to act' and the use is locative, the CIF predicts:

- a probability of *como* 'like' of 2.5% for Mexican Spanish, but
- a probability of *komo* 'like' of 21.3% for Huasteca Nahuatl.

However, the (blue) arrow means that, when the verb lemma is still *actuar* 'to act' but the use is now non-locative, the CIF predicts:

- a probability of *como* 'like' of 13.8% for Mexican Spanish, but
- a probability of *komo* 'like' of 43.7% for Huasteca Nahuatl.

This plot contains a lot of information at multiple levels of granularity, but two very general and straightforward observations are the following: First, most arrows go up, which is the CIF's/plot's way of saying that, on the whole, the CIF sees structure in the data that associates Mexican Spanish with *como* 'like' (as opposed to *como si* 'as if'): In 39 out of 48 cases, the switching from Huasteca Nahuatl to Mexican Spanish makes *como* 'like' more, not less, likely; we will discuss individual verbs/verb classes in more detail below (Section 3.1).

Second, most arrows do not cross the horizontal $y = 0.5$ line where the predicted probability of *como* 'like' increases to above 50%, i.e., where the prediction of the CIF changes from *como si* 'as if' to *como* 'like' — which means, most verb lemmas and locative vs. non-locative differences come with different degrees of the same preference — but in the following lemma-locative combinations, Huasteca Nahuatl and Mexican Spanish do not just differ in degree of preference, but in actual preference:

- changing the language from Huasteca Nahuatl to Mexican Spanish changes the prediction from *como si* 'as if' to *como* 'like' only when there is no locative, namely for *brincar* 'to jump' (see, e.g., the second blue log arrow from the left pointing upwards), *llorar* 'to cry', *pensar* 'to think,' *saltar* 'to jump,' *sentir* 'to feel,' and *ser* 'to be,' but never with locatives;
- changing the language from Huasteca Nahuatl to Mexican Spanish changes the prediction from *como* 'like' to *como si* 'as if' only for *cantar* 'to sing' with locatives, never anywhere else.

of *como* 'like' for each combination of the three predictors and, therefore, also allows the analyst to see how the preferences of combinations of LOCATIVE:LEMMA vary in the two languages, as the subsequent discussion will illustrate in more detail; this procedure is preferable over doing separate analyses for each language (see Gries 2021: 314–318).

2.2.2 Prototypes

In addition to the above, alternation contexts often benefit from the identification and interpretation of the prototypes of the alternants. The general definition of prototype is the following: The prototype P of a category C is the abstract combination of features $F_{1\text{-}n}$ that have the highest cue validity for C, where (i) "abstract combination of features" means that they might actually never all come together in real life — they are possible hypotheticals — and (ii) and cue validity can in turn be defined as follows: A feature F_x (e.g., 'having a beak') has a high cue validity for a category C (e.g., 'birds') if many/most members of the category have the feature and most/many non-members of C do not have F. For this example, most birds (C) have beaks (F) and most non-birds do not have beaks (animals that are not birds but have beaks make up a small set of some turtles, octopuses, platypuses/echidnas, and pufferfish).

The study can extend this kind of logic to alternation scenarios following the logic of Gries (2003) and operationalize a prototype of a(n alternation) category as the abstract combination of features that leads to the highest predicted probability of a category. In this case, that means identifying the combinations of verb lemmas and locative absences/presences that lead to:

- the highest predicted probabilities of *komo* 'like' for Huasteca Nahuatl;
- the highest predicted probabilities of *komo si* 'as if' for Huasteca Nahuatl;
- the highest predicted probabilities of *como* 'like' for Mexican Spanish;
- the highest predicted probabilities of *como si* 'as if' for Mexican Spanish.

The strongest predictions of *komo* 'like' for Huasteca Nahuatl (all with predicted $p \geq 0.9$) all arise only with non-locatives and the following verbs (in descending order of prototypicality for *komo* 'like'): *ka* 'to hear,' *nekui* 'to smell,' *ita* 'to see,' *huelita* 'to look,' and *ipan* 'to seem'; note that these are actually also all attested in the data. By contrast, the strongest predictions of *komo si* 'as if' for Huasteca Nahuatl (all with predicted $p \geq 0.9$) arise — with one exception — only with locatives and the following verbs (in descending order of prototypicality for *komo si* 'as if'): *ixehua* 'to behave,' *yolpoloa* 'to act,' *yelia* 'to behave,' *kamohui* 'to speak,' *kamohui* 'to speak' (without locatives), and *nemiyotl* 'to show'; here, the combinations with *kamohui* 'to speak'+locative and *nemiyotl* 'to show' are only hypothetical.

The strongest predictions of *como* 'like' for Mexican Spanish (all with predicted $p \geq 0.9$) all arise only with non-locatives and the following verbs (in descending order of prototypicality for *como* 'like'): *caminar* 'to walk,' *oler* 'to smell,' *ver* 'to see,' *sentir* 'to feel,' *escuchar* 'to hear,' *mirar* 'to look,' *venir* 'to come,' *salir* 'to leave,' *parecer* 'to seem,' *ser* 'to be,' *correr* 'to run,' *sonar* 'to sound,' and

brincar 'to jump,' where the verb lemmas *escuchar* 'to hear,' *mirar* 'to look,' *parecer* 'to seem,' and *sonar* 'to sound' are hypotheticals. By contrast, the strongest predictions of *como si* 'as if' for Mexican Spanish (all with predicted $p \geq 0.8$) arise with locatives and the verbs *ser* 'to be' and *comportar* 'to behave' (unattested) (in descending order of prototypicality for *como si* 'as if').

3. Discussion

3.1 ICE discussion

The most interesting ICE results (Figure 1) can be summarized in the following way. First, mistaken identity verbs, such as *actuar/yolpoloa* 'to act,' *comportar/ixehua* 'to behave,' *hablar/kamohui* 'to speak,' *mostrar/nemiyotl* 'to show,' and *portar/yelia* 'to behave' always prefer pretence *como si* 'as if,' with locatives and without them in Mexican Spanish and in Huasteca Nahuatl, as in (5) and (6):

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

b. *como si* 'as if' construction with LOC NP
se comporta como si estuviera en su casa.
PRON 3SG.behave.PRS as if 3SG.be.SUBJ LOC 3SG.POSS house
'He behaves as if he were at his house.'
(191 18-10-10 MX El Imparcial de Oaxaca)

(6) a. *komo si* 'as if' construction with NP
ixehua-k komo si el-s se tsopilo-tl
behave-PFV as if be-IRR a vulture-ABS
'He behaved as if he were a vulture.' (The storm story-07/16/2022)

b. *komo si* 'as if' construction with LOC NP
yolpoloa-k komo si el-s pa i-cha.
act-PFV as if be-IRR LOC 3SG.POSS-house
'He acted as if he were at his house.'
(The sun and the moon story-07/17/2022)

This indicates that speakers of Huasteca Nahuatl borrowed not only the Mexican Spanish clause-linking device *como si* 'as if,' but also the use of mistaken identity predicates in locative and non-locative constructions. This aligns with Hypothesis 1 in that we predicted that mistaken identity verb lemmas will occur equally frequent in both locative and non-locative pretence constructions in Huasteca

Nahuatl and Mexican Spanish due to intense language contact of Huasteca Nahuatl with Mexican Spanish, a hypothesis that was grounded in the fact that the meaning of Mexican Spanish pretence *como si* 'as if' constructions is that of 'to imitate, pretend, aspire to the behavior of something/someone' and that, accordingly, the meaning of mistaken identity verbs harmonizes with the semantics of this construction. From an ecological competition model (Mark Lindsay & Mark Aronoff 2013), this lexico-syntactic transference from Mexican Spanish to Huasteca Nahuatl is reasonable given that Huasteca Nahuatl does not contain a native construction used to express pretence. In the literature, it has been shown that the internal structure of recipient languages may play a role in facilitating the borrowing of clauses-linking devices (e.g., conjunctions) from a donor language. The absence of explicit ways of expressing a semantic relation in a recipient language may have provided a niche for the newly interpreted linking devices to fill (Marianne Mithun 1992: 126) along with the use of other constructional properties.

Second, the epistemic verbs *escuchar/ka* 'to hear,' *mirar/huelita* 'to look,' *oler/nekui* 'to smell,' *parecer/ipan* 'to seem,' *sonar/tsatsi* 'to sound,' and *ver/ita* 'to look' always prefer similative *como* 'like' constructions, with locative and without them, in Mexican Spanish [see (7)] and Huasteca Nahuatl [see (8)]. Some of these verb lemmas can be characterized as perception verbs (e.g., *escuchar/kaki* 'to hear). It has been shown that the 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 the experience and understanding of the world and metaphorical mappings (e.g., the frequent metaphorical mappings of understanding is seeing; obeying is hearing; conserving is touching; suspecting is smelling; see Iraide Ibarretxe-Antuñano 1999; George Lakoff & Mark Johnson 1980).

(7) a. *como* 'like' construction with NP
suena como una buena idea.
3SG.sound.PRS like a good idea
'It sounds like a good idea.' (1000 19-02-11 MX TekCrispy)

b. *como* 'like' construction with LOC NP
huele como en un estadio de futbol.
3SG.smell.PRS like LOC a stadium of soccer
'It smells like (we were) in a soccer stadium.'
(41 18-12-03 MX Diario de Querétaro)

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

b. *komo* 'like' construction with LOC NP
ki-ka komo pa ne kuatitla tepeti-pa.
3SG.OBJ-hear like LOC DEM woodland hill-LOC
'It sounds like (we were) at a hill.' (My work story-07/15/2022)

However, as can be seen in Figure 1, the red arrow indicates that with locatives, Mexican Spanish prefers the epistemic verb lemmas mentioned above in similative *como* 'like' constructions more than Huasteca Nahuatl (the red arrow points downwards). On the other hand, the blue arrow shows that without locatives, Mexican Spanish prefers the epistemic verb lemmas mentioned above in similative *como* 'like' constructions less than Huasteca Nahuatl (the blue arrow points up, mostly just a bit, but more notably with *sonar* 'to sound'). This does not align with Hypothesis 2 in that we predicted that epistemic verb lemmas will occur equally frequent in both locative and non-locative similative constructions in Huasteca Nahuatl and Mexican Spanish due to intense language contact of Huasteca Nahuatl with Mexican Spanish, a hypothesis that was grounded in the fact that the meaning of Mexican Spanish similative *como* 'like' constructions is that of 'to give the same appearance as something/someone' and that, accordingly, the meaning of epistemic verb lemmas would harmonize with the meaning of this construction.

The question is: Why are Huasteca Nahuatl epistemic *komo* 'like' constructions with locatives in (8b) not as frequent as in Mexican Spanish? From an ecological competition model, we hypothesize that this is the result of constructional rivalry (Benoît Leclercq & Cameron Morin 2023: 11–12). In Huasteca Nahuatl, epistemic *komo* 'like' meanings with locatives can also be expressed with constructions like in (9):

(9) *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)

The example in (9) should be understood as: lit. 'it feels like being in this place reaches the same feeling as that of being in his house.' In this construction, the verb *temanti* 'to reach' functions in a similar way as *komo* 'like.' The use of this motion verb in the expression of similative meanings can be interpreted as being part of a more general process whereby languages use a spatial metaphor (sometimes called fictitious motion) to refer, not to the motion of an agent, but to the (metaphorical) motion in time of a situation.

Formal approaches to language contact have accounted for the use of one construction over the other as a blocking effect (e.g., Ian Roberts & Anna Roussou 2003). Put another way, in language contact situations, the avoidance of grammatical doublets has been attributed to a blocking effect (Anthony Kroch 1994) in that two constructions with similar functions (i.e., a native and a borrowed construction) cannot co-exist in a recipient language (i.e., it is important identifying one form as the definite winner of the competition and the other as the loser). However, from a Usage-Based CxG perspective, this will not always be the case. It is well-known that in many recipient languages around the world, constructions with similar functions (i.e., a native and a borrowed construction) can co-exist in their grammars. Due to their competition, one of them may develop additional functions. For instance, it has been shown that many South American and Mesoamerican languages have borrowed the Spanish clause-linking device *pero* 'but' for expressing contrast (e.g., *the man is short, but the kid is short*) although they already contain a native way of expressing this meaning (Christel Stolz & Thomas Stolz 1996a: 100, 1996b: 269). In this scenario, the form of a construction, along with their function(s) and context(s) of use may be replicated in the target language relatively intact. Interestingly, in a number of South American and Mesoamerican languages, the loanword *pero* 'but' has developed other functions besides those attested in Spanish. In Mosetén (isolate) *pero* 'but' expresses not only contrast but also a change in topic (Jeanette Sakel 2007: 570), a phenomenon known as discourse widening (Mithun 2025). Accordingly, constructional rivalry does not necessarily give rise to a winner or a loser from a Usage-Based CxG perspective.

3.2 Prototype discussion

Now turn the attention to the prototype results presented in Section 2.2.2. This analysis is important for the following reason. As is shown in this section, Mexican Spanish contains prototypical similative constructions that do not appear with epistemic verb lemmas. Accordingly, the question is: Does Huasteca Nahuatl contain similative prototypes that do not occur with epistemic verb lemmas?

As for Mexican *como* 'like' constructions, all prototypical *como* 'like' constructions are non-locatives and contain not only epistemic verb lemmas: *oler* 'to smell,' *ver* 'to see,' *sentir* 'to feel,' *escuchar* 'to hear,' *mirar* 'to look,' *parecer* 'to seem,' *sonar* 'to sound,' as in (10), but also non-epistemic verb lemmas: *caminar* ' to walk,' *venir* 'to come,' *salir* 'to leave,' *correr* 'to run,' and *brincar* 'to jump,' as in (11). In a similar fashion, all prototypical Huasteca Nahuatl similatives are non-locatives. However, unlike Mexican Spanish, they only contain epistemic verb

lemmas: *kaki* 'to hear,' *nekui* 'to smell,' *ita* 'to see,' *huelita* 'to look,' and *ipan* 'to seem,' as in (12) (except for the non-epistemic verb lemma *nenemi* 'to walk').

(10) *se ve como un auto nuevo.*
PRON 3SG.see.PRS like a car new
'It looks like a new car.' (2791 17-08-05 MX Milenio.com)

(11) *salió como un vagabundo.*
3SG.leave.PST like a tramp
'He left like a tramp.' (2571 17-10-17 MX IGN Latinoamérica)

(12) *ki-ipan komo ueno tekita.*
3SG.OBJ-seem like good job
'It seems like a good job.' (The stupid man story-07/15/2022)

For those verb lemmas that are prototypical in non-locative similative constructions in both languages: *oler*/*nekui* 'to smell' and *ver*/*ita* 'to see,' this is not surprising given that they are epistemic and harmonize with the semantics of similative constructions. However, *caminar*/*nenemi* 'to walk' is an unexpected verb lemma in that it does not seem to align with the semantics of similative constructions, as in Examples (13) and (14) from Mexican Spanish and Huasteca Nahuatl:

(13) *camina como alguien que tuviera dinero.*
3SG.walk.PRS like someone REL 3SG.have.SUBJ money
'He walks like someone who has money.' (641 16-10-01 MX Sipse.com)

(14) *nenemi-ki komo toto-tl.*
walk-PFV like bird-ABS
'He walked like a bird.' (The animals of the forest story-07/15/2022)

Prototypical Mexican Spanish non-locative similatives involving epistemic verb lemmas can be characterized as semantically transparent in that epistemic verb lemmas clearly align with the semantics of this construction. On the other hand, non-epistemic verb lemmas in prototypical Mexican Spanish non-locative similatives can be characterized as semantically opaque. As was shown above, prototypical Huasteca Nahuatl non-locative similatives only arise in semantically transparent scenarios (e.g., epistemic verb lemmas except for *caminar* 'to walk'). From a theoretical perspective, this is expected given that languages emerging from language contact have shown to only borrow semantically transparent patterns rather than semantically opaque patterns (Sterre Leufkens 2013).

The discussion now turns to the *como si* 'as if' construction prototypes. In Mexican Spanish, all these prototypes are locatives and only contain mistaken identity verbs: *ser* 'to be' and *comportar* 'to behave,' as in (15). In a similar fashion,

all prototypical Huasteca Nahuatl *komo si* 'as if' constructions are locatives and only arise with mistaken identity verbs (except for *kamohui* 'to speak' that may occur with locative and non-locatives). Besides *kamohui* 'to speak,' other mistaken identity verbs that appear in prototypical *komo si* 'as if' constructions are *ixehua* 'to behave,' *yolpoloa* 'to act,' *yelia* 'to behave,' and *nemiyotl* 'to show,' as in (16).

(15) *se comporta como si estuviera en la cárcel.*
PRON 3SG.behave.PRS as if 3SG.be.SUBJ LOC DET jail
'She behaves as if she were at jail.' (600 12-03-06 MX El Siglo de Torreón)

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

The pretence prototypes in both Mexican Spanish and Huasteca Nahuatl only arise with mistaken identity verb lemmas and locatives. Accordingly, prototypes in both languages reflect semantic transparency. However, Huasteca Nahuatl has developed a more complex repertoire of pretence prototypes involving semantic transparency than Mexican Spanish. From a theoretical perspective, the fact that a recipient language (Huasteca Nahuatl) has developed more prototypes involving semantic transparency than a donor language (Mexican Spanish) is surprising and goes against the claim that contact simplifies the grammar of a recipient language (e.g., Wouter Kusters 2008). As argued by Peter Trudgill (2009: 99), this is "due to the relative inability of adult humans to learn new languages perfectly." In the process of learning a language, adults simplify its grammar by, for example, eliminating verbs that express more or less the same meaning. Accordingly, the fact that Huasteca Nahuatl pretence constructions contain more semantically transparent prototypes than Mexican Spanish is unexpected.

3.3 Implications

The findings of the present study have different methodological and theoretical implications to language contact studies.

It has been shown that one of the grammatical categories that is by far the most susceptible to being borrowed is that of clause-linking devices (Yaron Matras 2007: 54). For instance, many American languages, particularly those indigenous to Middle and South America, have borrowed clause-linking devices directly from Spanish or Portuguese (Mithun 2012: 6). Turkic languages have also received a great deal of attention in that that they have borrowed conjunctions, particularly from Arabic and Persian (Anthony P. Grant 2012: 314). Many Siberian languages have borrowed conjunctions from Russian, which has in turn gradually led to

a partial loss of their native morphological means (Gregory D.S. Anderson 2005: 223). Several Highland East Cushitic languages in Southwest Ethiopia have borrowed switch-reference converbs from neighboring North Omotic languages (Yvonne Treis 2012: 80). However, most of these studies, if not all, have only focused on clause-linking devices without paying attention to other constructional properties (e.g., verb lemmas appearing in dependent and main clauses). Theoretically, the present study has shown that the rigid partition of structural levels (lexicon, syntax, morphology, etc.) that has been adopted by traditional models of language contact (e.g., Sarah Grey Thomason & Terrence Kaufman 1988) proves inadequate for describing complex language situations in which different levels interact with one another in language use. This is in line with recent Usage-Based CxG approaches that have insisted on the need for an integrative, non-modular approach to language contact (e.g., Boas & Höder 2018).

From a theoretical perspective, our study has also important implications to pattern and matter replication studies. In the literature, it has been shown that speakers of recipient languages may replicate grammatical forms from a donor language with native material. This is known as pattern replication (Sakel 2007). In this scenario, only the patterns of the other language are replicated, i.e., the organization, distribution, and mapping of grammatical or semantic meaning, while the form itself is not borrowed (Matras 2007). Put another way, no phonetic substance is involved. On the other hand, there are cases in which speakers of recipient languages may borrow grammatical markers with their exact forms from a donor language (there may be minor differences in substance, as speakers of recipient languages integrate the sounds into their native phonological systems). This process is known as matter replication (Sakel 2007). While previous studies have demonstrated that recipient languages may either show matter or pattern replication, the present research has shed light on the fact that in lexico-syntactic transference, matter and pattern replication can co-exist in the same construction. In Huasteca Nahuatl, the verb lemmas that occur in similative and pretence constructions are the result of pattern replication and the similative and pretence markers are the result of matter replication.

Methodologically, the present study has developed a method that can not only be used for exploring lexico-syntax transference in language contact situations, but also can be used, by other linguists, to investigate other possible combinations of structural levels, e.g., semantico-syntactic transference, syntactic-pragmatic transference (Clyne 2003).

4. Concluding remarks

Using the method of CIFs, this paper has demonstrated that speakers of Huasteca Nahuatl have not only borrowed similative and pretence constructions from Mexican Spanish, but also lexical preferences of the first slot of these constructions. Based on two evaluation steps: (i) an assessment of ICE and (ii) the identification of constructional prototypes separately for Mexican Spanish and Huasteca Nahuatl, we were able to provide a novel and quantitatively rigorous way to explore how constructional templates and their lexical preferences diffuse in language contact situations.

First, the ICE results show that speakers of Huasteca Nahuatl borrowed not only the Mexican Spanish clause-linking device *como si* 'as if,' but also the use of mistaken identity predicates (e.g., *act, behave*) in locative and non-locative constructions. The study argues that this lexico-syntactic transference from Mexican Spanish to Huasteca Nahuatl makes sense given that Huasteca Nahuatl does not contain a native construction used for expressing pretence. As for similative constructions, the ICE results indicate that speakers of Huasteca Nahuatl borrowed the Mexican marker *komo* 'like' and the use of epistemic verb lemmas (e.g., *seem, look*) in both locative and non-locative instances. However, the results also shed light on the fact that Huasteca Nahuatl epistemic *komo* 'like' constructions with locatives are not as frequent as in Mexican Spanish. The study explains that this may be the result of constructional rivalry in that epistemic *komo* 'like' meanings with locatives can also be expressed with another construction in Huasteca Nahuatl.

Second, the prototypes identified in Mexican Spanish and Huasteca Nahuatl also shed light on other important details regarding the language contact situation between Mexican Spanish and Huasteca Nahuatl. Mexican Spanish contains prototypical non-locative similatives involving not only epistemic verb lemmas (semantically transparent), but also non-epistemic verb lemmas (semantically opaque). On the other hand, Huasteca Nahuatl only contains prototypical non-locative similatives with epistemic verb lemmas (except for *caminar* 'to walk'). This aligns with previous studies that have shown that recipient languages tend to borrow semantically transparent patterns rather than semantically opaque patterns (Leufkens 2013). As for pretence constructions, prototypes in both Mexican Spanish and Huasteca Nahuatl only arise with mistaken identity verb lemmas and locatives. Accordingly, prototypes in both languages reflect semantic transparency. Interestingly, Huasteca Nahuatl has developed a more complex repertoire of pretence prototypes involving semantic transparency than Mexican Spanish. This goes against previous studies that have proposed that language contact simplifies the grammar of a recipient language (e.g., Kusters 2008).

Where to go from here, what are the next steps? The present study has restricted the attention to lexico-syntactic transference. However, it remains to be explored whether speakers of Huasteca Nahuatl have also borrowed other constructional properties from Mexican Spanish similative-pretence similative constructions. For instance, similative and pretence constructions must be characterized as a subtype of comparative construction. They are qualitative comparative constructions in that "they bring together two terms of the comparison on the basis of similarity or likeness" (Fuchs 2014: 133). Similative and pretence constructions contain two NPs, one of which is the object of comparison (locative and non-locative NP), while the other functions as the yard-stick of the comparison. The questions are: How does the semantics of these NPs interact with the verb lemmas of similative and pretence constructions? And, is it possible to identify preferred and dispreferred co-occurrences of these features in Mexican Spanish and Huastec Nahuatl? Methodologically, how can the study explore the diffusion this domain? In the recent Usage-Based CXG, different studies have made inroads into the analysis preferred and dispreferred co-occurrences of features (types and antitypes, respectively) in individual languages (e.g., Olguín Martínez & Gries 2024), but not with respect to how they diffuse through language contact. This is a very promising for future research.

Finally, the study again acknowledges the potential for corpus/genre mismatches arising from the data availability situation for both Mexican Spanish and Huasteca Nahuatl. While the approach of using news data is a frequent one in corpus studies and while we have confidence in our results, it goes without saying that, ultimately, replication will be necessary, which will hopefully also help make the methodological approach promoted here more widespread in studies of language contact from an integrative, non-modular perspective.

Abbreviations

1	first person
2	second person
3	third person
ABS	absolutive
DEF	definite
DEM	demonstrative
DET	determiner
DIR	directional
IRR	irrealis
LOC	locative
PFV	perfective
PL	plural
POSS	possessive
PRON	pronominal
PRS	present
SBJ	subject
SG	singular
SUBJ	subjunctive

References

Anderson, G. D. S. (2005). *Language contact in South Central Siberia*. Wiesbaden: Harrassowitz Verlag.

Batchelor, R., & Pountain, C. J. (2005). *Using Spanish: A guide to contemporary usage (2nd edtion)*. Cambridge: Cambridge University Press.

Béchet, C. (2020). An empirical perspective on the contact between English and French: A case study on substitutive complex prepositions. *Linguistics Vanguard*, *6*(2), Article 20180051.

Boas, H. C., & Höder, S. (2018). *Constructions in contact: Constructional perspectives on contact phenomena in Germanic languages*. Amsterdam: John Benjamins.

Boas, H. C. & Höder, S. (2021). Widening the scope: Recent trends in constructional contact linguistics. In H. C. Boas & S. Höder (Eds.), *Constructions in contact 2: Language change, multilingual practices, and additional language acquisition* (pp. 1–13). Amsterdam: John Benjamins.

Bullock, B. E., Serigos, J., & Toribio, A. J. (2021). Exploring a loan translation and its consequences in an oral bilingual corpus. *Journal of Language Contact*, *13*(3), 612–635.

Campbell, L. (1987). Syntactic change in Pipil. *International Journal of American Linguistics*, *53*(3), 253–280.

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

Deshors, S. C., & Gries, S. Th. (2016). Profiling verb complementation constructions across New Englishes: A two-step random forests analysis to *ing* vs. *to* complements. *International Journal of Corpus Linguistics*, *21*(2), 192–218.

Field, F. W. (2002). *Linguistic borrowing in bilingual contexts*. Amsterdam: John Benjamins.

Flores Farfán, J. A. (2010). Sociolinguistics in Mexico: Defining new agendas. In M. Ball (Ed.), *The Routledge handbook of sociolinguistics around the world* (pp. 34–41). London: Routledge.

Fuchs, C. (2014). *La comparaison et son expression en français* [*Comparison and its expression in French*]. Paris: Ophrys.

Grant, A. P. (2012). Contact, convergence, and conjunctions: A cross-linguistic study of borrowing correlations among certain kinds of discourse, phasal adverbial, and dependent clause markers. In C. Chamoreau & I. Léglise (Eds.), *Dynamics of contact-induced language change* (pp. 311–358). Berlin: Mouton de Gruyter.

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

Gries, S. Th. (2021). *Statistics for linguistics with R: A practical introduction (3rd revised and extended edition)*. Berlin: De Gruyter Mouton.

Gries, S. Th., & Hilpert, M. (2008). The identification of stages in diachronic data: Variability-based neighbour clustering. *Corpora*, *3*(1), 59–81.

Gries, S. Th., & Adelman, A. S. (2014). Subject realization in Japanese conversation by native and non-native speakers: Exemplifying a new paradigm for learner corpus research. In J. Romero-Trillo (Ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New empirical and theoretical paradigms* (pp. 35–54). Cham: Springer.

Hakimov, N., & Backus, A. (2021). Usage-based contact linguistics: Effects of frequency and similarity in language contact. *Journal of Language Contact*, *13*(3), 459–481.

Hernandez-Mena, C. (2019). *TEDx Spanish corpus. Audio and transcripts in Spanish taken from the TEDx Talks*. Mexico: Universidad Nacional Autonoma de Mexico.

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

Hickey, R. (2010). *The handbook of language contact*. Malden: Wiley-Blackwell.

Hill, J. H., & Hill, K. C. (1986). *Speaking Mexicano: Dynamics of syncretic language in Central Mexico*. Tucson: The University of Arizona Press.

Hilpert, M. (2006). Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory*, *2*, 243–256.

Hilpert, M., & Östman, J.-O. (Eds.). (2016). *Constructions across grammars*. Amsterdam: John Benjamins.

Höder, S. (2012). Multilingual constructions: A diasystematic approach to common structures. In K. Braunmüller, & C. Gabriel (Eds.), *Multilingual individuals and multilingual societies* (pp. 241–257). Amsterdam: John Benjamins.

Höder, S. (2014). Constructing diasystems: Grammatical organisation in bilingual groups. In T. A. Åfarli & B. Mæhlum (Eds.), *The sociolinguistics of grammar* (pp. 137–152). Amsterdam: John Benjamins.

Hothorn, T., & Zeileis, A. (2015). Partykit: A modular toolkit for recursive partytioning in R. *The Journal of Machine Learning Research*, *16*(1), 3905–3909.

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

Kroch, A. (1994). Morphosyntactic variation. In K. Beals (Ed.), *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory* (pp. 180–201). Chicago: Chicago Linguistics Society.

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

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

Lara, L. F., Medina Urrea, A., Rosales Martínez, A., Diez Sánchez, C. F., & Serralde Galicia, J. L. (2018). *El Corpus del español mexicano contemporáneo*. https://cemcii.colmex.mx/

Leclercq, B., & Morin, C. (2023). No equivalence: A new principle of no synonymy. *Constructions*, *15*(1), 1–16.

Leufkens, S. (2013). The transparency of creoles. *Journal of Pidgin and Creole Languages*, *28*(2), 323–362.

Lindsay, M., & Aronoff, M. (2013). Natural selection in self-organizing morphological systems. In N. Hathout, F. Montermini, & J. Tseng (Eds.), *Morphology in Toulouse: Selected Proceedings of Décembrettes 7* (pp. 133–153). Munich: Lincom.

Matras, Y. (2007). The borrowability of structural categories. In Y. Matras & J. Sakel (Eds.), *Grammatical borrowing in cross-linguistic perspective* (pp. 31–73). Berlin: Mouton de Gruyter.

Mithun, M. (1992). External triggers and internal guidance in syntactic development: Coordinating conjunction. In M. Gerritsen & D. Stein (Eds.), *Internal and external factors in syntactic change* (pp. 89–130). Berlin: Mouton De Gruyter.

Mithun, M. (2012). Exuberant complexity: The interplay of morphology, syntax, and prosody in Central Alaskan Yup'ik. *Linguistic Discovery*, *10*(1), 5–26.

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

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

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

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

Olguín Martínez, J., & Gries, S. Th. (2024). *If not for-if it weren't/wasn't for* counterfactual constructions: A multivariate extension of collostructional analysis. *Cognitive Semantics*, *10*, 158–189.

Olguín Martínez, J., & Gries, S. Th. (2025). The similative-pretence alternating pair and filler-slot relations: A revised version of distinctive collexeme analysis. *Constructions and Frames*, *17*(1), 65–91.

Olko, J. (2020). Nahuas and Spaniards in contact: Cross-cultural transfer as seen through the Nahuatl lexicon. In A. Brylak, J. Madajczak, J. Olko, & J. Sullivan (Eds.), *Loans in colonial and modern Nahuatl: A contextual dictionary*, (pp. 10–58). Berlin: Mouton de Gruyter.

Olko, J., Borges, R., & Sullivan, J. (2018). Convergence as the driving force of typological change in Nahuatl. *STUF-Language Typology and Universals*, *71*(3), 467–507.

Roberts, I., & Roussou, A. (2003). *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.

Royo Viñuales, V., & Van linden, A. (2024). Beyond hypothetical manner: A functional typology of insubordinate *como si*-clauses. *Folia Linguistica*.

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

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). Amsterdam: John Benjamins.

Sommerer, L., & Smirnova, E. (Eds.) (2020). *Nodes and networks in diachronic Construction Grammar*. Amsterdam: John Benjamins.

Stolz, C., & Stolz, T. (1996a). Funktionswortentlehnung in Mesoamerika: Spanisch-amerindischer Sprachkontakt [Function word borrowing in Mesoamerica: Spanish-Amerindian language contact]. *STUF-Language Typology and Universals*, *49*(1), 86–123.

Stolz, C., & Stolz, T. (1996b). Transpazifische Entlehnungsisoglossen: Hispanismen in Funktionswortinventaren beiderseits der Datumsgrenze [Transpacific borrowing isoglosses: Hispanisms in function word inventories on both sides]. In N. Boretzky, W. Enninger, & T. Stolz (Eds.), *Areale, Kontakte, Dialekte. Sprache und ihre Dynamik in mehrsprachigen Situationen: Beiträge zum 10. Bochum-Essener-Symposium vom 30. 06.-01.07.1995 an der Universität GH Essen* (pp. 262–291). Bochum: Universitätsverlag Brockmeyer.

Strobl, C., Rothacher, Y., Theiler, S., & Henninger, M. (2024). Detecting interactions with random forests: A comment on Gries' words of caution and suggestions for improvement. *Corpus Linguistics and Linguistic Theory*.

Thomason, S.G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Traugott, E.C. (2003). Constructions in grammaticalization. In B.D. Joseph & R.D. Janda (Eds.), *The handbook of historical linguistics* (pp. 624–647). Oxford: Blackwell.

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

Treis, Y. (2012). Switch-reference and Omotic-Cuhistic language contact in Southwest Ethiopia. *Journal of Language Contact*, 5(1), 80–116.

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

Trujillo, R. (1990). Sobre la explicación de algunas construcciones de 'como' [An explanation regarding constructions with como 'like']. *Verba*, (17), 249–266.

Whorf, B.L. (1946). The Milpa Alta dialect of Aztec (with notes on the Classical and the Tepoztlan dialects). In H. Hoijer (Ed.), *Linguistic structures of Native America* (pp. 367–397). NewYork: Viking Fund Foundation.

Wiesinger, E. (2021). The Spanish verb-particle construction [V *para atrás*]: Disentangling constructional contact and change. In H.C. Boas & S. Höder (Eds.), *Constructions in Contact 2: Language change, multilingual practices, and additional language acquisition* (pp. 139–187). Amsterdam: John Benjamins.

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

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

Zenner, E., Backus, A., & Winter-Froemel, E. (2019). *Cognitive contact linguistics: Placing usage, meaning and mind at the core of contact-induced variation and change*. Berlin: Mouton de Gruyter.

Address for correspondence

Jesús Olguín Martínez
Illinois State University
Campus Box 4240
Normal, IL 61790-4240
US
jfolgui@ilstu.edu
https://orcid.org/0000-0002-4555-4213

Co-author information

Stefan Th. Gries
Justus-Liebig University Giessen
stgries@linguistics.ucsb.edu
https://orcid.org/0000-0002-6497-3958

Publication history

Date received: 4 March 2025
Date accepted: 6 June 2025