Here, there and everywhere: An experimental investigation of the semantic features of indexicals*

Mora Maldonado and Jennifer Culbertson

October 1, 2020

Abstract

Languages use indexical expressions to refer to properties that are dependent on the context of speech. Person indexicals (e.g., 'I'), for example, refer to the participants in the conversation, whereas spatial indexicals (e.g., 'here') specify spaces in relation to the location of the utterance. In this paper, we focus on deixis of person and space, and adopt a learning approach to investigate whether these two indexical categories have a common semantic basis. We present three artificial language experiments that test whether learners assume that new partitions of person and spatial indexicals share the same set of features. Our results suggest that learners tend to generalize a non-native person contrast from a pronominal system to a locative system, but do not extrapolate to the same extent in the opposite direction. Moreover, learners are sensitive to the naturalness of these systems, generalizing only when patterns of homophony between the forms in one system are feature-based, not random.

1 Introduction

Languages have strategies to categorize entities, spaces, times and objects as a function of their relation to the speech context. Personal pronouns, for example, serve to refer to individuals as a function of their role in the context of speech: there is the speaker, the addressee and others, who play no active role in the conversation. Similarly, locative expressions can be thought as ways of categorizing spatial positions relative to the speech act (among other things): for example, the English adverb 'here' can be used to refer to the location where the utterance takes place. These lexical items, whose meaning depends on a speech context, are known as *indexicals*.

In this paper, we focus on personal and spatial deixis; that is, indexical expressions that refer to participants and spaces relevant to the utterance. As we explain in §1.1, person and spatial indexicals have been argued to have a common semantic basis, being defined by (roughly) the same primitives. There is however some disagreement about whether both sets of indexicals should be unified in this way: in §1.2, we review evidence from languages where person and spatial indexicals seem to require different semantic primitives. As we spell out in §1.3, in this work we adopt an artificial language learning approach to tackle this question, investigating whether learners treat these two indexical categories as involving the same semantic primitives.

*Acknowledgments to be added.

1.1 A single basis for locative and person systems

Indexicals have been extensively studied in different traditions in linguistics and philosophy, but these vary in the attention paid to how spatial and personal deixis relate to one another. In standard semantic approaches, indexicals have traditionally been modelled as functions on contexts, which specify the speaker, time, location, etc. of a context of utterance (Kaplan, 1989). For example, words like 'I' or 'here' would (roughly) have the denotations below, where *speaker* and *location* are functions that pick up the speaker and the location from the context. In what follows, we will often refer to these contextual parameters as the *semantic primitives* that define indexical expressions.

(1) a. $[\![\text{I}]\!]^c = \mathit{speaker}(c)$
b. $[\![\text{here}]\!]^c = \mathit{location}(c)$

Notably, the meaning of locative and pronominal expressions in (1) is defined in terms of different semantic primitives (i.e., *speaker* and *location*), suggesting that, besides being context-dependent, person and spatial indexicals do not necessarily have a common semantic basis.

Other linguistic approaches have explored the relationship between personal and spatial deixis by comparing how different natural languages organize the space of person and locative meanings (Anderson & Keenan, 1985; Imai, 2003; Song, 2011). For example, based on observed similarities in the types of partitions attested in person and locative systems, Harbour (2016) argues that the same morphological and semantic features underlie both types of deixis. Simplifying, Harbour (2016) proposes that both person and spatial indexicals are built from the same three ontological primitives: *speaker*, *addressee* and *other(s)*. In contrast to (1) then, no independent *location* primitive is posited.[1] Instead, locative indexicals like 'here' are represented in terms of the same contextual parameter *speaker* that defines the first person pronoun 'I', plus an additional spatial specification, as schematically illustrated in (2).[2] The semantic space that person and locative systems cover can then be thought as exploiting analogous primitives, as in Table 1.

(2) a. $[\![\text{I}]\!]^c = \mathit{speaker}(c)$
b. $[\![\text{here}]\!]^c = \text{space close to } \mathit{speaker}(c)$

In order to account for how these categories are grammaticalized into person and locative systems across languages, Harbour proposes two (morphosyntactic) binary features $\pm$AUTHOR and $\pm$PARTICIPANT, also shared across spatial and personal domains (see Table 1). The $\pm$AUTHOR feature distinguishes between individuals or spaces as a function of whether they include (or are

[1] An analogous argument has been made with respect to the need for an *addressee* primitive to distinguish first and second person indexicals. The second person pronoun 'you' could be defined as picking out the '*addressee*(c)' or 'the person who *speaker*(c) is talking to'. The reader is referred to (Deal, 2019; Sudo, 2012) for discussions about this issue.

[2] More specifically, Harbour proposes that locatives differ from personal pronouns in that they involve an additional syntactic head χ that dominates person features. Semantically, χ takes the set of individuals satisfying a particular person specification, and yields, instead, the set of spaces in the vicinity of those individuals. For example, the English locative 'here' refers to spaces near an individual satisfying the feature +AUTHOR (i.e., the speaker).

(i) a. $[\![\text{I}]\!]^c = [\![+\text{AUTHOR}]\!]^c = \mathit{speaker}(c)$
b. $[\![\text{here}]\!]^c = [\![\chi(+\text{AUTHOR})]\!]^c = \text{close to } \mathit{speaker}(c)$

(a) Person systems

Space (of possible referents)	AUTHOR bipartition (Damin pronouns)		PARTICIPANT bipartition (Winnebago pronouns)	
speaker	+AUTH.	'n!aa'	+PART.,	'nee'
addressee	−AUTH.	'n!uu'	+PART.,	'nee'
other	−AUTH.	'n!uu'	−PART.,	''ee'

(b) Locative systems

Space (of possible referents)		AUTHOR bipartition (English locatives)		PARTICIPANT bipartition (Bulgarian locatives)	
	speaker	+AUTH.	'here'	+PART	'tik'
space near...	addressee	−AUTH.	'there'	+PART	'tik'
	other	−AUTH.	'there'	−PART	'tam'

Table 1: Examples of person and locative person partitions

close to) the speaker, whereas the ±PARTICIPANT feature contrasts groups depending on whether they include (or are close to) the participants in the conversation (i.e., speaker and addressee).[3]

Languages vary with respect to which of these two features are active (if any) in each indexical domain. For example, English arguably relies on both features to make a three-way person distinction (e.g., 'I', 'you','they'), whereas it only makes a two-way AUTHOR-based locative distinction, contrasting between spaces near the speaker ('here') and far from them ('there') (e.g., Author bipartition in Table 1a). In contrast, languages like Bulgarian have an English-like person system but make a two-way PARTICIPANT-distinction in their locative system (e.g., Participant bipartition in table 1a). Languages can also exhibit a locative tripartition, parallel to the classic three-person system attested in English: Korean, for instance, distinguishes between the vicinity of the speaker ('yeki'), of the addressee ('keki') and of neither of them ('ceki'). This kind of system would make use of both ±AUTHOR and ±PARTICIPANT features.[4]

Importantly, by using the same features for person and spatial domains, Harbour's system

[3] These two features are not sufficient to describe all locative systems attested in the world's languages. Locative systems can involve additional distinctions beyond the ±AUTHOR and ±PARTICIPANT features, as noted by Harbour himself (see also Imai, 2003, for an overview of possible systems). Some languages have been argued to have distance-based locative systems, where distinctions seem to be made mainly on the basis of the distance of the object with respect to the speaker (other individuals in the speech context are irrelevant). For example, Modern Japanese uses three-way locative system that distinguishes between spaces at a small, medium or large distance from the speaker (e.g., 'ko', 'so', 'a') (Anderson & Keenan, 1985; Harbour, 2016; Imai, 2003). Crucially, even if these distance-based systems are not fully person-based, they still rely on person-categories, as the location of speaker and addressee plays a role in the determination of categories.

[4] The reader is referred to Harbour (2016, Ch. 4 and 7) for a complete typological overview of the attested person and locative systems. Table 1 is a simplification of the possible systems. Besides bipartitions and tripartitions, locative and person systems can also make an additional clusivity contrast. Languages like Waray-Waray, for example, distinguish between spaces near the speaker alone ('a(a)di'), near the addressee alone ('a(a)da′'), near both speaker and addressee ('a(a)nhi'), and far from both of them ('a(a)dtu'). A similar pattern is found in the Illocano pronominal system (Cysouw, 2003).

predicts that the same constraints that limit possible person systems should also limit locative systems. For example, Harbour predicts that the only possible bipartitions of the indexical space should be either AUTHOR or PARTICIPANT based. That is, all two-way person and locative systems should be characterized by using one of these features. This prediction seems to be validated cross-linguistically: only these two bipartitions are systematically attested for personal pronouns and locatives in the typology, as illustrated in Table 1.[5]

1.2 The challenge from indexical shifting

The idea that both sets of indexicals should be unified in the way proposed by Harbour is challenged by evidence coming from indexical shifting (Deal, 2019). Indexical shift refers to a phenomenon attested in certain languages whereby indexical expressions embedded under attitude (e.g., *believe*) or speech (e.g., *say*) verbs do not get their referent from the utterance context but rather from the reported or attitudinal context (Anand & Nevins, 2004; Deal, 2020; Schlenker, 2003). In Zaziki (Indo-Iranian language spoken mostly in Turkey), for example, a first person pronoun embedded in an attitudinal context (e.g., under the predicate *say*) can either refer to the actual speaker as in (3a) or to the author of the reported speech as in (3b). Indexicals which are anchored to the reported speech act are considered to be *shifted*. This is not a possibility in languages like English, where indexicals always have to be anchored to the actual speech act (i.e., they cannot shift).

(3) Hɛseni va kɛ ɛz dɛwletia.
Hesen said that I rich.be-PRS
a. Hesen said that I am rich.
b. Hesen said that Hesen is rich. Zazaki (Anand & Nevins, 2004, ex. 4)

Crucially for our purposes, the possibility of shifting is not categorical (all or nothing) but it appears to be determined by specific semantic properties of the indexical expression. As new indexical-shifting languages were discovered, it became clear that indexical shifting does not operate on a fully indexical-by-indexical basis, but it's however sensitive to *classes* of indexicals. In her typology of indexical shifting, Deal (2020) notices that, while all indexicals *can* in principle shift, different languages show different shifting patterns. For example, languages like Zazaki and Nez Perce allow shifting of all person and locative indexicals, whereas, in Uyghur, first and second person indexicals may shift, but locatives cannot (Anand, 2006; Deal, 2017; Sudo, 2012). Interestingly, no attested language allows locative shift without first and second person shift.

Besides this cross-linguistic variation, Deal (2020) also observes *within-language* variation with respect to which indexical expressions may shift within a single embedded clause. In Zazaki, all indexicals within a clause have to shift together; that is, all indexicals have to be anchored to the same context of evaluation (Anand & Nevins, 2004).[6] In contrast, in Nez Perce, there can be

[5] This is based on a method of classifying partitions in which all relevant person or spatial deixis paradigms within a language (e.g., personal pronouns, paradigms of person agreement, etc.) are super-imposed, and therefore a language counts as using a particular feature if any paradigm in the language uses that feature. That means that other bipartitions might arise on the level of the individual paradigm, as instances of syncretism (Cysouw, 2003): for example, English is taken to distinguish three person categories (i.e., person tripartition) but shows syncretism between first and second person in present tense agreement (i.e., it has bipartition in that paradigm).

[6] The pattern attested in Zazaki was originally used to motivate the 'Shift together' principle, which stated that all indexicals within a speech-context domain must pick up their reference from the same context (Anand & Nevins,

person shift without locative shift within a single clause. The opposite pattern, however, does not hold: locative shift without person shift is not possible in Nez Perce.[7]

In most current views of indexical shift, these facts are accounted for by assuming that: (a) indexicals are functions that pick out contextual parameters (*á la* Kaplan, see entries in (1)); and (b) shifting operates on specific contextual parameters. Simplifying, shifting is then taken to occur when a certain contextual parameter is overwritten with information from the reported speech act. Importantly, this overwriting process is assumed to operate independently for each contextual parameter (e.g., speaker, addressee, etc).[8] As an example, consider shifting of first person indexicals, which are defined in terms of 'the speaker in c'. By manipulating the context against which the *speaker* parameter is evaluated, all first person indexicals in the embedded clause will refer to the attitude holder, rather than to the utterance speaker. This, however, will arguably not affect other indexicals in the clause, which are not defined by the *speaker* contextual parameter. Languages may then vary depending on whether they allow changing the interpretation of all or only some contextual parameters (e.g., Zazaki vs. Uryghur).

As the attentive reader might have noticed, this account of indexical shift has important implications for our original question, as indexicals that have a common semantic basis are predicted to pattern together with respect to shifting. If the same primitives underlie personal and spatial deixis, as proposed by Harbour (2016), one might expect first person indexicals and locatives defined in terms of a *speaker* primitive to shift together in languages that allow indexical shifting (*mutatis mutandis* second person indexicals and locatives defined by an *addressee* primitive). As observed, this is not the case, neither at the language level (e.g., Uyghur) nor within a single clause (e.g., Nez Perce). As a result, in this theory of indexical shift, personal and spatial indexicals need to involve different semantic primitives (e.g., *speaker* vs. *location*).

Before moving on, let us note that these indexical shifting patterns constitute a challenge for views such as Harbour's under the assumption that shifting is conditional on specific contextual parameters. Alternative accounts of indexical shift may not necessarily raise this issue (binding-based account of indexical shift; Schlenker, 2003; Von Stechow, 2002, see Fn.19).

1.3 Our goals

Here we aim to bring behavioral evidence from learning to bear on the question of whether spatial and personal indexicals should receive a unified treatment in terms of their features. According

2004). This principle was later reformulated in light of new data (e.g., Nez Perce), relativizing the generalization to *classes* of indexicals (see Deal, 2017, for discussion).

[7]No language has been discovered where locative shift without person shift is possible, neither generally at the language level nor within a single clause. This has been taken by Deal (2020) as evidence for an implicational hierarchy with respect to which indexicals can shift: "Within and across languages, the possibility of indexical shift is determined by the hierarchy **1** $>$ **2** $>$ **locative**. Indexicals of a certain class undergo shift in a particular verbal complement only if indexicals of classes farther to the left undergo shift as well." (Deal, 2020, Ch.3, p.66). This asymmetry is not immediately relevant for our purposes but we will briefly come back to it in the General Discussion.

[8]To be more specific, this view is known as the operator-based account of indexical shift precisely because it posits the existence of a silent context-shifting operator which scopes over the embedded predicate and manipulates a certain contextual parameter in its scope (Anand, 2006; Anand & Nevins, 2004). In principle, there are as many shifty operators as contextual parameters. For example, the *speaker* (or *author*) operator will force its complement to be interpreted relative to a context whose speaker value has been replaced by the speaker value of the reported context (Deal, 2020).

to the theory given in Harbour (2016), spatial indexicals use the same set of features as person indexicals. In contrast, research on indexical shift tends to rely on standard semantic assumptions (Kaplan, 1989) and assume that deixis of person and space involve different semantic primitives, and thus different features (Anand, 2006; Anand & Nevins, 2004; Deal, 2020).

From these two alternative approaches, we derive predictions for learning under the general hypothesis that if two systems share underlying structural and/or semantic primitives, then a priori, a learner should assume those systems are organized according to the same rules or distinctions in a language. Put another way, learners should generalize across domains, rather than assuming different structures or distinctions. This prediction follows from independent evidence suggesting that learners tend to generalize or extend known rules and distinctions to new items and domains (e.g., in natural acquisition and in laboratory experiments: Finley & Badecker, 2009; Hupp et al., 2009; Linzen & Gallagher, 2017; Marcus et al., 1992; Myers & Padgett, 2014; Saffran & Thiessen, 2003; Yang, 2002; Yang & Montrul, 2017, among many others), combined with the broader assumption that generalization likelihood is affected by the similarity of the relevant items or domains (see Dautriche et al., 2016; White, 2014; Xu and Tenenbaum, 2007 for various evidence for this assumption in learning).[9] One can conceive of this general pressure to re-use structures, rules, and distinctions across domains under the umbrella of simplicity biases in learning (e.g., Chater & Vitányi, 2003; Culbertson & Kirby, 2016). Applied to the case at hand, if spatial and person indexicals use the same set of features, then, absent evidence to the contrary, learners are predicted to assume a single underlying partition or set of features for both kinds of indexicals. By contrast, if the two systems are distinct, i.e. do not share features, then learners cannot straightforwardly generalize from one to the other (at least not based on features).[10]

The present study uses artificial language learning experiments (Culbertson, 2012; Maldonado & Culbertson, In press) to test the hypothesis above. In Experiment 1, we investigate whether, when learners are trained on a pronoun system in an artificial language, they generalize the person distinctions made in that system to a new locative system. Experiment 2 reverses the directionality of the generalization, testing whether learners generalize a locative distinction to a person system. Finally, Experiment 3 explores whether any logically possible pronominal system can serve as a model for learning a locative system.

2 Experiment 1

In Experiment 1, we used an artificial language learning paradigm to investigate whether learning a novel locative system is influenced by prior knowledge of a pronominal system. This experiment, including all hypotheses, predictions, and analyses, was preregistered. More details about the

[9]Similar assumptions are made in the structural priming literature with respect to the so-called ‘lexical boost’, whereby priming effects are stronger when lexical content is repeated (Pickering & Ferreira, 2008; Traxler et al., 2014).

[10] In principle, this learning-based hypothesis could also have implications for expected typological patterns. Learners are often assumed to drive linguistic change over time by altering the language in response to their preferences during acquisition (Clark & Roberts, 1993; Prince & Smolensky, 1997). One would then expect that, if learners prefer to uniformly organize locative and person systems, such bias might have an impact in the typology, resulting in a within-language tendency to use the same contrasts across domains. That is, languages should be more likely to make the same partition in their locative and person systems than a different one. The typological data available to us at the moment, however, does not allow us to assess whether this is the case.

design and the experiment can be found in the associated OSF repository.[11]

2.1 Methods

2.1.1 Design

The experiment uses a version of the extrapolation paradigm (e.g., Culbertson & Adger, 2014; Maldonado & Culbertson, In press; Wilson, 2006), in which learners are trained on input that is ambiguous between two systems of interest, and then tested on stimuli which reveal which of the two systems they have inferred. Here, participants will be trained on an ambiguous locative system. Two locative forms are used, one of which refers to spaces in the vicinity of the speaker (e.g., 'hiti'), the other to spaces in the vicinity of other, non speech-act participants (e.g., 'hodo'). Crucially, a third meaning was held out: during training, participants were not exposed to any situation where a locative form was used to refer to spaces in the vicinity of the addressee. In the critical test phase, participants must extrapolate by using one of the forms they were trained on to express this held out meaning.

Different patterns of extrapolation indicate what bipartition of the locative space participants have inferred. For example, if learners use the locative form that refers to spaces in the vicinity of the speaker (i.e. 'hiti') to express the held-out meaning, then they have inferred an participant bipartition (i.e., forms distinguish between spaces as a function of whether they are in the vicinity of speech-act participants). Instead, if learners use the form that refers to spaces in the vicinity of others (i.e. 'hodo') for this new meaning, they have inferred an author bipartition (i.e., forms distinguish between spaces as a function of whether they are in the vicinity of speaker).

Importantly, we manipulated whether participants are additionally pre-trained on a partition of the person space (i.e., a pronominal system). Participants in the experiment are randomly assigned to one of three conditions. A baseline condition, in which they are only trained on the ambiguous locative system described above, and two critical conditions in which they are first trained on a complete pronominal bipartition. In Condition 1 (Author pre-training), participants were first trained on a pronominal author bipartition; in Condition 2 (Participant pre-training), they were first trained on a pronominal participant bipartition. All three conditions are illustrated in Figure 1.

By evaluating whether participants' likelihood of inferring either an author or a participant bipartition of the locative space is affected by pre-training on a pronominal system, we can test the predictions of the accounts outlined above. If learners assume that the same features underlie both indexical systems, then, all things equal, they should extrapolate the locative forms according to the pronominal bipartition they were taught. We would then expect subjects in Condition 1 to be more likely to infer an author bipartition of the locative space than subjects in Condition 2 (and vice-versa for the participant bipartition). The Baseline condition allows us to measure the relative preference for the two locative bipartitions in the absence of any influence from a pre-trained pronominal system. This is particularly relevant in light of participants' (native) experience with English: English features an author bipartition in its locative system (Harbour, 2016, see also Anderson and Keenan, 1985; Imai, 2003 for a similar distinction), so one might expect participants to have an a priori preference for these bipartitions.

[11] The original pre-registration of this experiment included both Experiment 1 and Experiment 3 together. We are breaking down these for presentation purposes.

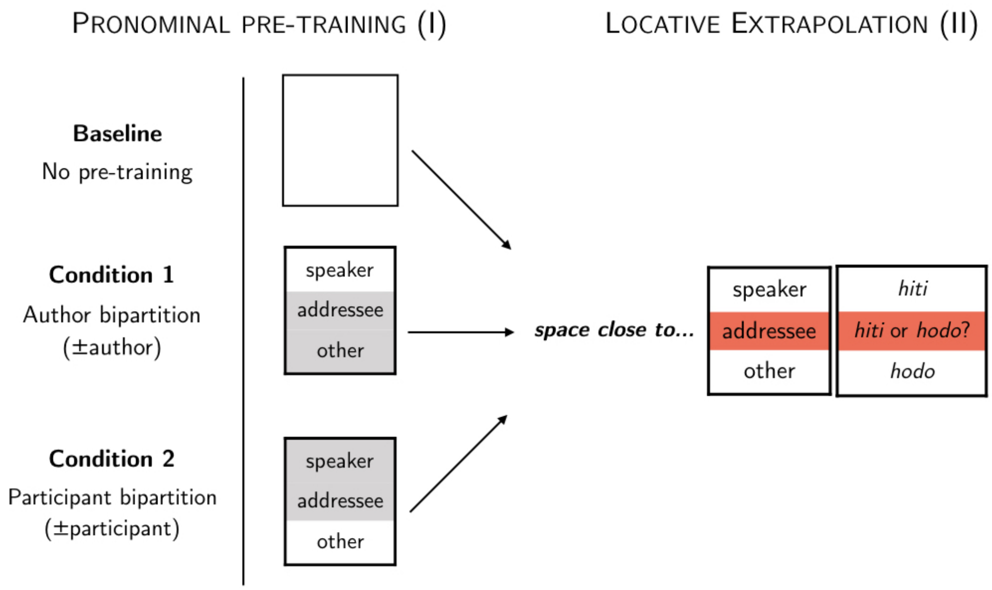


Figure 1: Conditions in Experiment 1. The left-side shows the personal pronoun system that learners were first trained on, depending on the condition (Author bipartition, Participant bipartition or no pronominal training). Cell colouring indicates the pronominal forms: cells with the same color are mapped into the same form. The right-side shows trained and held-out spatial categories, in white and red respectively. Each trained meaning/category was mapped into a different locative form ('hiti' or 'hodo'). Participants must use one of these forms to express the held-out meaning at test.

2.1.2 Materials

The lexical items in the locative training phase, 'hiti' and 'hodo', were created to facilitate learning of locative meanings. First, they are similar to one other, along the lines of the morphologically related forms 'here' and 'there' in English. Second, they were not randomized: 'hiti' was always used for spaces near the speaker and 'hodo' for spaces near non-participants in the conversation. The reasoning behind this choice was to exploit potential sound symbolism between locations and front/back vowels (Rabaglia et al., 2016; Woodworth, 1991).

The lexical items in the pronominal training phase (for Conditions 1 and 2) were taken from Maldonado and Culbertson (In press). For each participant, the four pronominal forms were randomly drawn from a list of 6 CVC non-words created following English phonotactics: 'kip', 'dool', 'heg', 'veek', 'tosh', 'lom'. Items were presented orthographically.

Pronouns and locative forms were used as one-word answers to questions presented in English. Visual stimuli were pictures of a family and their pets. The family consisted of three sisters and their parents. Each family member had a clearly-defined role in the conversational context. The two older sisters were always speech act participants (speaker or addressee). The parents and the third little sister were in the background serving as additional others, without participating in the conversation (see upper panels in Figure 2). In both parts of the experiment, the speaker and addressee roles switched several times to highlight that the words were dependent on contextually-determined speech-act roles. The questions for the pronoun training phase targeted sub-groups of family members (e.g., 'Who will be rich?'). The reference of the pronoun was expressed by highlighting family members, as exemplified in Figure 2 (see Procedure below for more details).

The family pets consisted of a cat, a parrot, and a turtle. The questions for the pronoun training targeted locations of the pets (e.g., 'Where is the cat/parrot/turtle?'). The reference of the locative was expressed by revealing the location of the pet, as exemplified in Figure 2 (see procedure below for more details).

2.1.3 Procedure

The general procedure was based on previous work on learning of pronominal systems presented in Maldonado and Culbertson (In press). Participants were first introduced to the family. In Conditions 1 and 2, following Maldonado and Culbertson (In press), the pronominal pre-training phase began with participants being told that they were going to see the sisters playing with a hat that had two magical properties: whoever wore it could see the future but would also talk in a mysterious ancestral language. In each trial, one of the sisters would ask a question about future events in the life of the family, and the sister wearing the hat would answer the question using a pronoun in the mysterious language. Participants were instructed to figure out the meanings of words in this new language. Participants were exposed to and then tested on four pronominal forms, used to cover six person/number categories (first, second and third persons, in the singular and plural).

Each exposure trial had two parts: a scene where a question is asked, and a scene where the question is answered with a pronominal form in the language. The reference of the pronoun was given by highlighting a subset of the family members. To check that participants were paying attention, they were then asked to select the pronominal form they had just seen from two alternatives, which were either the two singular or the two plural pronominal forms. There were 30 exposure trials in total (4 repetitions for singular categories, 6 repetitions for plural categories). Testing trials consisted of a question scene, followed by a scene highlighting the referent(s), but no pronominal form. Participants had to pick the word corresponding to the meaning from two alternatives, which were either the two singular or the two plural pronominal forms. Testing was divided in two blocks: there was a first block of 18 testing trials for singular categories (6 repetitions), and then a second block of 36 testing trials for both singular and plural categories (6 repetitions). Participants were given feedback on all their answers.

Participants were required to learn the full pronominal system: only those who responded accurately on more than 2/3 of the testing trials were allowed to continue with the experiment, passing to the locative phase.

The locative extrapolation phase was the first phase in the Baseline condition and the second phase in Conditions 1 and 2. Participants were introduced here to the family's pets: a cat, a parrot and a turtle. Participants were told that these animals were very naughty and kept getting lost in the family backyard. Participants were told that whoever was wearing the magical hat could also use it to find missing objects or animals.[12] The sister without the hat would ask where one of their pets was, and the sister with the hat would answer in the mysterious language, as the location of the pet in the scene would simultaneously be revealed (see lower panel in Figure 2). Participants were given a hint that the answer words were not hiding places or objects but always indicate the location of the pet. They were also given an example trial with an English locative ('over here').

Participants were then trained on two locatives forms in the language, one used to refer to spaces in the vicinity of the speaker, and another for spaces in the vicinity of the parents. Each

[12] Participants in the Baseline condition were introduced here to the magical hat which was described as having two magical properties: whoever wore it could find missing objects but would also talk in a mysterious ancestral language.

training trial involved a scene where a question is asked, and a scene where the question is answered with one of the two locative forms in the language (Figure 2E-G). To check that participants were paying attention, they were then asked to select the locative form they had just seen from two possible alternatives. There were a total 12 training trials (6 repetitions per form).

After this initial training, participants were tested on these locative forms. As in the pronominal phase, they were presented with the location of one of the pets, and they were asked to pick, between the two locative forms, the one that they would use to refer to that location. This testing block included trials both for the locative meanings they had seen (trained categories) as well as for the held-out category. There were 36 testing trials (8 repetitions of the held-out meaning, 14 repetitions of each taught meaning). Participants received no feedback during testing.

Speaker and addressee roles switched several times during both experimental phases. This change was induced by swapping who had the magical hat, and serves to reinforce the context dependent meaning of the forms, ensuring that participants cannot develop strategies based on the specific characters or locations, or on visual properties of the stimuli. An illustration of alternative trials and a more detailed description of the design are given in the Supplementary Materials.

2.1.4 Participants

A total of 187 English-speaking adults (Condition 1: 72; Condition 2: 64; Baseline: 51) were recruited via Amazon Mechanical Turk. Participants who performed accurately on the pronominal phase of the experiment passed to the second phase (Condition 1: 52; Condition 2: 57). Participants who completed only the first part of the experiment ($\sim$15 minutes) were paid 3 USD; participants who pass to the second part of the experiment ($\sim$10 minutes) were paid a bonus of 2.5 USD.

As per our pre-registered plan, participants were excluded from the analyses if they failed to correctly answer at least 2/3 of the taught forms during testing, which corresponds to 10 out of 14 correct responses per form. This resulted in the analysis of 132 participants in total (Condition 1: 43; Condition 2: 48; Baseline: 41).

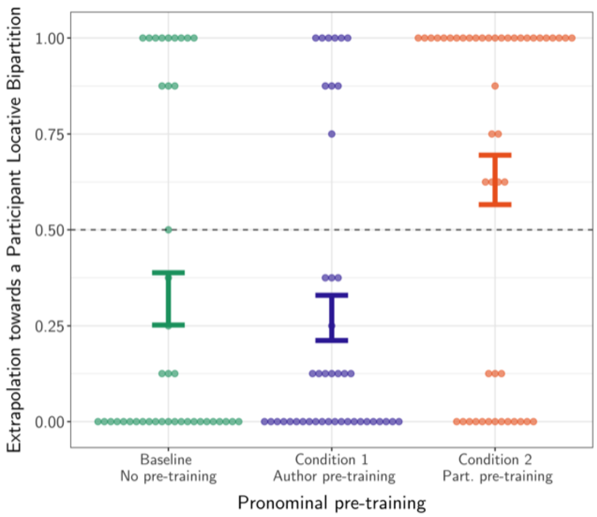
2.2 Results

Recall that in the locative phase participants were taught two locative forms, one for spaces near the speaker ('hiti') and another one for spaces far from both speaker and addressee ('hodo'). At test, they were asked to select between these two locative forms to refer to spaces in the vicinity of the addressee. We are interested on whether participants' responses for this held-out meaning were influenced by the pronominal training they received in the first part of the experiment (with the Baseline condition as a control).

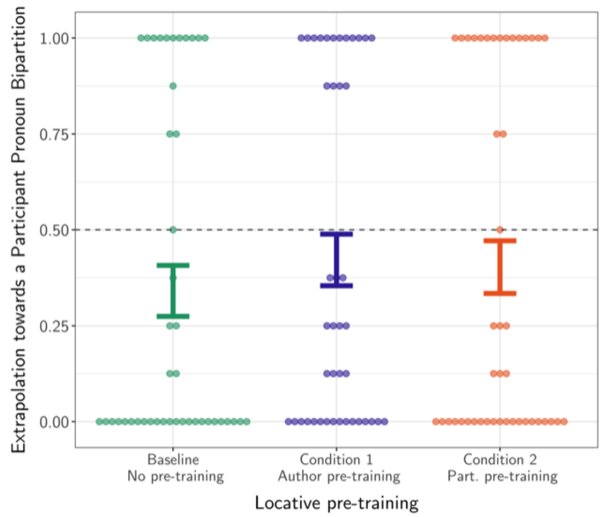
Figure 3a shows the proportion of responses to the held-out meaning compatible with a participant locative bipartition as a function of pronominal training (i.e, whether learners were pre-trained on an author pronominal bipartition, a participant pronominal bipartition or they had no pronominal training at all). Responses were treated as compatible with a participant locative bipartition when the form chosen in held-out trials—for spaces close to the addressee—was the same one as the form used for spaces close to the speaker (i.e. 'hiti').

We ran a logistic mixed-effects regression model predicting the responses in held-out trials by Condition (levels: Condition 1, Condition 2 and Baseline). Fixed effects were treatment coded

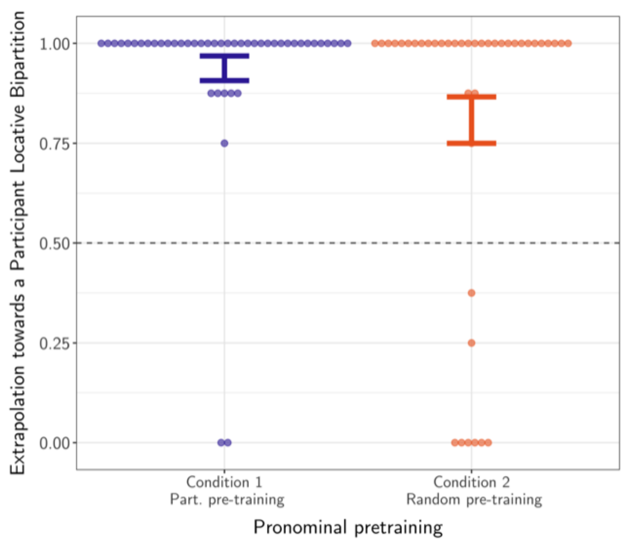
Figure 2: Experiment 1 example trials. Panels A-D show trials for the pronominal phase in Condition 2 (participant bipartition), including an example question (A), and responses for first, second and third person categories (B-D). Panels E-H show examples of locative trials: example question (E), exposure to the trained locative forms (F-G) and held-out trial (H), where participants have to choose between the two locatives which one they would use to refer to spaces near the addressee.



(a) Experiment 1



(b) Experiment 2



(c) Experiment 3

Figure 3: Response for held-out meanings in each experiment. (a) Experiment 1: Proportion of responses compatible with a participant *locative* bipartition (vs. author bipartition) after each type of pronominal pre-training. (b) Experiment 2: Proportion of responses compatible with a participant *pronoun* bipartition responses (vs. author bipartition) by locative pre-training. (c) Experiment 3: Proportion of responses compatible with a participant *locative* bipartition (vs. random bipartition) by pronominal pre-training. Error bars represent standard error on by-participant means; dots represent individual participant means.

(with Baseline condition as baseline). Random by-participant intercepts were also included.[13] This model revealed that the log odds of inferring a participant bipartition of the locative space (i.e., choosing the locative form 'hiti') in the Baseline condition were significantly below chance

[13] We use the standard alpha level of 0.05 to determine significance, and follow the standard practice for generalised linear models of obtaining p-values based on asymptotic Wald tests (type III). In simplified models, p-values were corrected for multiple comparison using a Bonferroni correction.

(intercept: $\beta = -8.6;\ p < .001$). The likelihood of deriving a participant bipartition in Condition 1 (Author pre-training) was not statistically different from the Baseline ($\beta = -0.01;\ p = .9$). By contrast, relative to the baseline, learners were significantly more likely to infer a participant bipartition in Condition 2 (Participant pre-training) : $\beta = 16.8;\ p < .001$.

A second, simplified model was fitted to directly compare Conditions 1 and 2, using treatment coding with Condition 2 as baseline. This model revealed that, after being trained on a participant partition of the person space, the likelihood of inferring a participant partition of the locative space was (i) above chance (intercept: $\beta = 7.9;\ p < .001$) and (ii) significantly higher than after being trained on an author person bipartition ($\beta = -16.16;\ p < .001$).

2.3 Discussion

The main goal of Experiment 1 was to test the general hypothesis that person and spatial deixis share semantic features. In this experiment, we assessed this hypothesis by exposing English-speaking participants to two locative forms in an artificial language, and then testing whether the locative system they inferred was influenced by the pronominal training they had in the first part of the experiment.

Our results first indicate that, in absence of any pronominal training (i.e., Baseline condition), learners are highly likely to infer an English-like locative system, defined by an author bipartition of the locative space. The existence of this native-language bias reinforces the idea that participants in the experiment understand the stimuli as we intend—i.e., in terms of a locative system.[14]

More importantly, learners in our experiment also show a bias towards using the same set of features to uniformly organise person and spatial indexical systems. After being trained on a participant bipartition of the person space (where first and second persons were expressed using a single homophonous form), learners were more likely to infer the same (non-native) partition in a new locative system. This behavior involves over-riding their native-language system. Perhaps surprisingly, the impact of pre-training was not symmetric: learners' preference for a native-like locative system was weakened by pre-training with a participant pronominal bipartition, but was not strengtheded by pre-training with an author pronominal bipartition. However, this may simply reflect a floor effect, as preferences in the Baseline condition are already quite strong.

In §1, we have observed that Harbour's system differs from standard semantic approaches *á la* Kaplan is that locative indexicals do not require specific semantic primitives, and can instead be built on person primitives. Our findings provide evidence that learners extend a person contrast to a new locative system, suggesting that locatives can indeed be represented by the same formal and semantic features that define person indexicals. However, in order to assess whether person and spatial indexicals are defined by a fully overlapping set of features, we need to investigate whether learners also extend a *locative* contrast to a new person system. Experiment 2 was designed to address this question.

[14]This is further confirmed by a debrief questionnaire, in which most participants reported having understood the new words as locatives.

3 Experiment 2

In Experiment 2, we tested whether pre-training on a *locative* bipartition influences learners' inferences about a new personal pronoun system. The design of the experiment is therefore similar to Experiment 1, but swaps the direction of influence tested. The preregistration for this experiment can be found here.

3.1 Methods

3.1.1 Design

In Experiment 2, English-speaking participants were exposed to a partial person partition in which two pronominal forms were used to convey the first and third singular person categories, with second person singular held out (no plural meanings were used). That is, one form was used to refer to the speaker and another one to refer to a non-participant in the conversation. Situations in which the referent was the addressee held out. In the critical test phase, participants had to extrapolate by using one of the forms they were trained on to express this held out meaning. Different patterns of extrapolation indicate whether learners have inferred an author or a participant bipartition of the person space.

As in Experiment 1, we manipulated what partition of spatial indexicals (if any) participants were pre-trained on. In the Baseline condition, participants were only trained in the ambiguous pronominal system described above. Note that since English does not use a bipartition of the person space, we do not have a specific prediction regarding what is learners' preferred person system in absence of locative training.[15]

In Conditions 1 and 2, participants were pre-trained on a complete locative system. Two locative forms (again 'hiti' and 'hodo') were used to cover three categories of spatial deixis: spaces near the speaker, near the addressee, and spaces distant from both of the speaker and the addressee (close to others). In Condition 1, they were trained on an author bipartition, where the contrast is between spaces close to or far from the speaker. In Condition 2, they were trained on a participant bipartition, where the contrast is between spaces near speech act participants (speaker or addressee) or far from them. A summary of the conditions is given in Figure 4.

3.1.2 Materials

Materials were the same as in Experiment 1. The form-to-meaning mapping for locatives is specified in Figure 4. Pronominal forms were drawn from the same list of 6 CVC items used in Experiment 1.

[15] Maldonado and Culbertson (In press) report an extrapolation experiment where participants have to assimilate the second plural person category with either the first or the third (Experiment 3, Condition 2). This is arguably be very similar to our Baseline condition. Given that Maldonado and Culbertson's design differs from ours in some important aspects—for example, it tests plural rather than singular pronouns—we opted by not using their results as a baseline for statistical comparison, and run our own baseline condition. For reference, however, Maldonado and Culbertson (In press) found that English-speakers have an overall preference for participant bipartitions of the person space (i.e., homophony between second and third person forms).

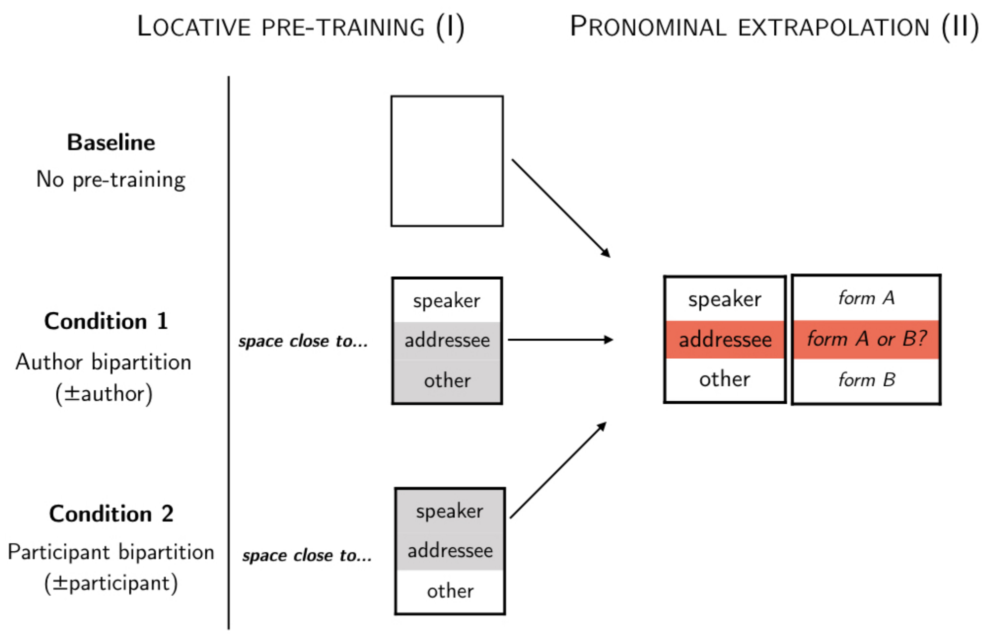


Figure 4: Conditions in Experiment 2. The left side shows the locative system that learners were first trained on: either an author or participant bipartition. Grey cells indicate use of the same locative form. The right side shows trained and held-out person categories, in white and red respectively. Each trained category was mapped into a different pronominal form. Participants could use either of these forms to express the held-out meaning.

3.1.3 Procedure

The procedure was analogous to Experiment 1, except that a *locative* training preceded a *pronominal* extrapolation phase. An illustration of the trial structure is given in Figure 5.

After being introduced to the general backstory (i.e., presentation of the family and their pets), participants were instructed to figure out the meanings of words in the new language. In the locative pre-training phase, participants were trained on two locative forms in the language, used to cover all spatial categories. The specific form-to-meaning mapping was determined by the condition. The structure of the trials was as in Experiment 1. There were a total of 30 exposure and 54 testing trials. Participants were given feedback on all their answers. Participants who responded accurately to more than 2/3 of the testing trials were allowed to move forward to the second, pronominal phase.

In the pronominal phase, participants were first trained on two pronominal forms, used to refer to the first and third person singular categories. The structure of the training trials was the same as in Experiment 1. There were a total of 12 training trials (6 repetitions per form/meaning). Participants were given feedback on all their answers. Participants were then tested on these pronominal forms. As in Experiment 1, they were presented with sets of referents, and they were asked to pick, between the two pronominal forms, the one that they would use to refer to that set of referents. This testing block included trials both for the pronominal meanings they had seen (first and third person) as well as for the held-out category (second person). There were a total of 12 training trials (6 repetitions per form/meaning) and 36 testing trials (8 repetitions of the held-out meaning,

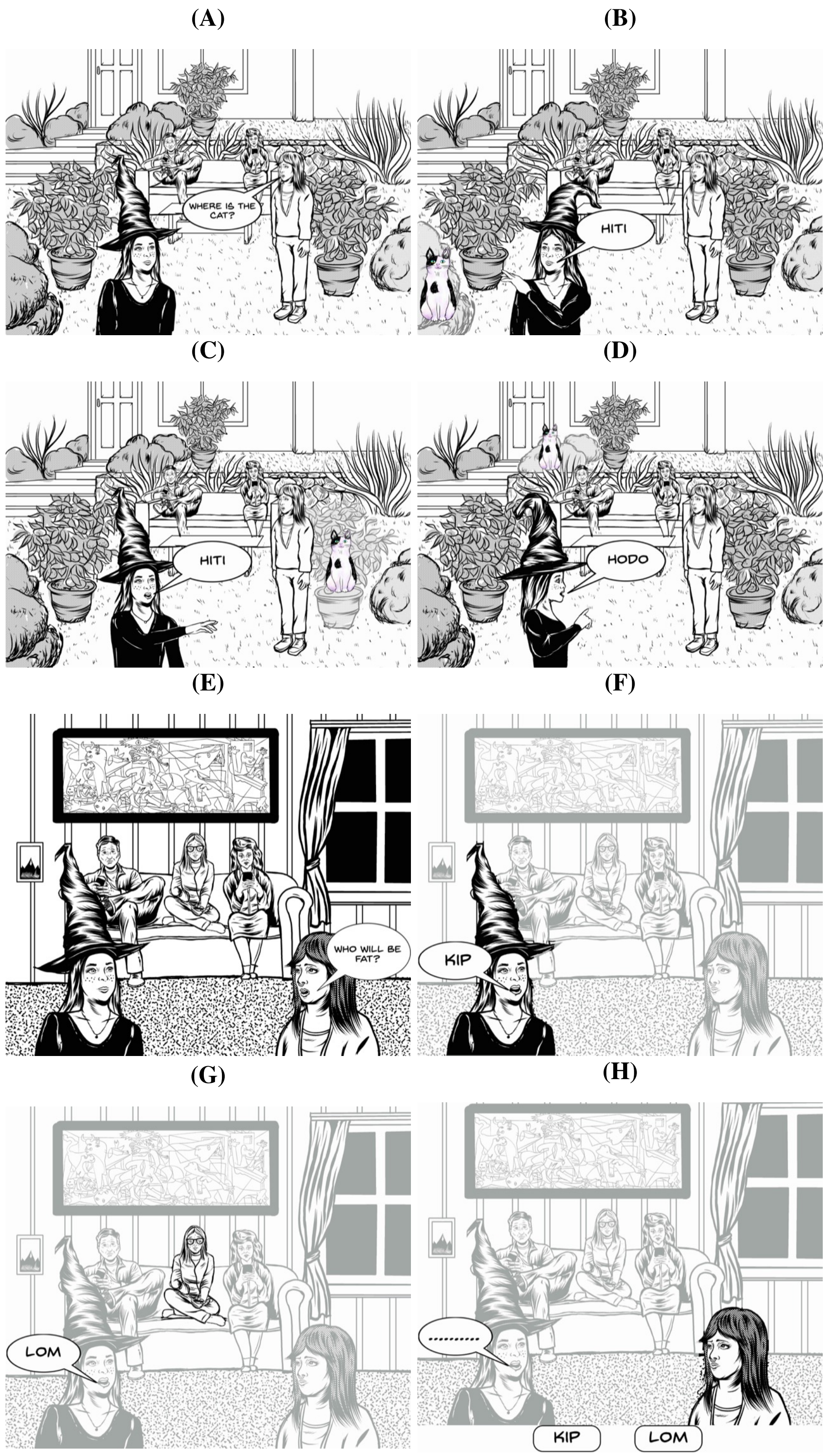


Figure 5: Experiment 2 example trials. Panels A-D show trials for the locative phase in Condition 2 (participant bipartition). Panels E-H show examples of pronominal phase, including training on first and third person categories (F-G) and held-out testing (H), where participants have to extrapolate to the second person category.

14 repetitions per taught meaning). No feedback was provided.

A more detailed description of the design is provided in the Supplementary Materials.

3.1.4 Participants

A total of 184 English-speaking adults (Baseline: 76; Condition 1: 53; Condition 2: 55) were recruited via Amazon Mechanical Turk. 102 participants passed to the pronominal phase in Conditions 1 and 2 (Condition 1: 50; Condition 2: 52). Participants were excluded from the analyses if they failed to correctly answer more than 2/3 of the taught forms during testing (i.e., 10 out of 14 correct responses per taught form). This resulted in the analysis of 88 participants in total (Baseline: 44; Condition 1: 43; Condition 2: 45).

3.2 Results

Figure 3b shows the proportion of responses to the held-out meaning compatible with a participant pronominal bipartition as a function of locative training (i.e, whether learners were pre-trained on an author locative bipartition or a participant locative bipartition). Responses were treated as compatible with a participant pronominal bipartition when the form chosen in held-out second person trials was the same one as the form used for first person.

A visual inspection of Figure 3b suggests that there is no clear stable pattern of responses: some learners consistently derive a participant bipartition, some an author bipartition, and others behave randomly. Crucially, the locative system participants were trained on does not seem to have a clear influence on the extrapolation patterns.

To evaluate this statistically, we ran a logit mixed-effect model analogous to the one used in Experiment 1, predicting the responses in held-out trials by Condition and including random-by-subject intercepts. Responses were coded as 1 if they matched the participant pronominal bipartition, and as 0 otherwise. This model reveals that the likelihood of inferring a participant bipartition of the person space in the Baseline condition was significantly below chance (intercept: $\beta = -8.68$; $p < .001$). Neither Condition 1 nor 2 were statistically different from the Baseline (Condition 1: $\beta = 1.36; p = 0.44$; Condition 2: $\beta = 0.67; p = 0.65$).

A simplified model was used to compare Conditions 1 and 2. After being trained on a participant locative bipartition in Condition 2, the likelihood of inferring a participant bipartition of the person space was not statistically higher than chance (intercept: $\beta = -1.94$; $p = .12$) nor significantly different from the pattern of responses in Condition 1 ($\beta = 0.7$; $p = .66$).

3.3 Discussion

Recall that in Experiment 1, we found that learners' inferences about a new locative system were influenced by pre-training on a pronominal system. Here, we tested the reverse, in order to evaluate whether learners generalize a locative contrast they have learned to a new person system. Our results failed to find any effect of locative pre-training on learners' inferences for personal pronoun systems: no significant difference was found between either pre-training condition (Conditions 1 and 2) and the baseline in terms of the inferred pronominal bipartition.

Moreover, our findings reveal that, in absence of locative training (cf. Baseline condition), English-speaking learners are more likely to infer an author- than a participant-based pronominal system. This roughly matches Maldonado and Culbertson's results, which used a similar experimental paradigm to show that participants had a (strong) bias in favor of an author person bipartition (2=3).[16] While, at first sight, learners in Conditions 1 and 2 might not appear to behave uniformly with respect to which non-native bipartition of the person space they preferred (see Figure 3b), the likelihood of inferring one specific pronominal system in these conditions is not statistically different from the baseline preference rate.

The contrasting results across these two experiments suggest that learners do not necessarily represent pronominal and locative indexical systems as using the same set of underlying features. More specifically, learners do not seem to define personal pronouns by using the the same features that define locatives, suggesting that the semantics of locatives cannot be fully reduced to the semantics of person indexicals.

One possible explanation for this difference between experiments lies on a potential asymmetry in how person and spatial indexicals are represented. Intuitively, spatial categories can be easily defined *in terms of* person categories: the spatial indexical 'here' in English can be treated as 'close to *me*', without any loss of meaning. The opposite, however, might not hold: it might not be so straightforward to define person categories in terms of spatial primitives or categories. Imagine, for example, that the first person pronoun 'me' were defined as 'the individual *here*'. Arguably, that would not necessarily guarantee that 'me' refers to the speaker in the context: one could use the pronoun to refer to someone else, who happens to be close to the speaker.

This asymmetry might have an impact on learners in our experiment: After being initially exposed to a new pronominal system, learners might then become more likely to define new locatives in terms of personal pronouns, while they might not be able to, or as likely to do the same the other way around. As we return to in the General Discussion (§5), this intuitive asymmetry may be accounted for either by positing an additional, specific locative primitive, or by making the representation of spatial indexicals more sophisticated, such that they might involve additional properties besides the ±PARTICIPANT and ±AUTHOR features (in lines of Harbour's proposal, see Footnote 2).

Before discussing the differences between Experiments 1 and 2 any further, we present a third experiment aimed at pushing the results of Experiment 1 further. If learners in Experiment 1 can extend a feature-based distinction from the person to the spatial domain by virtue of the fact that both domains sharing a set of features, then distinctions which are *not* feature-based should not get carried over from one domain to another. To test this, we compare learners' inferences for locative systems following pre-training on either a pronominal bipartition with feature-based or random homophony.

4 Experiment 3

Experiment 3 replicates the design of Experiment 1: participants were first taught a complete person pronominal bipartition, then exposed to part of a locative bipartition, and finally tested on how they extrapolate two locative forms they have learned to the remaining locative meaning. Here, we compare pre-training on a participant pronominal bipartition to pre-training on a bipartition with random homophony.

[16]Maldonado and Culbertson's experiment tests *plural* rather than *singular* person categories. This difference might be enough to explain why the extrapolation pattern in the present study is weaker than the one in their experiment.

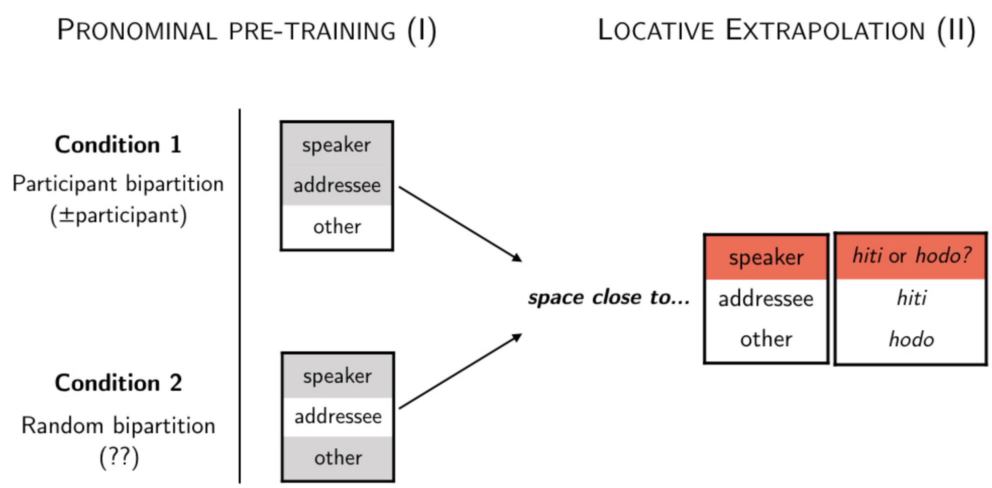


Figure 6: Conditions in Experiment 3. Left-side shows the personal pronoun system learners were first trained on, depending on the condition (participant or random bipartition). Right-side shows trained form-to-meaning mappings for locative categories in white and the held-out meaning in red.

4.1 Methods

4.1.1 Design

As before, we manipulated what pronominal bipartition learners were pre-trained on: either a participant bipartition (Condition 1), or a *random* bipartition, where the same pronominal form was used to refer to the speaker and to other non-participants (Condition 2). This partition is random in the sense that the two homophonous meanings do not belong to a natural class (i.e., do not share feature values), at least under most theories of person (Harbour, 2016; Harley & Ritter, 2002).

In the locative phase, learners were then taught that the locative ‘hiti’ referred to spaces near the addressee, and that ‘hodo’ referred to spaces near others. Participants had to then extrapolate by using one of the locative forms they were trained on to express the held-out meaning: spaces near the speaker. Different patterns of extrapolation indicate whether learners have inferred a participant or a random locative bipartition. Note that, by making this modification from Experiment 1, we can also directly investigate the preferences of English-speaking learners when they have to choose between two non-native locative systems.

Experiment 3 was identical to Experiment 1, except for the differences in form-to-meaning mappings specified above (see Figure 6). The experiment was originally pre-registered with Experiment 1.

4.1.2 Materials

Materials were the same as in Experiment 1, except that in this case the locative form ‘hiti’ was always used for spaces in the vicinity of the *addressee* rather than the speaker.

4.1.3 Procedure

The procedure was identical to Experiment 1 (see Supplementary Materials for more details).

4.1.4 Participants

134 English speakers were recruited and compensated as per Experiment 1. Participants were randomly allocated to Conditions 1 or 2 (Condition 1: 64; Condition 2: 70). 100 participants completed both pronominal and locative phases (Condition 1: 57; Condition 2: 43). As before, the data from participants who did not correctly answer 2/3 of the taught locative forms during testing were excluded from the analyses. This resulted in the analysis of 46 participants in Condition 1 and 41 in Condition 2.

4.2 Results

Figure 3c shows the percentage of responses to held-out meanings compatible with a participant locative bipartition as a function of training (i.e., whether learners were pre-trained on a participant pronominal bipartition, or a random bipartition). Responses were treated as compatible with a participant locative bipartition when the form chosen in held-out trials—for spaces close to the speaker—was the same one as the form used for spaces near the addressee (i.e., 'hiti'). Participants had a very strong preference for the participant bipartition across conditions. Indeed, a logistic mixed-effects regression model predicting response by Condition (treatment coded with Condition 1 as baseline) confirms that responses compatible with a participant bipartition were produced significantly above chance in Condition 1 (intercept: $\beta = 9.9;\ p < .001$), with no effect of Condition ($\beta = 0.94;\ p = .6$).

4.3 Discussion

Results from Experiment 3 suggest that learners do not generalize *all* person contrasts to a novel locative system: participants in Condition 2 failed to infer the same random partition for locatives that they were taught for pronouns. Instead, learners consistently inferred a participant bipartition of spatial deixis across the two conditions.

Unlike natural-class based bipartitions, random bipartitions of the kind tested here are difficult to derive by most theories of person, and in general typologically rare or unattested (Harbour, 2016; Harley & Ritter, 2002). In addition, they have been shown to be independently difficult to learn (Maldonado & Culbertson, In press). This difference in learnability is also in evidence here. Recall that in order to control for differences in learning of the pronominal bipartitions, only participants who successfully learn the relevant pronominal partition for their condition go on to complete the locative phase. 89% of participants met our criterion in Condition 1 (participant bipartition), while only 61% met this criterion in Condition 2 (random bipartition). This suggests that English-speaking learners distinguish among non-native bipartitions, and in particular, find unnatural bipartitions more difficult to learn.

After controlling for this effect of pronominal bipartition learning, failure to find any effect of pre-training therefore suggest that learners may generalize across domains when prior knowledge is based on which features are relevant to determining homophony. Put another way, they generalize from a person partition to a spatial one only when the relevant person partition is a natural one. The result in this case, is that regardless of training condition, participants infer a participant locative bipartition rather than a random one.

5 General Discussion

Theories of indexicality have been built on different sources of evidence, leading to conflicting assumptions regarding the semantic primitives underlying personal and spatial deixis. Based on typological evidence suggesting similarities in attested partitions of pronominal and spatial indexicals, Harbour (2016) posits a set of shared features and thus a unified treatment of indexicals. However, this view is challenged by evidence coming from indexical shifting languages (Anand, 2006; Anand & Nevins, 2004; Deal, 2020), where the two types of indexicals do not appear to behave in a uniform way. Here we have presented a series of behavioral experiments aimed at assess whether there is evidence from learning that person and space deixis are represented in terms of a common semantic basis.

In the artificial language learning experiments reported above, English-speaking participants were taught either a pronominal or a locative system in a new language, and were tested on whether they extended the contrasts learned in one domain to the other. We predicted that if learners represent personal and spatial indexicals using the same set of semantic primitives, then they would generalize a new indexical system from one domain to the other. Put another way, we treat the likelihood of extrapolating between domains as an indication of the semantic similarity between person and locative systems.

In Experiment 1, we found that learners indeed generalize a learned non-native pronominal system to a locative system. This supports the hypothesis that the same features that define personal pronouns can be used to represent locatives, in line with Harbour (2016). In this theory, a participant bipartition, for example, is built on a single $\pm$PARTICIPANT feature, regardless of whether it applies to the individuals themselves or to the space close to them. Participants in Experiment 1 would thus be exploiting the common features between person and spatial indexicals: they infer that if a novel pronominal system features a certain set of contrasts, this will be reused in the locative system of that language. This inference can be thought of as a way of minimizing the number of contrasts in the language, in line with a general simplicity bias in learning (Culbertson & Kirby, 2016). Our results could alternatively be seen as a kind of *priming* effect (McNamara, 2005), whereby the exposure to a certain pronominal partition (set of contrasts) *primes* or facilitates the use of the same partition in a new locative system. Since priming only occurs between structurally or semantically related stimuli, interpreting our results as evidence of a general learning bias or as a priming effect would not make a crucial difference for our purposes: in any case it would reflect shared primitives.[17]

The fact that participants in Experiment 3 do not extrapolate an unnatural *random* pronominal bipartition to a locative system provides additional evidence for the role of primitive features in

[17]Note that both learning biases and priming effects may be quite weak; for example, learning biases can be overcome with additional evidence from the input, and priming is not necessarily long-lasting (see Bock & Griffin, 2000; Branigan et al., 1999; Chang et al., 2006, for discussion). However, there is evidence from computational models of language evolution that weak biases can come to have relatively strong effects on typological distributions over the course of generations (Thompson et al., 2016).

learning and extrapolating indexicals. In previous work, Maldonado and Culbertson (In press) have found that random bipartitions of the pronominal space are more difficult for learners to acquire. Here we show that even when such a system *is* acquired (recall that participants must successfully learn the pronominal bipartition to advance to the locative phase of the experiment), it is not used to infer a locative partition. What drives learners' generalization patterns in Experiment 1 thus appears to be the possibility of using analogous *features* to organize the two domains. When there is no feature-based generalization possible, learners do not simple extrapolate superficial homophony patterns.

Of course, the results of Experiment 2 suggest a more complex picture. While learners extrapolate from pronominal to locative systems, suggesting that shared primitives *can* be used to make inferences in learning, they do not extrapolate in the opposite direction, i.e., from locative to pronominal system. In §3 above, we suggested that this difference could be related to an asymmetry in how personal and spatial indexicals are related to one another. We presented there an intuitive example showing that while locatives can easily be defined in terms of person primitives, it is not straightforward to define personal pronouns in terms of locative primitives. This intuitive asymmetry is to some extent already presupposed in Harbour's approach to indexical systems. The idea behind Harbour's claim that deixis of person and space 'shares a common semantic basis' (Harbour, 2016, p.170) is that locatives are defined in terms of *person* features. As a response to this claim, the indexical shifting evidence in Deal (2019) is explicitly used to support the existence of a specific *location* primitive (whereas the existence of at least one person primitive is taken for granted).[18] In what follows, we sketch two alternative accounts of our results that exploit how this intuitive asymmetry has been integrated into existing theories.

One possible account of our results combines Harbour's proposal with standard assumptions about the characterization of spatial indexicals. Under this view, there is only one way of defining person indexicals, whereas locatives can be represented either by the same primitives that define personal pronouns (e.g., as 'space close to the speaker') or by a specific location primitive (e.g, as 'the location of the utterance'). Thus, learners have a choice with respect to how they organize locative systems: they can either deploy ±PARTICIPANT and ±AUTHOR features or some distinctive location-based feature. In Experiment 1, exposure to a person system encourages (or primes) learners to represent locatives through person features. Instead, when learners are pre-trained on a locative system in Experiment 2, the choice between these two ways of characterizing locative systems remains unconstrained. Only learners who represent locatives in terms of person features would be expected to extrapolate the locative partition to a pronominal system. As a result, the influence of the pre-training is weaker in Experiment 2 than in Experiment 1.

However, the effect of locative pre-training in Experiment 2 was not simply *weaker* than in Experiment 1, instead there was no difference between learners' inferences after this training and their inferences in the baseline condition, where there was no pre-training. While this could reflect a very weak effect that we were not able to detect, it is on the face of it somewhat problematic

[18] It's worth noting that an asymmetry between personal and spatial indexicals is also attested in the typology of indexical shifting languages. As observed in Fn. 7, locatives shift in a certain language only if person indexicals shift. Some theories have explained this implicational hierarchy by positing a fixed functional sequence for shifty operators (Deal, 2017, 2020), whereby the operator that shifts locatives can only be stacked hierarchically above an operator that shifts person indexicals. The results reported here suggest that this asymmetric relation, by which the status of locatives might somehow depend on the characterisation of person indexicals, may be encoded elsewhere in the grammar, not only as a functional sequence.

for the account just sketched. One possibility is to additionally assume that these two alternative representations of locative expressions differ in terms of naturalness, such that, in absence of a bias, English-speakers would be more likely to rely on a location-specific representation of spatial indexicals than on a person-based one. This would explain why we fail to observe an effect of locative pre-training in Experiment 2.

Before moving towards a second possible account of our results, let us note that this first proposal has the potential of making interesting cross-linguistic predictions. In particular, this account assumes that the two alternative representations of locatives are available for *English* speakers, leaving open the possibility that they may not be available for speakers of other languages. Indeed, languages may differ with respect to whether they make use of both person- and location-based representations or only one of them. Importantly, parametrizing the availability of these two representations of locatives could account for the indexical shift patterns discussed in §1.2, while retaining some of the typological predictions made by Harbour (2016). For example, languages where spatial and person indexicals do not shift together (e.g., Uyghur) might represent locatives strictly in terms of a distinctive location-based feature. If this were the case, this variation may have repercussions for learning a new language. For instance, we would predict distinct behavior in our experiments if learners are speakers of an indexical shifting language such as Uyghur or Nez Perce.

There is a second possible account of our results that is also worth entertaining. Namely, the difference between experiments might be a product of the fact that spatial indexicals have a richer structure than person indexicals. According to Harbour (2016), locatives are built on person indexicals plus some supplementary structure, which contributes the meaning of 'space close to' or 'the vicinity of'. As observed in Footnote 2, this additional piece of content is encoded in Harbour (2016) in a syntactic head χ, which scopes over person features. For example, for a pronoun defined as $+$PARTICIPANT, the corresponding locative would be defined as $\chi(+$PARTICIPANT$)$. If this additional structure in the representation of spatial indexicals (i.e. the presence of the χ operator) matters for learning, this could potentially drive our results. Specifically, learners might find it difficult to extrapolate from *embedded* features. That is, participants in Experiment 2 may not extend either $\pm$AUTHOR nor $\pm$PARTICIPANT contrasts from locative to person systems precisely because these features are embedded under an operator, which makes them "invisible" for extrapolation purposes. By contrast, Experiment 1 involves extending the unembedded features defining personal pronouns to embedded positions in locatives.

To summarize the two potential accounts we have proposed to account for our results: the first posits a specific location primitive at the cost of modifying Harbour's original proposal, the second is compatible with Harbour (2016) as is, but requires further assumptions linking generalization in learning of the embedded status of features.[19] While the current results do not definitively support one of these accounts over the other, there are some testable predictions which could help adjudicate between them. The first proposal predicts possible differences between speaker populations (e.g., based on a parameterized possibility of locative representations). In contrast, the second proposal makes an experimental prediction regarding other classes of indexicals (e.g.,

[19] By adopting Harbour's theory, one leaves unexplained the indexical shifting patterns mentioned in §1.2. The way we see it, explaining these patterns under Harbour's view might require abandoning the idea that shifting operates on specific contextual parameters. Instead, one might adopt an account of shifting that relies on idiosyncratic differences at the level of the indexical entries (e.g., a shiftable first person indexical in some languages but not in all of them; see Anvari, 2019). A full proposal along these lines is beyond the scope of this paper.

temporal): the same asymmetry is expected to arise for any two classes of indexicals where one is defined in terms of the other.

Lastly, it is worth saying a few words about how our results relate to the typology of person and locative systems. Our study was built on the assumption that, if the same feature underlie personal and spatial deixis, then learners should show a bias towards uniformly organizing both systems. A priori, one would expect this bias to show up in the typology: that is, one might expect a tendency for languages to show the same partitions in the locative and in the person domain (see Footnote 10). Whether this is in fact the case remains to be determined. However, in light of our results, one might want to revise this predictions. In particular, given the *asymmetric* nature of the learning bias in our study, one might no longer predict a uniform organization across domains. Instead, there may be an implicational cross-linguistic tendency regarding the contrasts made in person and locative systems within a language. For example, locative systems may make systematic use of a subset of the features at play in person systems. This is attested in English, which makes more distinctions in the person than in the locative domain.

6 Conclusion

Theories of indexicals differ in whether they treat personal and spatial indexical as defined by the same primitives. Typological evidence suggests similarities among the attested partitions of pronouns and locatives, leading Harbour (2016) to posit a single set of primitives shared across both domains. By contrast, evidence from indexical shifting suggests that these two types of indexical are distinct in important ways (Deal, 2019). Here we investigated the hypothesis that these two domains share primitives by testing whether when exposed to a new language, learners assume that the same features underlie both kinds of indexicals. We found that learners have a bias towards organizing a new locative system by using known person distinctions, but do not generalize locative distinctions to a new pronominal system.

Our findings do not provide unequivocal evidence with respect to whether different semantic primitives for person and location are required. However, the fact that locative systems can be characterized by similar properties to the ones that define personal pronouns suggests that person and spatial deixis *can* indeed have a common semantic basis, at least for English-speaking learners. Our experiments therefore support a theory of indexicality that allows for the possibility that person features play a role in the representation of locative indexicals.

References

Anand, P. (2006). *De de se* (PhD Thesis). Ph. D. thesis, MIT Cambridge, MA.

Anand, P., & Nevins, A. (2004). Shifty operators in changing contexts, In *Semantics and Linguistic Theory*.

Anderson, S. R., & Keenan, E. L. (1985). Deixis. *Language typology and syntactic description*, *3*(4), 259–308.

Anvari, A. (2019). *The Ban Against Illeism and Indexical Shift in Farsi*. Retrieved September 2, 2020, from https://ling.auf.net/lingbuzz/004823

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning. *Journal of Experimental Psychology: General*, 177–192.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (1999). Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review*, *6*(4), 635–640.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), pmid 12517354, 19–22. https://doi.org/10.1016/s1364-6613(02)00005-0

Clark, R., & Roberts, I. (1993). A Computational Model of Language Learnability and Language Change. *Linguistic Inquiry*, *24*(2), jstor 4178813, 299–345.

Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, *111*(16), 5842–5847. https://doi.org/10.1073/pnas.1320525111

Culbertson, J. (2012). Typological Universals as Reflections of Biased Learning: Evidence from Artificial Language Learning: Typological Universals as Reflections of Biased Learning. *Language and Linguistics Compass*, *6*(5), 310–329. https://doi.org/10.1002/lnc3.338

Culbertson, J., & Kirby, S. (2016). Simplicity and Specificity in Language: Domain-General Biases Have Domain-Specific Effects. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01964

Cysouw, M. (2003, July 3). *The Paradigmatic Structure of Person Marking*. Oxford, UK, OUP Oxford.

Dautriche, I., Chemla, E., & Christophe, A. (2016). Word Learning: Homophony and the Distribution of Learning Exemplars. *Language Learning and Development*, *12*(3), 231–251. https://doi.org/10.1080/15475441.2015.1127163

Deal, A. R. (2017). *Shifty asymmetries: Universals and variation in shifty indexicality*. http://linguistics.berkeley.edu/~ardeal/papers/Deal-indexicals-2017.pdf

Deal, A. R. (2019). Person features and shiftiness. The Alphabet of Universal Grammar, London, UK. https://ling.auf.net/lingbuzz/005433

Deal, A. R. (2020). *A theory of indexical shift*. MIT Press. Retrieved March 1, 2020, from http://linguistics.berkeley.edu/~ardeal/papers/Deal-theory-of-indexical-shift.pdf

Finley, S., & Badecker, W. (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language*, *61*(3), 423–437. https://doi.org/10.1016/j.jml.2009.05.002

Harbour, D. (2016). *Impossible persons*. Cambridge, MA, MIT Press.

Harley, H., & Ritter, E. (2002). Person and Number in Pronouns: A Feature-Geometric Analysis. *Language*, *78*(3), 482–526. https://doi.org/10.1353/lan.2002.0158

Hupp, J. M., Sloutsky, V. M., & Culicover, P. W. (2009). Evidence for a domain-general mechanism underlying the suffixation preference in language. *Language and Cognitive Processes*, *24*(6), 876–909. https://doi.org/10.1080/01690960902719267

Imai, S. (2003). *Spatial Deixis*. State University of New York at Buffalo. Buffalo, NY.

Kaplan, D. (1989). Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives. *Themes from Kaplan. Oxford*, 481–563.

Linzen, T., & Gallagher, G. (2017). Rapid generalization in phonotactic learning. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1), 24. https://doi.org/10.5334/labphon.44

Maldonado, M., & Culbertson, J. (In press). Person of interest: Experimental investigations into the learnability of person systems. *Linguistic Inquiry*, 1–71. https://doi.org/10.1162/ling_a_00406

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, *57*(4), pmid 1518508, 1–182.

McNamara, T. P. (2005). *Semantic Priming: Perspectives From Memory and Word Recognition*. Psychology Press.

Myers, S., & Padgett, J. (2014). Domain generalisation in artificial language learning. *Phonology*, *31*(3), 399–433. https://doi.org/10.1017/S0952675714000207

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological bulletin*, *134*(3), 427.

Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science*, *275*(5306), 1604–1610.

Rabaglia, C. D., Maglio, S. J., Krehm, M., Seok, J. H., & Trope, Y. (2016). The sound of distance. *Cognition*, *152*, 141–149. https://doi.org/10.1016/j.cognition.2016.04.001

Saffran, J. R., & Thiessen, E. D. (2003). Pattern Induction by Infant Language Learners. *Developmental Psychology*, *39*(3), 484–494.

Schlenker, P. (2003). A Plea For Monsters. *Linguistics and Philosophy*, *26*, 29–120.

Song, J. (Ed.). (2011). *The Oxford handbook of linguistic typology*. Oxford University Press.

Sudo, Y. (2012). *On the Semantics of Phi Features on Pronouns*. MIT. Cambridge, MA.

Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, *113*(16), pmid 27044094, 4530–4535. https://doi.org/10.1073/pnas.1523631113

Traxler, M. J., Tooley, K. M., & Pickering, M. J. (2014). Syntactic Priming During Sentence Comprehension: Evidence for the Lexical Boost. *Journal of experimental psychology. Learning, memory, and cognition*, *40*(4), pmid 24707789, 905–918. https://doi.org/10.1037/a0036377

Von Stechow, A. (2002). Binding by verbs: Tense, person and mood under attitudes, In *Proceedings of NELS*.

White, J. (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition*, *130*(1), 96–115. https://doi.org/10.1016/j.cognition.2013.09.008

Wilson, C. (2006). Learning Phonology With Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science*, *30*(5), 945–982. https://doi.org/10.1207/s15516709cog0000_89

Woodworth, N. L. (1991). Sound symbolism in proximal and distal forms. *Linguistics*, *29*(2). https://doi.org/10.1515/ling.1991.29.2.273

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological review*, *114*(2), 245.

Yang, C. (2002). *Knowledge and learning in natural language*. Oxford University Press on Demand.

Yang, C., & Montrul, S. (2017). Learning datives: The Tolerance Principle in monolingual and bilingual acquisition. *Second Language Research*, *33*(1), 119–144. https://doi.org/10.1177/0267658316673686