Learnability as a window into universal constraints on person systems*

Mora Maldonado and Jennifer Culbertson

Centre for Language Evolution
University of Edinburgh, Edinburgh, UK
mora.maldonado@ed.ac.uk

Abstract

Zwicky (1977) made the following observation regarding the cross-linguistic distribution of person systems: Languages that do not have a dedicated form for the inclusive meaning (speaker+addressee) always assimilate it into the first person and never into the second or third. Different theories of person have been put forward to explain this asymmetry, positing different kinds of semantic constraints on the human capacity to categorize the person space. Here we use an Artificial Language Learning methodology to investigate whether learners are sensitive to these typological asymmetries. Our results reveal that (1) learners prefer person systems where there is homophony between inclusive and first person meanings; and (2) they find it easier to learn that the inclusive is a form of 'you' than a form of 'them'. Given that second-inclusive and third-inclusive homophony patterns are both absent in the typology, our findings suggest that not all typological tendencies are built equal: while some might be seen as the result of strong, universal cognitive constraints on grammar, others should be modeled as weaker learning biases.

1 Introduction

In a classic paper from 1977, Zwicky made the following observation regarding the cross-linguistic distribution of person systems: Languages that do not have a dedicated phonological form for an inclusive person (speaker+addressee) always assimilate it into the first person and never into the second or third. That is: in these languages (e.g. English), the *you and us* inclusive meaning is expressed as a form of 'us', and never as a form of 'you' (or 'them').[1] The basic distinction between the typological patterns is illustrated in Table 1.

At first glance, Zwicky's generalization is quite surprising. Most feature-based approaches to person systems (e.g. binary features account in Table 1, see [5] for discussion) assume that the inclusive person shares features with both the first exclusive (e.g. +speaker) and the second exclusive (e.g. +addressee). Indeed, a number of languages have inclusive pronouns that can be morphologically decomposed into first plus second forms (e.g., Bislama, 11), suggesting that there is (at least at some level) a semantic overlap between second and inclusive categories.[2] This leads naturally to the expectation that languages should be as likely to assimilate the

*This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 757643).

[1] Zwicky's generalization holds for languages taken as a system and not for individual paradigms within a language, which might show accidental homophony. For example, one can find languages in which some paradigms collapse inclusive and second persons to the exclusion of the first exclusive (and third); see [12] and examples therein. None of these languages, however, systematically group the inclusive and the second person in all its paradigms; instead, they have some paradigms in which the two differ. This is crucially different from *true* three-person systems like English, where first person collapses exclusive and inclusive in *all* its paradigms.

[2] Note that there are also inclusive languages where the verbal agreement of the inclusive form is shared with the second person and not with the first (e.g. Ojibwa pronouns, [13] from [20]).

1_{ST}	+speaker, –addressee
INCL	+speaker, +addressee
2_{ND}	–speaker, +addressee
3_{RD}PL	–speaker, –addresee

Binary-features account of person systems

(a) First-Inclusive	(b) Second-Inclusive	(c) Third-Inclusive
1_{ST}PL (grayed)	1_{ST}PL	1_{ST}PL
INCL (grayed)	INCL (grayed)	INCL (grayed)
2_{ND}PL	2_{ND}PL (grayed)	2_{ND}PL
3_{RD}PL	3_{RD}PL	3_{RD}PL (grayed)
attested	**unattested**	**unattested**

Table 1: Illustration of Zwicky's observation. Grayed cells are mapping to a single pronominal form, white cells to different and distinct forms.

inclusive with the second person as they are to assimilate it with the first. In contrast, no theory would predict the inclusive meaning to be homophonous with the third person, as the inclusive and the third person do not have any features in common (although see discussion of Tupinamba in [18] for a potential exception).

A number of theories of person have been developed with an eye toward accounting for Zwicky's observation. These theories take one of two general approaches: either maintain the traditional set of features (e.g. Table 1), but add default feature specifications, or posits a different set of features. For example, Harley and Ritter [13] take the first approach. Following ideas from phonology [7], they put forward a universal feature geometry for person, with primitives similar to the binary-features account but also hierarchical relations between them (see also 4, 15, 8 for similar approaches). In their system, the primitive features Speaker and Addressee are dependent nodes of the feature Participant. However, they specify the Speaker feature as less marked than the Addressee feature (see Figure 1). Consequently, in languages without an inclusive distinction, a preference for assimilating the inclusive meaning into the first person is expected, as they share the default feature. Defaults can be overridden, therefore the second-inclusive homophony pattern can still arise in contrast to a third-inclusive system which is predicted to be impossible.

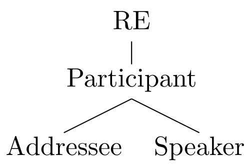


Figure 1: Feature geometry account in Harley and Ritter (2002)

The second approach is adopted by Harbour [12], who posits a different set of binary features, ±author and ±participant. While the features themselves denote lattices of possible entities (speaker, addressee, etc.), the values of the features are modelled as complementary operations on lattices. Because features are similar to functions, languages can differ not only in which features are active, but also in the order of feature composition.[3] Importantly, the absence of a ±addressee feature in the system creates an inherent asymmetry between speaker and addressee discourse roles. This asymmetry is essential to derive Zwicky's observation as a *strong* constraint on possible person systems: Systems showing systematic homophony between the first and inclusive person are derived, but homophony between inclusive and second or third person are both impossible. For example, a language which makes use of the ±author feature only will have a bipartition of the person space in which first and inclusive persons are homophonous, and second and third are homophonous. A language in which both ±author and ±participant are active, with the ±participant composing last, will have a first-inclusive tripartition (indeed this is the only tripartition of the space generated by this theory). A language where ±participant feature composes first will have not a second-inclusive, but a quadripartition of the space (see [12] for details), where each category is mapped into a unique phonological form. Without a corresponding ±addressee feature, there is no way to have any partition which picks out the set of categories including the addressee. [4]

The theories outlined above account for Zwicky's observation, but differ critically in how second-inclusive and third-inclusive are treated. Under Harley and Ritter's system [13], third-inclusive is singled out as underivable, while second-inclusive is possible but more marked than first-inclusive. By contrast, the system proposed by Harbour [12] takes as its starting point the idea that only first-inclusive can be generated by the grammar. Based on the typology alone, it is impossible to adjudicate between these theories: both second- and third-inclusive patterns are unattested. Further neither theory provides an explicit mechanism for linking the feature-based representations (and operations) they posit to typology. The implicit link is *learnability*; only a subset of possible person partitions are learnable by humans, or alternatively, some are learned readily while others are more difficult (e.g., can be learned but require substantially more evidence). However, there are many factors that can shape typological distributions—from learning biases to historical accidents, to genetic relations between languages, to pathways of diachronic change. In the domain of person, there is currently no evidence that learning is the crucial driving force.

The present paper follows a growing body of work using artificial language learning to provide new sources of behavioral data to test theoretical claims made on the basis of typological data (for reviews see [9, 10, 16]). Here, we investigate learners' sensitivity to predicted asymmetries among non-inclusive paradigms. We train English-speaking learners on an artificially-constructed personal pronoun paradigm in which the inclusive is a form of 'us' (first-inclusive), a form of 'you all'(second-inclusive) or a form of 'them'(third-inclusive). Participants are taught one of these three paradigms and are tested on how accurately they are able to learn it (within a set number of trials).

Given that English features first-inclusive homophony *and* it is the only tripartition sys-

[3]For the sake of brevity, here we are simplifying Harbour's account. More specifically, assume a person space that contains a speaker (i), an addressee (u) and an undefined number of others (o) as basic entities in the ontology. Harbour's proposed ontology consists on *egocentrically* nested subsets, such that the smallest subset in the ontology contains the speaker alone: $\{i\} \subset \{i, iu, u\} \subset \{i, iu, u, o\}$. The ±author and ±participant features will then denote part of this ontology by denoting the semi-lattices $\{i\}$ and $\{i, iu, u\}$ respectively. When the author feature is set up in its positive value, it will denote a semi-lattice containing only those elements containing i (i.e. $\{i, iu\}$). Otherwise, it will denote those elements that *exclude* the speaker: $[\![$ -author $]\!] = \{i, iu, \}$

[4]An intermediate proposal can be found in Ackema and Neeleman's system [3]. In their proposed feature structure, "there is no natural class (...) that comprises the first person inclusive and the second person, but not the first person exclusive" (p.910,[2]). However, second-inclusive patterns can still be obtained by incorporating an impoverishment rule in the system. This is not possible for inclusive-third homophony, creating an asymmetry between the two unattested patterns.

tematically attested in the typology, learners are predicted to prefer such paradigms over the alternatives. This mainly serves as a sanity check that participants understand the task and are able to learn a language that has the same structure as their own. Regarding second-inclusive and third-inclusive homophony patterns—both unattested in the typology—the accounts outlined make different predictions. If both of these patterns are directly ruled out by the grammar (*à la* Harbour), learners should be equally unlikely to learn either of them. By contrast, if learners are sensitive to the semantic commonalities between the inclusive and the second person (e.g. + addressee), a second-inclusive system should be easier to learn than a third-inclusive one [13, 15]. This pattern of results would moreover suggest that any apparent asymmetry between first-inclusive and second-inclusive languages should not be encoded as a hard constraint on person systems (contra Harbour).

2 Methods

This experiment, including all hypotheses, predictions, and analyses, was preregistered https://osf.io/5h4m6.

2.1 Design

Participants were randomly assigned to one of three possible conditions, which correspond to the (a-c) patterns in Table 1. Participants in all conditions were taught three pronominal forms mapped into four *plural* person categories (first exclusive, inclusive, second exclusive, and third). Each condition instantiated a different form-to-meaning mapping: the pronominal system could assimilate the inclusive meaning into the first plural person (First-Inclusive condition), into the second plural person (Second-Inclusive condition), or into the third plural person (Third-Inclusive condition).

Participants in all three conditions were also exposed to three additional distinct pronominal forms corresponding to the first, second, and third *singular* persons. Participants' learning of these forms was used as an exclusion criteria (see below).

2.1.1 Materials

The language consisted of 6 different pronominal forms: 3 forms were used for the plural pronouns (critical categories), and 3 different forms were used for the singular pronouns (filler categories). For each participant, these 6 lexical items were randomly drawn from a list of 8 CVC non-words created following English phonotactics: 'kip', 'dool', 'heg', 'rib', 'bub', 'veek', 'tosh', 'lom'. Items were presented orthographically.

To express the pronoun meanings, we commissioned a cartoonist to draw scenarios involving a family of three sisters and their parents. Each family member has a clearly-defined role in the conversational context. The two older sisters are speech act participants (in all scenarios they are either speaker or addressee). The third (little) sister was spatially close, but never a speech act participant. The parents were seated in the background (serving as additional others).

Pronouns were used as one-word answers to questions like 'Who will be rich?'. Meanings were expressed by visually highlighting subsets of family-members, as in Table 2. In some cases, more than one pattern of visual highlighting could match the target meaning, options were then randomly selected. An example illustrating the 1^{st} plural trial is provided in Figure 2. All questions were English interrogative sentences of the form 'Who will...?', which were randomly drawn from a list of 60 different tokens.

Category	Highlighted set
1^{st}_{SG}	speaker
2^{nd}_{SG}	addressee
3^{rd}_{SG}	one other
1^{st}_{PL}	speaker, other(s)
INCL	speaker, addressee (other(s))
2^{nd}_{PL}	addressee, other(s)
3^{rd}_{PL}	multiple others

Table 2: Highlighted family members for each person category. 1^{st}, 2^{nd} and 3^{rd} plural categories randomly include one or two additional others; the inclusive category could refer to speaker and addressee alone or include as well one or two others.

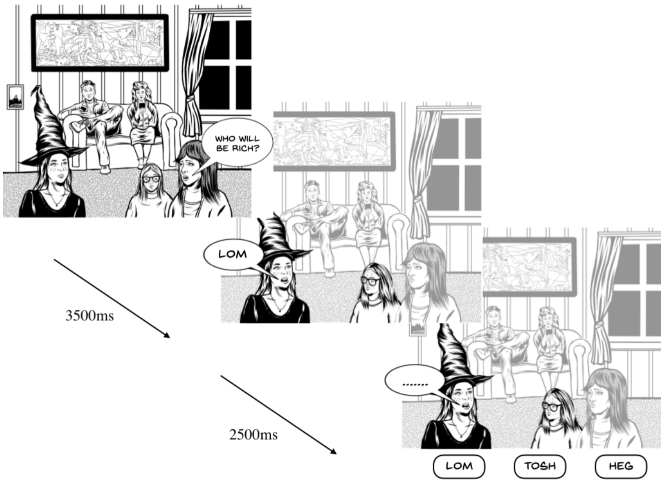
2.1.2 Procedure

Participants were first introduced to the family, including the names of the sisters, and were told 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. Participants were instructed to figure out the meanings of words in this new language. They were given a hint that the words were not names, but they always refer to people present in the scene. In addition, the speaker and addressee roles switched several times during the experiment to highlight that the words were dependent on contextually-determined speech-act role. This was induced by swapping who had the magical hat.

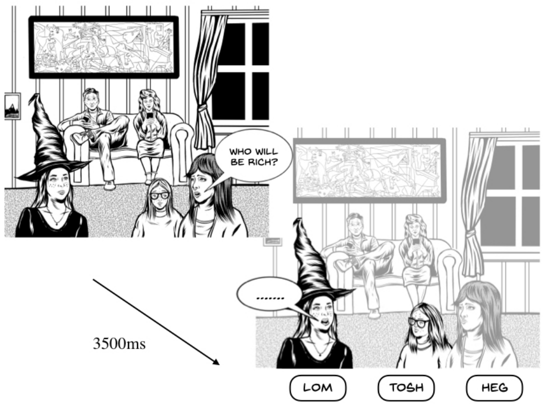
The experiment had two phases, each composed of exposure and testing blocks (e.g., **??**). Exposure trials (e.g., Figure 2a) had two parts: a scene where a question was asked, and a scene where the question was answered with a pronoun form in the language. To check that participants were paying attention, they were then asked to select among three different alternatives the one corresponding to the pronominal form they had just seen. Testing trials (e.g., Figure 2b) consisted of a question, just as in exposure trials, followed by a scene where a set of individuals was highlighted (reference of the pronoun) without presenting the pronoun form. Participants had to pick the correct word for that meaning among three different options.

During the first phase, participants were trained and tested on the three singular pronouns. There were a total of 12 exposure and 12 testing trials (4 repetitions per form/meaning). Participants who responded accurately to at least 2/3 of the testing trials in this phase (corresponding to 8 correct responses) passed to the second critical phase. In the critical phase, participants were exposed to and tested on the mapping between three plural pronouns and four person meanings. This phase was comprised of two alternating exposure and testing blocks. There were a total of 24 exposure trials (6 repetitions per meaning) and 48 testing trials (12 repetitions per meaning).

The complete experimental session lasted approximately 20 minutes. The order of presentation of meanings was fully randomized within exposure and testing blocks for each participant. Participants were given a debrief at the end of the experiment to check how they interpreted the forms they were trained on. For example, participants in the Second-Inclusive condition would describe the critical form as 'me and you or you all' or as 'group containing Ann or Mary', whereas participants in the First-Inclusive condition would just use the pronoun 'us'.



(a) Exposure



(b) Testing

Figure 2: Illustration of Exposure and Testing trials for the 1stPL category.

2.1.3 Participants

A total of 320 English-speaking adults were recruited via Amazon Mechanical Turk (First-inclusive group: 109, Second-inclusive group: 101, Third-inclusive: 110).[5] 167 participants responded accurately on more than 8 singular testing trials and were allowed to continue with the *critical* plural pronoun phase, according to our pre-registered plan (First-inclusive group: 57, Second-inclusive group: 55, Third-inclusive: 55). Participants who passed the singular test phase were paid 3.5 USD for their participation and 1 USD otherwise.

[5]The number of participants reported here does not include workers who were excluded for not being self-reported native speakers of English (10) and participants who fail to pass an attention check (AC) included at the very beginning of the experiment (35). This AC was added to exclude participants who had not read the instruction or were bots [19]. While these participants could in principle have been excluded based on their performance singular testing trials, the AC allowed us to filter them out in advance, distinguishing them from participants who just found the experiment hard.

3 Results

Mean accuracy rates on testing trials during the *critical* phase are given in Figure 3. The effect of Condition and Block on accuracy rates was analyzed using logit mixed-effect models in R [1].[6] The model included the maximal random effect structure, random intercepts per subject and slopes per block [following **?**]. The standard alpha level of 0.05 was used to determine significance, and p-values were obtained based on asymptotic Wald tests.

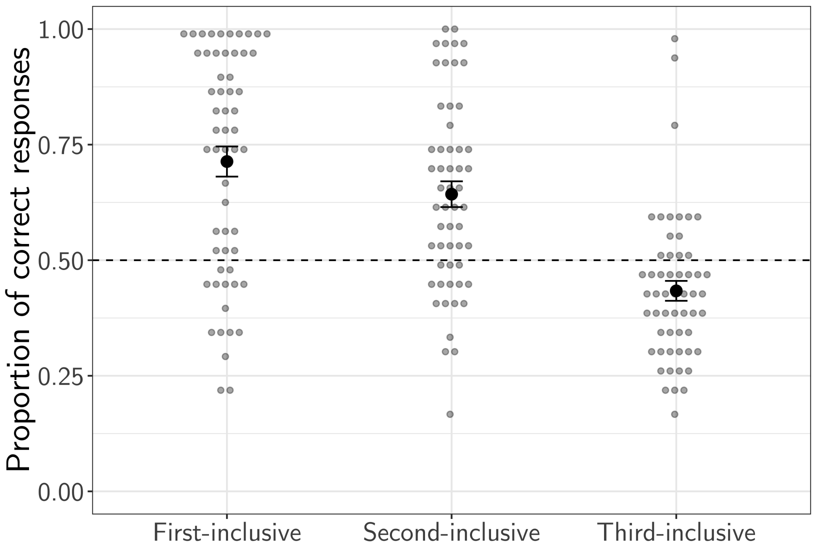


Figure 3: Accuracy rates in *critical* testing trials by condition. Error bars represent standard error on by-participant means; gray dots represent individual participant means.

We first compared First-Inclusive with Second-Inclusive and Third-Inclusive (contrasts were treatment coded, with First-Inclusive and Block 2 as baselines). The model revealed that the proportion of accurate responses in the First-inclusive condition (Block 2) was significantly above chance (intercept: $\beta = 1.59$; $p < .001$). In addition, accuracy rates in the First-inclusive condition were significantly higher than in the Third-inclusive condition ($\beta = -1.83$; $p < .001$), and marginally different from accuracy rates in the Second-inclusive condition ($\beta = -0.572$; $p = .055$).

A second model was fitted to test the difference in accuracy between Second-inclusive and Third-inclusive conditions (contrasts were treatment coded, with Second-inclusive and Block 2 as baseline). Accuracy rates in the Second-inclusive condition (Block 2) were found to be significantly above chance (intercept: $\beta = 0.97; p < .001$) and higher than the ones in the Third-inclusive condition ($\beta = -1.2; p < .001$).[7]

[6] There were two testing blocks in the critical phase, each preceded by an exposure block. Participants were generally expected to improve with accumulated exposure, but this improvement could vary across conditions. Each model included the effect of Block on accuracy, as well as the interaction with Condition. However, we report here only simple effects regarding the second testing block. The complete model output can be found in the OSF repository.

[7] Following our preregistration, we ran a second version of each of these models, restricting the analysis to inclusive meanings. The idea behind this move was that our hypotheses target specifically the inclusive category

4 Discussion

Our findings first confirm that English-speaking participants are most successful at learning a new language that features native-like homophony between inclusive and first person meanings. This result confirms that participants understand the task and are able to learn a language that has the same structure as their own.

If, as predicted by Harbour [12], there is additionally a hard constraint on tripartitions which derives only first-inclusive homophony, then we might expect the preference for English-like systems to be very strong indeed. However, the difference in accuracy between First-inclusive and Second-inclusive conditions was only marginally significant. This result is consistent with the idea that learners are sensitive to the semantic overlap between inclusive and second person, as predicted by [5, 13, 15]. In line with these accounts, participants in the Second-Inclusive condition may treat these as a natural class, relying on a shared feature, ±addressee, to learn this new language. This result supports a theory in which first- and second-inclusive patterns are both generated by the grammar, but the latter is just dispreferred (contra 12, and possibly 2, 3, see Footnote 4).

Importantly, we also found that learners have a bias against systems that assimilate the inclusive into the third person. This result reveals that second- and third-inclusive systems, despite being unattested in the typology, are not equal from a learnability perspective: there is a stronger pressure against third-inclusive than against second-inclusive homophony. This is again as predicted by theories like [5, 13, 15] (but arguably also by [3]) which posit that, unlike first and second person, third person differs from the inclusive in more than one feature and therefore does not form a natural class with it (cf., e.g., [18]). As a result, homophony between inclusive and third persons is not predicted to occur systematically, though it might arise accidentally.

Returning to Zwicky's observation, our findings suggest that the typological asymmetry between alternative non-inclusive systems may result from multiple interacting factors, not just learnability as tested in our experiment. In the typology first-inclusive systems are attested systematically, while second- and third-inclusive systems are not; in our experiment, third-inclusive systems were clearly dispreferred, but the advantage for first-inclusive over second was much weaker. While this is consistent with a theory positing weak learning biases (i.e., constraints that can be over-ridden given sufficient evidence) which penalize third-inclusive most, it still leaves the typological data partially unexplained.

One possibility is that there is an additional weak bias, not at play in our experiments, which further advantages first-inclusive systems. An obvious such candidate is a general *egocentric* bias, i.e., increased importance or salience of the speaker to him or herself. Indeed, a version of this bias has been argued to hold during early pronoun acquisition [6, 14, 17]. If individuals (and potentially children more so than adults) perceive the world as a function of their presence in it, then they may be more likely to adopt categorization systems which preserve this distinction. This would lead to an stronger asymmetry between first-inclusive and second-inclusive systems. There is good reason to suspect that any obvious egocentric bias is weakened in the context of our experiment, where participants are passive learners and do not themselves feature in the meanings they are learning—our learners are never themselves speakers. Future work will investigate whether a stronger egocentric bias, and therefore a stronger preference for first-inclusive paradigms, arises when the learner plays a more active role in the conversation.

Before concluding, let us entertain an (partial) alternative interpretation of our results.

(expressed by the ambiguous pronoun in each system). The output of these models follows the pattern of results described above.

Recall that Zwicky's generalization holds specifically for languages as a system and not for particular paradigms. In our experiment, we teach participants a pronominal system, which in principle may be just one paradigm of many in the language. Arguably, our participants might be treating the pronominal system we teach them as an instance of accidental syncretism within an inclusive language (i.e., a language that makes an inclusive/exclusive distinction). If this were the case, theories like Harbour's could still account for our results. However, given that our participants are speakers of a non-inclusive language, this seems very unlikely. Further, this would likely not predict the difference we find between second- and third-exclusive.

5 Conclusion

In this study, we used an artificial language learning paradigm to test the different predictions made by theories of person systems. These theories were designed to, among other things, explain Zwicky's generalization: when a language does not have a dedicated pronominal form for the inclusive, it will assimilate the inclusive with the first person (not the second or the third). We targeted two influential approaches which differ in the set of universal person features they posit. According to Harley and Ritter [13], first-inclusive homophony is predicted to occur systematically, second-inclusive homophony is marked but possible, and third-inclusive homophony is impossible. For Harbour [12], only first-inclusive homophony can be generated by the grammar. Our results revealed that English-speaking learner have a slight preference for first-inclusive over second-inclusive paradigms, but clearly disprefer third-inclusive. Our findings support theories like Harley and Ritter's, which posit semantic overlap between second and inclusive persons which makes homophony between these categories possible. Thus apparent regularities in typology are not necessarily the result of strong constraints on possible grammars (*à la* Harbour). Here we have argued that (morpho-)semantic universals such as Zwicky's observation might be best explained in terms of weak learning biases (i.e., sensitive to cross-category similarity, and the importance of the ego), which can be overcome by learners given enough evidence.

References

[1] R: A language and environment for statistical computing. 2013.

[2] Peter Ackema and Ad Neeleman. Person features and syncretism. *Natural Language & Linguistic Theory*, 31(4):901–950, November 2013.

[3] Peter Ackema and Ad Neeleman. *Features of Person: From the Inventory of Persons to Their Morphological Realization*, volume 78. MIT Press, 2018.

[4] Susana Bejar. *Phi-Syntax: A theory of agreement*. PhD thesis, University of Toronto, Toronto, 2003.

[5] Jonathan David Bobaljik. Missing persons: A case study in morphological universals. *The Linguistic Review*, 25(1-2), January 2008.

[6] Rosalind Charney. Speech roles and the development of personal pronouns. *Journal of child language*, 7(3):509–528, 1980.

[7] G. N. Clements. The Geometry of Phonological Features. *Phonology Yearbook*, 2:225–252, 1985.

[8] Elizabeth Cowper and Daniel Currie Hall. Argumenthood, pronouns, and nominal feature geometry. *Determiners: Universals and variation*, 147:97–120, 2009.

[9] Jennifer Culbertson. 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, May 2012.

[10] Jennifer Culbertson and Kathryn Schuler. Artificial Language Learning in Children. *Annual Review of Linguistics*, 5(1):353–373, January 2019.

[11] Daniel Harbour. Descriptive and explanatory markedness. *Morphology*, 21(2):223–245, May 2011.

[12] Daniel Harbour. *Impossible Persons*. Linguistic Inquiry Monographs. The MIT Press, Cambridge, MA, 2016.

[13] Heidi Harley and Elizabeth Ritter. Person and Number in Pronouns: A Feature-Geometric Analysis. *Language*, 78(3):482–526, 2002.

[14] Katherine A. Loveland. Learning about points of view: Spatial perspective and the acquisition of 'I/you'. *Journal of child language*, 11(3):535–556, 1984.

[15] Martha McGinnis. On Markedness Asymmetries in Person and Number. *Language*, 81(3):699–718, 2005.

[16] Elliott Moreton and Joe Pater. Structure and Substance in Artificial-phonology Learning, Part I: Structure. *Language and Linguistics Compass*, 6(11):686–701, 2012.

[17] Morgan Moyer, Kaitlyn Harrigan, Valentine Hacquard, and Jeffrey Lidz. 2-year-olds' comprehension of personal pronouns. In *Online Proceedings of the 29th Annual Boston University Conference on Language Development*, page 11, Boston, 2015.

[18] Aryon D Rodrigues. You and I = Neither You nor I: The personal system of Tupinamba. page 13, 1990.

[19] Steven V. Rouse. A reliability analysis of Mechanical Turk data. *Computers in Human Behavior*, 43:304–307, February 2015.

[20] Linda J. Schwartz and Timothy Dunnigan. Pronouns and pronominal categories in Southwestern Ojibwe. *Pronominal systems*, pages 285–322, 1986.