Something about *us*: Learning first person pronoun systems

Mora Maldonado (Mora.Maldonado@ed.ac.uk)
Centre for Language Evolution
University of Edinburgh

Jennifer Culbertson (Jennifer.Culbertson@ed.ac.uk)
Centre for Language Evolution
University of Edinburgh

Abstract

Languages partition semantic space into linguistic categories in systematic ways. In this study, we investigate a semantic space which has received sustained attention in theoretical linguistics: person. Person systems convey the roles entities play in the conversational context (i.e., speaker(s), addressee(s), other(s)). Like other linguistic category systems (e.g. color and kinship terms), not all ways of partitioning the person space are equally likely. We use an artificial language learning paradigm to test whether typological frequency correlates with learnability of person paradigms. We focus on first person systems (e.g., 'I' and 'we' in English), and test the predictions of a set of theories which posit a universal set of features (±exclusive, and ±minimal) to capture this space. Our results provide the first experimental evidence for feature-based theories of person systems.

Keywords: artificial language learning; categorization; person systems; extrapolation; typology; linguistic universals

Introduction

One of the fundamental goals of cognitive science is to understand how human languages carve up semantic space into linguistic categories. Research on the typology of categorization systems, from colour names, to noun classification and kinship terms suggests that not all systems are equally likely.

For example, despite some cross-linguistic variation, certain ways of carving up the continuous color space into linguistic categories are much more common than others. This has been argued to provide evidence for a universal basis for color categorization, reflecting properties of the human perceptual system (Kay & Regier, 2007; Zaslavsky, Kemp, Tishby, & Regier, 2018; Gibson et al., 2017). Similar arguments have been made to explain the distribution of kinship systems across languages (Kemp & Regier, 2012; Kemp, Xu, & Regier, 2018).

Here, we focus on a semantic space which has garnered substantial attention in theoretical linguistics: person systems (e.g., Zwicky, 1977; Harley & Ritter, 2002; Harbour, 2016; Ackema & Neeleman, 2018). Such systems–exemplified in pronoun paradigms (e.g. 'me', 'you', 'her')–describe how languages categorize entities as a function of their role in the context of a speech event (i.e., speaker(s), addressee(s), other(s)). Like color and kinship systems, person systems have long been observed to exhibit constrained variation.

The person space

Research on the typological distribution of person systems has hypothesized an inventory of four discrete categories: first exclusive (speaker only), first inclusive (speaker and addressee), second (addressee) and third (other) (Harley & Ritter, 2002; Cysouw, 2003; Bobaljik, 2008). The interaction with number multiplies the possible distinctions.

Here, we focus specifically on *first* person systems, as they allow us to investigate a contrast that is not instantiated by English (1^{st} inclusive vs. 1^{st} exclusive). Theories of first person systems have posited two binary features, one for person (±addressee) and one for number (±minimal) (Bobaljik, 2008; Cysouw, 2011; Harley & Ritter, 2002).[1] This two-feature system is designed to instantiate all first person categories, as illustrated in Figure 1.[2]

A language which takes advantage of the maximal 4-way contrast will have a person *paradigm* with 4 distinct forms (e.g., Ilocano pronouns). Alternatively, the contrast between some cells can be neutralized within a paradigm, in which case different cells will use the same form. Such paradigms exhibit *homophony*.

Homophony which neutralizes one of the two hypothesized features–person or number–has been called *systematic homophony* (Harbour, 2008; Baerman, Brown, Corbett, et al., 2005). For example, a paradigm that neutralizes only the person contrast (keeping the number one) would have just two pronominal forms, one for both minimal inclusive and exclusive, and another for

[1] The ±minimal feature encodes an asymmetry in status between the minimal group consisting of the speaker and addressee, and a larger group including others. This is used rather than the more intuitive singular/plural contrast to distinguish between the two inclusive categories.

[2] The two-feature system in Figure 1 is a simplification of current proposals for the complete person space (i.e. including 2^{nd} and 3^{rd} persons). Most approaches rely on the existence of at least two different person features and three number features (Bobaljik, 2008; Harbour, 2016; Bobaljik & Sauerland, 2018).

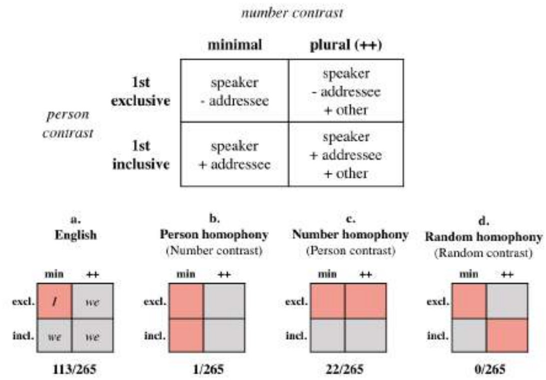


Figure 1: First person system (top) and four possible paradigms obtained by homophony (bottom, a-d) along with typological counts (Cysouw, 2003).

the two non-minimal (plural) categories ('Person homophony', Figure 1b). A paradigm that neutralizes only number (keeping the person contrast) would have one inclusive and one exclusive form ('Number homophony', Figure 1c). Homophony of both features is also possible, as in English ('we' for inclusive and exclusive plural, and minimal inclusive). Finally, a paradigm can partially neutralize one feature, for example, number homophony in the inclusive, but two distinct exclusive forms.

Random homophony patterns, not based on feature neutralization, are in principle also possible. For example, minimal exclusive and plural inclusive could share the same form, minimal inclusive and plural exclusive another ('Random homophony', Figure 1d).

Feature-based theories of person systems (cf. Figure 1) predict that systematic homophony is a natural consequence of feature-neutralization (or loss), and should arise regularly and be (easily) learnable. By contrast, they argue that there is no linguistic basis for random homophony, which is expected to arise only by historical accident, and be less readily learnable. Intuitively, there is nothing which ties together homophonous cells in a random homophony paradigm, therefore they should be less natural for learners. Notably, these theories are formulated on the basis of typological samples of person paradigms (the largest of which include <300 languages). Interestingly, while their predictions hold when considering complete paradigms, it is less clear for first person systems. According to Cysouw (2003), most of the possible paradigms for the 4-cell 1^{st} person space have not been documented. Among those that are attested, the skew is zipfian: the English-like pattern (Figure 1a) is by far the most frequent, the next most common systems have partial or complete number homophony (e.g., Figure 1c). Unexpectedly, both random *and* person (only) homophony appear to be very rare (see also Sauerland & Bobaljik, 2013; Baerman et al., 2005).

Experimental goals and predictions

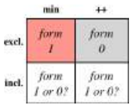
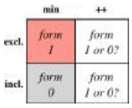
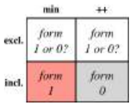
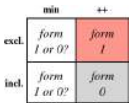
The principal goal of this paper is to set out a method for investigating person systems experimentally. The first step we take here is to test whether some first person paradigms are more natural than others. Our measure of naturalness will be learners' likelihood of inferring the relevant paradigm. We will test three main hypotheses: the first is a sanity-check, and the second two are derived from the theories outlined above in combination with the typology.

The first hypothesis is that, all things equal, learners generally assume a new language to have the same structure as their own. Learners in our experiment are native English speakers, therefore this predicts that they will be most likely to infer a first person paradigm that is English-like in its homophony pattern. The second hypothesis is that typologically frequency is correlated with learnability (Culbertson, 2018). This predicts that learners will be more likely to infer a paradigm characterized by number homophony than person or random homophony.[3] The third hypothesis is that there is a universal set of person/number features, as in (3), which learners are sensitive to regardless of their native language. This predicts that natural homophony patterns–which neutralize one specific feature–should be more likely to be inferred by learners than random homophony.

To test these predicted patterns of inference, we use an artificial learning paradigm in which learners are required to generalize (or extrapolate) from ambiguous evidence (a.k.a 'Poverty-of-the-Stimulus design, Wilson, 2006; Culbertson & Adger, 2014). Participants are trained on two cells of a first person paradigm, and must then use the forms they have learned to express all the cells in the paradigm. In other words, they must extrapolate the forms they have learned to the remaining two categories. For example, if a learner is trained on two distinct forms for exclusive minimal (speaker only) and exclusive plural (speaker plus others), they will be tested on the two remaining categories that include the addressee. If they use the plural form for both new categories, then they have inferred an English-like paradigm. Different patterns of extrapolation would indicate person or random homophony (as described in detail in Table 1).

[3] In principle this also predicts that learners should be most likely to infer an English-like paradigm, since this pattern of person *and* partial-number homophony is much more common. However, we cannot test this prediction with English-speaking learners.

Table 1: Summary of conditions.

Condition		Critical training set	Critical held-out set	Compatible paradigms
(1)		excl.min, excl.++	incl.min, incl.++	**English-like**, Person Hom., Random Hom.
(2)		excl.min, incl.min	excl.++, incl.++	**English-like**, Number Hom., Random Hom.
(3)		incl.min, incl.++	excl.min, excl.++	**Person Hom.**, Random Hom.
(4)		excl.++, incl.++	excl.min, excl.min	**Number Hom.**, Random Hom.

Methods

This experiment, including all hypotheses, predictions, and analyses, was preregistered.[4]

Participants

A total of 332 English-speaking adults were recruited via Amazon Mechanical Turk (female = 152). Participants were paid 2 USD for their participation which lasted approximately 15 mins. Per our pre-registered plan, participants were excluded if (a) their accuracy rates during exposure training were below 80%, or (b) their accuracy rates for trained cells during the test phase were below 66%. This resulted in analysis of 181 participants (Conditions 1: 46; Condition 2: 50; Condition 3: 49; Condition 4: 36).[5]

Design

Participants were randomly assigned to one of four possible conditions, summarized in Table 1. Conditions differed in which subset of two first person categories was trained (*critical training set*) and held-out (*critical held-out set*). This determines which alternative full paradigms are consistent with the two categories participants have learned. Conditions 1 and 2 are consistent with an English-like pattern (or systematic homophony). Conditions 3 and 4 are each consistent with one type of systematic homophony, and random homophony.

All participants were additionally exposed to another four pronominal forms which mapped into the second and third person singular and plural categories. These forms were used as controls.

Table 2: Highlighted family members for each category.

Category	Highlighted set
1st excl.min	speaker
1st incl.min	speaker, addressee
1st excl.pl	speaker, other(s)
1st incl.pl	speaker, addressee, other(s)
2nd sg.	addressee
2nd pl.	addressee, other(s)
3rd sg.	one other
3rd sg.	multiple others

Materials

The language consisted of 6 different pronoun forms, used for the control categories (2nd sg/pl, 3rd sg/pl), plus the critical first person forms. 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 highlighting subsets of family-members, as in Table 2.[6] An example illustrating 1st incl.min 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.

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, and an example trial with an English pronoun ('her').[7]

[4] Maldonado, M., & Culbertson, J. (2019, January 29). Extrapolation to bipartitions. https://doi.org/10.17605/OSF.IO/J2RCN.

[5] High accuracy rates on trained critical items were required because extrapolation of these forms is not interpretable if participants have not learned them.

[6] To ensure that forms were not associated with specific quantities, in all non-minimal categories, pronouns randomly referred to two or three individuals. Third person singular meanings were always expressed with a female other.

[7] In addition, the speaker and addressee roles switched during the experiment to highlight that the words were de-

Figure 2: Trial example in the test phase for the inclusive minimal category.

The experiment had two phases. In the training phase, participants were taught the pronouns in the control and training sets (6 person categories). Each training trial had two parts: a scene where a question is asked, and a scene where the question is answered with a pronoun form in the language (cf. Figure 2). There were 12 training trials (2 repetitions per form). Participants were given feedback on their answers.

After this initial training phase, participants were given an initial test of the trained forms. Each trial consisted of a question and answer scene, as in training, followed by a 'what if?' scene in which a new set of individuals was highlighted. They were asked to pick the correct word for that meaning among two options. There were 16 such trials (2 repetitions per control form, 4 per critical training form). Participants were given feedback on their answers. The test phase involved a similar procedure but included trials for the two remaining critical categories, i.e. the held-out set. This phase consisted of 48 trials (6 repetitions per form). Participants received no feedback during this phase.

The experimental session lasted approximately 15 minutes. The order of presentation of meanings was fully randomized within training and test phases for each participant.

Results

Recall that participants were taught two pronominal forms (coded as forms 1 and 0), which they had to use to describe both a critical trained set of first person meanings, and a held-out set. Figure 3 shows the proportion of trials on which participants chose the 'form 1' (pronoun) for each first person category during the test phase. Choice of the same form across categories indicates homophony. A visual inspection of Figure 3 suggests that participants in Conditions 1 and 2 are consistently using one form for 1st excl.min., and the other for the remaining three categories: this indicates inference of an English-like paradigm. Participants in Conditions 3 and 4 appear somewhat noisier in their responses, however, distinct patterns are evident. In Condition 3, one form is used for the two minimal categories, and the other for the plurals (consistent with person homophony). In Condition 4, one form is used for the two exclusive categories, and, at least for some participants, the other form is used for the two inclusive categories (consistent with number homophony).

Following our pre-registered plan, we conducted three analyses to evaluate these patterns statistically[8]:

Prediction 1: Preference for L1 pattern Figure 3 suggests that participants in Conditions 1 and 2 are more likely to infer a stable pattern, as predicted if an L1-like pattern is easier to learn. To test this, we used *joint entropy* of the two held-out categories to measure how variable participants' are in their mapping of the taught forms for the two held-out categories. The entropy value for a given category indicates the degree of uncertainty or variability in the responses. The joint entropy will therefore reveal the level of variability or uncertainty for each of the two held-out categories, with higher joint entropy values for participants who are less consistent in their answers. We fit a simple linear regression model predicting joint entropy by Condition (4 levels, treatment coded, Condition 1 as baseline). No random effects were included in the model, as each participant had a single joint entropy value associated. As predicted, joint entropy rates were significantly higher for Conditions 3 and 4 (intercept= .28; vs. 3: $\beta = .639$

pendent on contextually-determined speech-act roles.

[8]All analyses used the lme4 package in R (Bates, 2010). The data and analyses script can be found here.

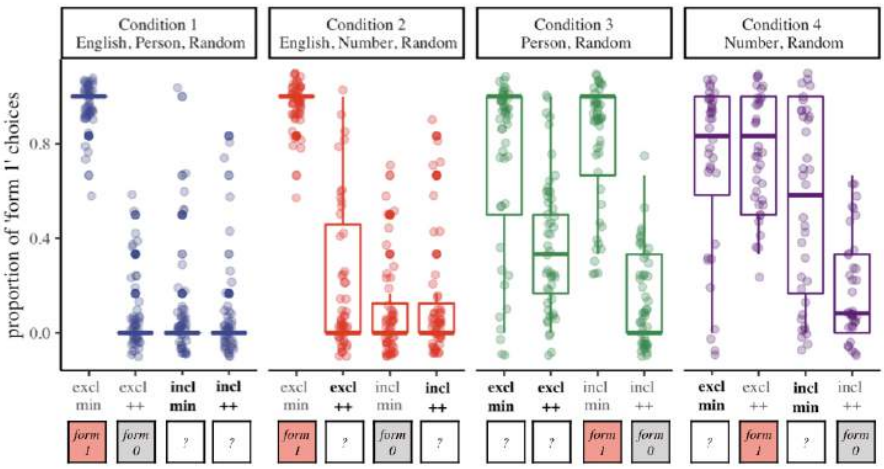


Figure 3: Proportion of 'form 1' choices for each first person category during the test phase. The held-out set for each condition is highlighted in bold-face. Choice of the same form (1 or 0) across categories indicates homophony. Dots are means of individual participants. Boxplots show by-participant means, quartiles, and range.

$\pm .12$, $p < .001$; vs. 4: $\beta = .447 \pm .13$, $p < .001$). Joint entropy rates for Condition 2 were marginally higher than Condition 1 ($\beta = .232 \pm .12$, $p = .055$).

Prediction 2. Preference for number over person homophony Figure 3 provides some evidence that participants in Conditions 3 and 4 are inferring paradigms with person and number homophony. Based on typological frequency, we predicted that number homophony should be more readily inferred than person homophony. If this is the case, then we should see a higher overlap in forms that share number in Condition 3 (columns in table 1) than forms that share clusivity in Condition 4 (rows). We measured this degree of homophony using the joint entropy between the relevant cells. For person homophony, we merged cells within a column and calculated joint column entropy. For number homophony, we merged cells within a row and calculated joint row entropy. The lower the joint entropy levels for a given homophony type, the more likely it is that participants are inferring a paradigm which neutralized that distinction. A simple linear regression model predicting joint entropy by Condition (2 levels, treatment coding, Condition 3 as baseline) revealed a marginally significant difference ($\beta = .207 \pm .11, p = .068$), with higher rates of person homophony (Condition 3) than number homophony (Condition 4). This fails to confirm our prediction.

Prediction 3. Preference for systematic over random homophony Finally, are participants in Conditions 3 and 4 in fact more likely to infer systematic rather than random homophony (as suggested by Figure 3)? To test this, the joint column/row entropy scores for systematic homophony computed above were compared to a random homophony score: the joint entropy of *all* alternative two-category combinations.[9] We ran separate mixed-effects models for Conditions 3 and 4, predicting entropy by homophony type (systematic vs. random) and including random intercepts per subject. We used likelihood ratio tests to compare these models to models with no fixed-effects. In both cases, entropy score for the systematic homophony pattern was significantly different from the random homophony score (person vs. random in Condition 3: $\chi^2 = 171.6$, $p < .001$; number vs. random in Condition 4: $\chi^2 = 84.4$, $p < .001$). This confirms that participants are more likely to use forms in a way that is consistent with systematic, not random homophony.

Discussion

In this experiment, we exposed English-speaking learners to sub-paradigms expressing person categories in a new language. We focused on first-person systems, which have been argued to have a universal basis in two features, encoding person and number. Participants were taught labels for two first person meanings, and asked to extrapolate to the two remaining meanings. We tested three hypotheses, designed to evaluate (1) whether learners were most likely to infer an English-like paradigm; (2) whether number homophony was more likely than person homophony (expected based on typological frequency); and (3) whether systematic homophony was more likely than random homophony

[9]For example, the joint column entropy in Condition 3 was compared to the joint entropy of each pair of diagonal and horizontal cells.

(predicted by feature-based theories).

Our results confirm that learners' are indeed highly likely to infer an English-like pattern when their training is consistent with this, producing systematic patterns of extrapolation from trained forms to new meanings. This result functions as a sanity check: it shows that participants are indeed understanding the stimuli in terms of a pronominal system.[10]

Our results also indicate that systematic homophony–which neutralizes either the number or person feature–is more natural than random homophony. This supports the claim that learners perceive the first person space as based on these two distinct features. Importantly, this finding cannot be accounted for solely based on experience with English. Inferring a person homophony pattern requires making a productive use of the ±minimal distinction. This is not the same number contrast made in English pronouns, which distinguish atomic (speaker only) and non-atomic entities (i.e., the more familiar singular/plural distinction). Similarly, inferring a number homophony pattern requires participants to learn and generalize the ±exclusive contrast, which is completely absent in English.

As for the typological difference between number (more common) and person (very rare) homophony, this does not appear to correlate with a learning difference in our task. Learners were, if anything, marginally more likely to infer paradigms characterized by person (Condition 3) rather than number homophony (Condition 4). One possibility is that, unlike random homophony, the rarity of person homophony in first person systems cross-linguistically is purely accidental, or reflects low sampling numbers. Indeed, person homophony is found for other parts of the person space (e.g., homophony of 1st and 2nd person in some languages). However, it may also reflect participants' experience of person homophony in English. Assuming that English encodes an atomic/non-atomic number distinction, it is possible to characterize English as a case of (only) person homophony (Harbour, 2016). In other words, English speakers have more experience with distinctions in number than in clusivisty. Indeed, a *posthoc* analysis shows that accuracy rates on trained categories (before exclusion) are higher in Condition 3 than 4 ($p < .001$), suggesting that the person distinction was harder to learn than the ±minimal distinction.

Finally, it is worth noting that differential sensitivity to person and number may also explain the marginal difference between Conditions 1 and 2. Both of these conditions allowed participants to generalize to an English-like paradigm, but they differed in whether a person or a number contrast was learned during the training phase (cf. Table 1). It could be that learning a new or unexpected distinction–between inclusive and exclusive minimal forms–led learners to be less likely to neutralize this feature in the plural.

[10]This is further confirmed by a debrief questionnaire, in which most participants reported having understood the new words as pronouns. For example, participants in Condition 4 have described the meaning of form 1 as 'Me or us not including you' and the meaning of form 0 as 'Us including you'.

Conclusion

In this study, we present the first experimental evidence for differences in learnability between alternative person paradigms. This was prompted by recent research in cognitive science on semantic spaces, and a lively literature in theoretical linguistics on the universal basis of person systems. We find, perhaps unsurprisingly, that English learners have a strong bias for first person paradigms that resemble their native language. They are more likely to infer paradigms analogous to English, and show a greater tendency to neutralize features that English also neutralizes (i.e. person). However, we also find that participants are sensitive to contrasts not found in their native language. Learners make productive use of both the ±minimal and ±exclusive distinctions, neither of which is present in English. Importantly, as predicted by feature-based theories of first person systems, learners were more likely to infer patterns which neutralized these features as compared to patterns in which featurally-unrelated cells were randomly homophonous. These initial results suggest that the paradigm we have developed can answer theoretically-motivated about how languages carve up the person space. Future work will target the full person paradigm, and incorporate recent insights about the potential role generally cognitive biases, such as simplicity, and communicative pressures like need probability (Kay & Regier, 2007; Kemp & Regier, 2012).

Acknowledgments

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

References

Ackema, P., & Neeleman, A. (2018). *Features of person: From the inventory of persons to their morphological realization* (Vol. 78). MIT Press.

Baerman, M., Brown, D., Corbett, G. G., et al. (2005). *The syntax-morphology interface: A study of syncretism* (Vol. 109). Cambridge University Press.

Bates, D. M. (2010). *lme4: Mixed-effects modeling with R Springer* (Tech. Rep.).

Bobaljik, J. D. (2008). Missing persons: A case study in morphological universals. *Linguistic Review*, *25*(1-2), 203–230.

Bobaljik, J. D., & Sauerland, U. (2018). *ABA and the combinatorics of morphological features. *Glossa: a journal of general linguistics*, *3*(1), 1–34.
Culbertson, J. (2018). Artificial language learning. In *Oxford handbook of experimental syntax* (pp. 1–26).
Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, *111*(16), 5842–5847.
Cysouw, M. (2003). *The Paradigmatic Structure of Person Marking* (Oxford University Press, Ed.).
Cysouw, M. (2011). The expression of person and number: A typologist's perspective. *Morphology*, *21*(2), 419–443. doi: 10.1007/s11525-010-9170-5
Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., ... Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*.
Harbour, D. (2008). On homophony and methodology in morphology. *Morphology*, *18*(1), 75–92.
Harbour, D. (2016). *Impossible persons*. MIT Press.
Harley, H., & Ritter, E. (2002). A Feature-Geometric Analysis. *Language*, *78*(3), 482–526.
Kay, P., & Regier, T. (2007). Color naming universals: The case of Berinmo. *Cognition*, *102*(2), 289–298.
Kemp, C., & Regier, T. (2012). Kinship Categories Across Languages Reflect General Communicative Principles. *Science*, *336*(6084).
Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, *4*, 109–128.
Sauerland, U., & Bobaljik, J. D. (2013). Syncretism Distribution Modeling: Accidental Homophony as a Random Event. In *Glow in asia* (pp. 1–27).
Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*.
Zaslavsky, N., Kemp, C., Tishby, N., & Regier, T. (2018). Color naming reflects both perceptual structure and communicative need.
Zwicky, A. M. (1977). Hierarchies of person. In *Papers from the 13th regional meeting of the chicago linguistic society* (pp. 714–733). Chicago.