Social structure and lexical uniformity: a case study of gender differences in the Kata Kolok community

Katie Mudd (katie.mudd@ai.vub.ac.be)
AI lab, Vrije Universiteit, Elsene, Belgium

Hannah Lutzenberger (h.lutzenberger@let.ru.nl)
Centre for Language Studies & Department of Language and Communication, Radboud University, Nijmegen, the Netherlands

Connie de Vos (C.L.G.deVos@tilburguniversity.edu)
Department of Communication and Cognition, Tilburg University, Tilburg, the Netherlands

Bart de Boer (bart@ai.vub.ac.be)
AI lab, Vrije Universiteit, Elsene, Belgium

Abstract

Language emergence is characterized by a high degree of lexical variation. It has been suggested that the speed at which lexical conventionalization occurs depends partially on social structure. In large communities, individuals receive input from many sources, creating a pressure for lexical convergence. In small, insular communities, individuals can remember idiolects and share common ground with interlocuters, allowing these communities to retain a high degree of lexical variation. We look at lexical variation in Kata Kolok, a sign language which emerged six generations ago in a Balinese village, where women tend to have more tightly-knit social networks than men. We test if there are differing degrees of lexical uniformity between women and men by reanalyzing a picture description task in Kata Kolok. We find that women's productions exhibit less lexical uniformity than men's. One possible explanation of this finding is that women's more tightly-knit social networks allow for remembering idiolects, alleviating the pressure for lexical convergence, but social network data from the Kata Kolok community is needed to support this explanation.

Keywords: social structure; input variability; lexical uniformity; language emergence; sign language; Kata Kolok

Introduction

The amount of variation that is observed in recently emerged languages appears to depend in part on the social network structure of the communities in which they are used. In *esoteric communities*, which are small, highly connected, insular communities, and in which contact with outsiders is infrequent, variation appears to be tolerated. In contrast, in *exoteric communities*, which are large and sparsely connected communities, variability appears to be much harder to maintain.

An example of differing amounts of variation in esoteric and exoteric communities comes from a comparison of two sign languages, which have emerged in different types of communities around the same time. Meir, Israel, Sandler, Padden, and Aronoff (2012) observed a high degree of variation in everyday concepts in Al-Sayyid Bedouin Sign Language (ABSL), a sign language which emerged in a small, insular community in the Negev Desert in Israel. For example, ABSL shows a high degree of lexical variation in everyday concepts such as *onion* and *morning*, with different signers using different variants to refer to these concepts. In comparison, Meir et al. (2012) observed that there is far less lexical variation in Israeli Sign Language (ISL), which emerged as a result of interaction in a community consisting of predominantly deaf individuals in Israel, resulting in a much larger community of users than the ABSL community. In comparison to ABSL, ISL seems to be more uniform, and signers seem to have converged more quickly on signs for everyday concepts. Crucially, the amount of variation present in these sign languages may be partially attributed to users adapting to the amount of variability they receive in their input (which is determined in part by social structure), and to their need to understand each other despite such variability.

In Kata Kolok (KK), a sign language which emerged in a relatively small, insular community in Bali, de Vos (2011) finds that there are few conventionalized color terms. Signers used other strategies, such as naming relevant objects (e.g. banana to indicate *yellow*) and pointing to objects in one's surroundings (e.g. pointing to a yellow piece of clothing to refer to *yellow*). Notice that these strategies require a shared context or face-to-face interaction. The lack of conventionalized color terms and substitutive strategies are not unique to KK, and have been attested across other esoteric communities (de Vos, 2011, p. 73). More generally, as seen in the previous example of ABSL, there seems to be a lack of pressure for lexical uniformity in languages emerging in esoteric contexts.

de Vos (2011) puts forth three reasons for why esoteric communities exhibit such a high degree of variation: first, face-to-face interaction facilitates the use of strategies like pointing, and generally relying on a shared environment. Second, because these communities are relatively small and community members frequently interact, individuals can remember each other's idiosyncratic expressions. Last, these types of sign languages are not typically used in deaf education. Standardized education may accelerate lexicalization and lexical uniformity.

Though it is likely that all three of these factors play a role in determining the degree of lexical variation in a community, here we focus on the second of these explanations, namely

that it may be easier to keep track of idiosyncrasies because of frequent contact in a small, insular community. It fits with an existing hypothesis that one of the drivers of linguistic uniformity in the early stages of language formation is the need to overcome high levels of input variability (Lev-Ari & Shao, 2017; Raviv, Meyer, & Lev-Ari, 2019). In esoteric communities, all individuals tend to interact with the same set of other individuals (i.e. the small set of members of the community) and therefore it is likely that each one receives similar linguistic input (Liu, Madhavan, & Sudharshan, 2005). Hence, it may be easier to keep track of idiosyncrasies. In contrast, in exoteric communities, each individual tends to interact with different sets of other individuals, and therefore keeping track of idiosyncrasies is not very useful. In such communities, the low level of shared context between members creates a pressure for convergence (Wray & Grace, 2007; Trudgill, 2011). Put simply, if people need to interact with individuals with whom they share little common ground on a regular basis, it is beneficial to use the same linguistic variant, so no time is spent having to decipher what a variant refers to.

The hypothesized link between social network structure, the ability to remember idiosyncrasies and emergence of linguistic conventions has been investigated in a recent computational model. Thompson, Raviv, and Kirby (2020) used an agent-based model to study the relationship between population size and lexical convergence. In this model, agents had individual-specific representations of the variants used by community members (i.e., idiolects) as well as population-level representations (i.e., the shared language), which updated as a result of interactions with other agents in the population. They focused on memory limitations and shared history, which dictated agents' productions: when communicating with a familiar agent, agents would choose the specific variant used by that agent; but when interacting with a stranger, agents would choose the most prevalent variant in the entire population. Crucially, agents only had a limited memory capacity, such that it was not possible to memorize all agents' productions across the community. In the initial stage, there were no conventions at all, so agents needed to improvise signals for a given meaning. This means that in the bigger populations, agents were exposed to more different signals than they could handle and quickly converged to a uniform language. In contrast, in the smaller population, fewer different signals were coined and therefore agents could memorize them (and their association with who uses them). This reduced the pressure to conventionalize. This leads to the counterintuitive situation that the population that initially has less variants ends up with more variation. This result is comparable to the high level of variation found for everyday referents in esoteric signing communities, such as in ABSL. To summarize, naturalistic data and computational models have shown that the amount of lexical uniformity in a community may be attributed in part to social network structure and to memory limitations, such that initial high variability in exoteric communities eventually leads to lower levels of lexical variation (i.e., more convergence), while relatively lower variability in esoteric communities actually preserves existing lexical variation.

In the current study, we expand on the idea that social network structure can lead to differences in the amount of variability in productions across a linguistic community. Here, we look at specific interaction patterns between different individuals within one community, instead of comparing individuals from two separate communities. We investigate what happens when subgroups of a community who share a language have different interaction patterns. Though not perfectly, this helps us to isolate the effect of the social network from the other proposed explanations, because the groups concerned share the same cultural environment, and for both groups standardization stemming from education cannot play a role.

Extrapolating from previous work, individuals within the same community may have denser or more sparsely connected networks, resulting in exposure to different amounts of input variability. In turn, these differences may affect their choice of lexical variants, such that those exposed to more variability would prefer producing more uniform variants (the most extreme case being where all individuals use one sign to refer to a concept). Supporting this hypothesis, Lev-Ari and Shao (2017) showed that differences in the social network (size, heterogeneity and density) of native English speakers in the US influenced their language skills: individuals with more heterogeneous networks (who were consequently exposed to more input variability) demonstrated better performance on lexical access and prediction tasks. Thus, there is preliminary evidence that interaction patterns within a community can affect linguistic behaviors. However, it is currently unclear how such differences in interaction patterns and social network structure affect variation within a community. To explore this question, we turn to a case study of lexical variation in KK. As a young language (compared to any spoken language), KK provides a unique window into the process of language evolution, and since subgroups of this community have different interaction patterns, we have the opportunity to see how such differences may in part affect the degree of lexical uniformity across groups.

The current study

KK is a sign language which emerged at least six generations ago in a Balinese village due to a high incidence of hereditary deafness (Winata et al., 1995). More than half of the village knows KK, with the majority of signers being hearing and fluent in KK to varying degrees (Marsaja, 2008). In this community, Marsaja (2008) found that more men are fluent signers than women, which he attributed to the strong patriarchal system in the community: "opportunities for females to go to school, leave the home, be involved in public activities, and interact freely with outsiders are considerably less than for males" (Marsaja, 2008, p. 115). Marsaja (2008, p.115-116) adds that the majority of women are housewives and

are dependent on their husbands or parents to attend certain events.

As such, these differences in social responsibilities affect interaction in a way that parallels characteristics of esoteric and exoteric communities. Specifically, the interaction patterns of men resemble exoteric communities with frequent interaction with a larger pool of community members, and the interaction patterns of women resemble esoteric communities with a smaller and more dense network structure. In this setting, the hypothesis that social network structure influences variability can investigated, while controlling for environmental and cultural factors. It predicts that women should have higher variability in their language use than men. However, because we are lacking extensive data on the social networks of women and men in this study and in the KK community at large, it is not possible to draw any conclusions about the nature of the relationship between social network and the amount of lexical variation. Rather, this study should be seen as a preliminary attempt to investigate this relationship. A possible confound is that women tend to lead linguistic innovation (Labov, 1990). Thus, if a higher degree of variation within their group is found it is possible that this is a result of women's propensity to innovate. To control for this, we would need to study women in different network types; this should be the subject of future work.

To test our prediction about social network structure and the degree of lexical variation, we conduct a reanalysis of a picture description task in KK (Mudd et al., 2020). A major shortcoming of the previous analysis was that only the first lexical variant produced in response to the stimulus was analyzed. Because participants typically produced several variants in response to stimuli, in the current study we analyze the majority of relevant variants produced by participants, and compare their productions for each stimulus using a distance measure called Jaccard distance (see Analysis). Specifically, we compare the amount of lexical variation between the women and men in order to test the effect of interaction patterns on the amount of linguistic variation in one community.

Method

Participants

46 participants from the KK community took part in a picture description task (see Table 1). Our sampling focused on maximizing the number of deaf participants from different village clans (geographical areas) and age groups. Additionally, we recorded the gender of the participant and if they have a deaf family member. Fluency was assessed by Lutzenberger, a fluent signer of KK. Although it is typical to just include deaf signers in studies of sign languages, we included both deaf and hearing participants in order to achieve a representative sample of the population. Similar to other shared sign languages, hearing signers build the core of the KK signing community. Marsaja's (2008) claims about the difference in the social networks of women and men in the KK community apply to both deaf and hearing individuals, hence we did not have a specific hypothesis pertaining to differences in the degree of lexical uniformity between deaf and hearing groups.

Materials and Procedure

The picture description task consisted of 36 culturally relevant stimuli from various semantic domains: animals, food, religion, colors and a miscellaneous category. Based on Lutzenberger and de Vos's knowledge of KK (both are fluent KK signers), as well as ongoing work documenting KK (Lutzenberger, 2021), we expected differences in the number of variants that each stimulus would elicit. For some stimuli, we expected few to no variants (e.g. *dog*), and for others we expected a larger number of variants (e.g. *dragonfruit*). A deaf research assistant from the KK community led the picture description task, asking participants what each stimulus is in KK. The participants' responses were videotaped.

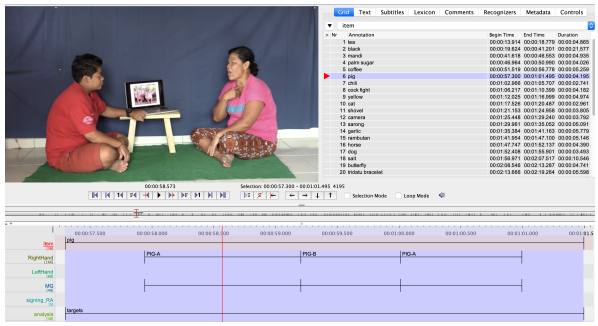

Figure 1: Example of annotation using ELAN: a participant responds to the stimulus *pig*

Analysis

As shown in Figure 1, the data from the picture description task were annotated using ELAN (Crasborn & Sloetjes, 2008). Mudd annotated the data using the KK dataset in Global Signbank as a reference, accompanied by a list of glosses that could be expected in response to each stimulus, which was provided by Lutzenberger (a fluent KK signer). Lexical variants were identified based on their underlying iconic motivation and mapping, such that signs were grouped together based on their iconic motivation and where on the body and hands a feature is mapped. For example, as shown in Figure 2, each of the three variants produced for the stimulus *pig* refer to different semantic components of a pig: how a pig is killed (PIG-1), how a pig eats (PIG-2) and the ears of a pig (PIG-3). Another example is from participant responses to the stimulus *butterfly*, where variants involving the same semantic content (the wings of the butterfly) were identified as being distinct variants for the analysis as they mapped onto different body parts (the hands or arms) (see Fig. 4 in Mudd et al., 2020).

All 46 participants responded to all 36 picture stimuli (1,656 trials in total). We removed trials from the analysis in the following cases: if a participant did not know the sign for

Table 1: Participant information by gender

		age			deaf/hearing		deaf family		signing skills				
		young	medium	old	deaf	hearing	yes	no	1	2	3	4	5
women	21	5	11	5	11	10	16	5	0	1	4	2	14
men	25	11	7	7	9	16	16	9	1	3	4	3	14
total	46	16	18	12	20	26	32	14	1	4	8	5	28

a stimulus, if a participant did not understand the stimulus and if a participant did not produce a valid target. All color stimuli (*black, white, red, yellow*) were removed prior to the analysis as they were confusing to participants. We considered signs produced to be a valid target if they are a relevant description of the stimulus, produced by three or more participants, with the goal of excluding gestures and productions improvised on the spot[1]. Finally, prior to the lexical distance analysis we removed four stimuli (*blayag*- steamed rice wrapped in a leaf, *cat, cow, dog*) which only elicited one target in response to the stimulus; because all participants produced the same target, there is no variation, i.e. maximal uniformity between participants for these stimuli. After removing the trials not suitable for analysis, we analyzed the remaining 1,166 trials. For each trial, all relevant parts of the response were annotated[2]. Hence, sequences of signs (see Figure 2 for two examples of participants producing sequences), compounds and collocations were included, with each individual sign as separate variants. Because it is not always obvious if a pair of signs are a compound, in this analysis, we chose to consider each sign as a separate. Determining the boundary between sequences, collocations and compounds should be the topic of a future work (see Tkachman & Meir, 2018 for more on this topic).

To see how much lexical variation exists in a group of participants, we calculate the Jaccard distance between responses to each stimulus between pairs of participants. As shown in Figure 2, the Jaccard similarity, also used in Horton's (2018) analysis of homesigners in Guatemala, is the number of targets in the intersection divided by the number of targets in the union. Jaccard distance equals 1 – Jaccard similarity. A Jaccard distance of 0 would mean that both participants produced identical responses to a stimulus, while a Jaccard distance of 1 would mean that there was no overlap between the responses produced by participants (i.e., maximally distant). All targets in responses are of identical weight in this analysis, meaning that no two targets are considered more similar or different, or weighted based on factors such as the frequency of production in the sample (see Lutzenberger, de Vos, Crasborn, & Fikkert, in press for an analysis of weighted form variation). See Figure 2 for an example of how Jaccard distance is calculated for the stimulus *pig* between participants B and D.

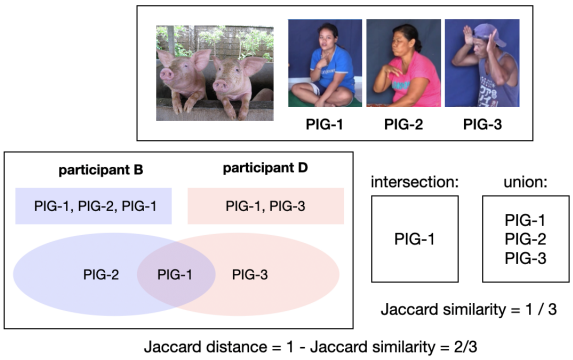

Figure 2: Calculating the Jaccard distance between participant B and D's response to the stimulus *pig*. Both their productions consisted of the target PIG-1 (intersection = 1) and all the targets produced by the pair are PIG-1, PIG-2, and PIG-3 (union = 3). The Jaccard distance between these productions is 2/3, yielded by 1 - Jaccard similarity, which is the intersection divided by the union (= 1 / 3).

We use a permutation test to determine if the ratio between the mean Jaccard distance between the 25 women and the 21 men is significantly different from randomly generated subsamples of the population (i.e., permutations). To elaborate, with the total sample of 46 participants, we sample without replacement a group of 25 and a group of 21 (irrespective of gender). For each group, we calculate the mean of the Jaccard distances. We then calculate the ratio between the mean Jaccard distance of the group of 25 and the group of 21. This process is repeated 1000 times. With these subsamples, the original statistic (ratio of the Jaccard distance between 25 women and 21 men) is compared to that of the generated subsamples. The null hypothesis is that the ratio between women and men is 1, and would be rejected if the original statistic falls outside of the 2.5 or 97.5 percentiles of the generated subsamples.

Results

The number of variants produced across all participants in response to stimuli ranged from 1 variant (*blayag, cat, dog, cow*) to 8 variants (*offering*). Focusing on the community as a

[1]Though, because of iconic affordances, we cannot rule out that several signers may have improvised the same production.

[2]Some longer responses were not fully annotated, for example if a participant was telling a lengthy anecdote

whole, we found that the majority of stimuli elicited multiple variants. In most trials, participants produced more than one target in response to a stimulus (M = 1.41, SD = 0.64), and women tended to produce more targets (M = 1.45, SD = 0.66) than men (M = 1.38, SD = 0.62). Across all participants, we found that the average Jaccard distance was 0.5 (SD = 0.09, min = 0.21, max = 0.79), indicating that on average half of the participants' lexical choices were shared across the community. Table 2 shows a subset of the distance matrix with 5 participants.

Table 2: Subset of the matrix used for analysis showing the average Jaccard distance over stimuli between pairs of participants. For example, participants B and D have a Jaccard distance of 0.68, which is fairly high compared to the Jaccard distance between other pairs of participants, meaning that their responses differ more than most of the responses between other pairs of participants. In contrast, participants A and E have a Jaccard distance of 0.28, meaning that their responses to stimuli were fairly similar.

	A	B	C	D	E
A	0	0.45	0.5	0.57	0.28
B	0.45	0	0.61	0.68	0.37
C	0.5	0.61	0	0.42	0.56
D	0.57	0.68	0.42	0	0.64
E	0.28	0.37	0.56	0.64	0

Next, we focus on the distances between women and between men in our study, which may provide an insight into if the amount of variation in their lexicons may be the result of the interaction patterns of men (more typical of exoteric communities) and the interaction patterns of women (more typical of esoteric communities). However, as previously stated, quantitative social network data from the KK community would be needed to back this claim. Figure 3 shows the average Jaccard distance per stimulus for women and men separately. For some stimuli, such as *coffee, slippers* and *shovel*, there is more variation between women than between men. For other stimuli, such as *sarong, dragonfruit* and *mango*, there is more variation between men than between women. Nevertheless, the average Jaccard distance between women is higher (M = 0.52, SD = 0.09) than between men (M = 0.47, SD = 0.08). Thus, on average there is more variation in responses between women than between men. The ratio between the mean Jaccard distances of women and men is 0.516:0.470, thus 1.097. We used a permutation test to determine if this ratio is significantly different from the ratios of the 1000 randomly generated subsamples. We find that the actual ratio lies outside of the 95% range of the ratios of the generated subsamples (2.5 percentile = 0.912 and 97.5 percentile = 1.096), and therefore we accept the alternative hypothesis that the ratio of the mean Jaccard distance of women and men is significantly different from 1.

To give a clear example of the difference in the amount of variation between women and men in this study, consider the productions for the stimulus *shovel*. In this case, there were two distinct responses to this stimulus: SHOVEL-1 (depicting holding the shovel handle) or SHOVEL-2 (depicting the blade of the shovel). The majority of men produced SHOVEL-1 (21 out of 25), while women did not show a strong preference for either variant (12 produced SHOVEL-1 and 8 produced SHOVEL-2). Therefore, in response to the stimulus *shovel*, women produced more variation (mean Jaccard distance of 0.51) than men (mean Jaccard distance of 0.28).

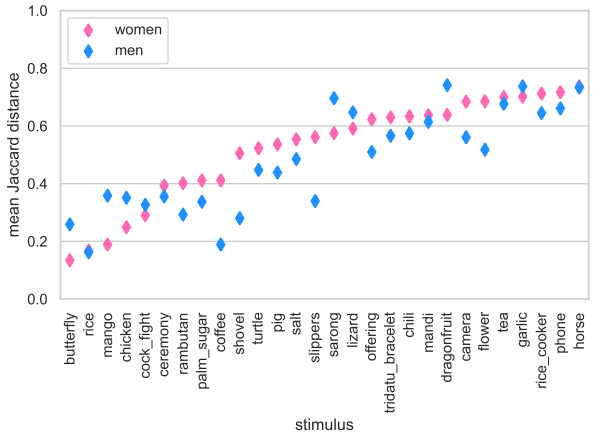

Figure 3: Mean Jaccard distance per stimulus for women (pink diamonds) and men (blue), with the average Jaccard distance across all stimuli being higher between women (M = 0.52, SD = 0.09) than between men (M = 0.47, SD = 0.08).

Discussion

Previous studies compared the amount of lexical variation present in two different communities as a function of their social structure, and showed that communities with wider social networks have a more uniform lexicon (e.g. Meir et al., 2012). Our study is the first to consider lexical variation within a community in which different subgroups (i.e., women and men) have different types of interaction patterns and thus different social network structures. Our work showed that subgroups of women and men in the same community produced different degrees of lexical uniformity: women produced more varied labels compared to men, who were more uniform in their productions.

One explanation for this trend, as put forth by Lev-Ari and Shao (2017) and Raviv et al. (2019), is that exposure to a high degree of input variability leads to more linguistic uniformity. In the KK community, as men have community-wide ties, they likely receive more varied input, which could have in turn led them to produce more uniform lexical variants. In contrast, women have tighter social networks and mostly interact with known individuals, leading them to receive less

varied input, which could have in turn led them to produce more varied signs. In line with de Vos' (2011) hypotheses about what drives lexical variation at the community level, the stable, shared context of women may have reduced the pressure for lexical uniformity and lexicalization at all.

It should be noted that it is possible that a higher degree of variation is tolerated in sign languages because of the iconic affordances of the manual-visual modality (Meir et al., 2012). Iconicity, referring to the degree by which a meaning can be retrieved from a form, depends on the unique linguistic, perceptual, sensory-motor and cultural experiences of an individual (Occhino, Anible, Wilkinson, & Morford, 2017). In other words, what is iconic for one individual might not be iconic for another individual even within a linguistic community. And given one's experience with a concept, one variant may be preferred over another.

For example, the variant PIG-1 (see Figure 2) refers to how a pig is killed. Knowledge of how pigs are killed is specific to communities that practice animal farming and use this technique to kill pigs, and even more specific to people who carry this out. Other variants produced in response to the stimulus *pig* are PIG-2, referring to how a pig eats, and PIG-3, referring to the ears of a pig. The majority of variants produced in response to this stimulus were PIG-1 (55%) and PIG-2 (32%). In the KK community, it is typically men who kill pigs, while feeding is typically but not exclusively done by women (Lutzenberger et al., in press). In looking at how women and men responded to the stimulus *pig*, it appears that individual experience is shaping their choice of variants[3]; men produce more PIG-1 (66% of their responses) than women (45% of their responses) and women produce more PIG-2 (39% of their responses) than men (22% of their responses). Further, men were more likely to produce a single target in response to *pig* (82% of all responses) while women were equally likely to produce one or two variants in response to *pig*. It should also be noted that women's responses to this stimulus were longer on average (M = 1.45, SD = 0.66) than men's (M = 1.38, SD = 0.62), but given the amount of variance, this may be coincidence.

In sum, the different experiences of individuals with a concept likely shapes lexical variant choice with regards to iconicity. Even when taking the prevalence of iconicity into account, social network structure is still expected to affect linguistic variation. But the precise contribution of shared context (permitting iconic form-meaning mappings to be understood) and social network structure on the degree of variation within a community are yet to be determined, and should be the topic of further study.

A second potential explanation of our finding, that women exhibit more lexical variation than men, is that women simply innovate more as discussed by Labov (1990). It is possible that the higher degree of variation found in the group of women is a product of their tendency to innovate and lead language change, as opposed to a result of their social network structure. With the current design, it is not possible to rule this out. However, as previously mentioned, the extent to which innovation vs. social network configuration drives the degree of lexical variation could be studied by considering women with different social network types.

One clear limitation of the current study is that it did not directly test the role of input variability and social network structure, but instead relied on the observation that women and men in the KK community exhibit different interaction patterns (Marsaja, 2008). Future work should include richer and up-to-date information about the exact social network structure of participants in order to determine the precise nature of their social networks. Following Lev-Ari and Shao (2017), we recommend collecting data on the size, heterogeneity and density of the social network of participants. Moreover, in contrast to large-scale studies comparing properties esoteric and exoteric communities (e.g. Lupyan & Dale, 2016), the groups which we have drawn parallels with esoteric and exoteric communities here are not independent: women and men in the KK community frequently communicate. It is unclear how the interaction between these groups would affect the degree of variation within them.

In addition, future work could focus on other sociolinguistic properties of participants besides their gender, such as their age (for examples, see Table 1). Older speakers may exhibit more linguistic variation than younger speakers. This hypothesis is in line with findings from iterated learning experiments, which show that unpredictable variation is typically reduced as a language ages and is transmitted over more generations (Smith & Wonnacott, 2010). Given the characteristics of our sample, age may be a confounding factor that, however, may represent the community adequately: the old age group includes more men than women. Undoubtedly, multiple sociolinguistic factors can interact when it comes to lexical preferences and the amount of variation in the lexicon, and as such should be the topic of future work. In a community of this size, it is a challenge to disentangle these factors.

Some features which may be of further interest also present possible confounds in this study because the groups are not entirely balanced across all sociolinguistic features[4]. As can be seen in Table 1, there are more hearing men than women, and there are more young and old men in the sample than women (who are mostly in the medium age category). Because of the complexity of the analysis, involving non-independent distance measures between participants, it is unclear how to tease apart these additional sociolinguistic features which could have an influence on the amount of variation observed within the groups of women and men.

[3]For another example of differences between the experiences of women and men shaping iconic affordances and thus the choice of preferred variants, see an analysis of shared homesign systems in Guatemala by Horton (2018, p. 200)

[4]In a small community balancing participants across all sociolinguistic fronts is virtually impossible. For example, as mentioned in Mudd et al. (2020) with regards to sampling by geographic unit in the KK community, no deaf participants in clan 9 were sampled because there are none.