Formal variation in the Kata Kolok lexicon

HANNAH LUTZENBERGER [ID]

CONNIE DE VOS [ID]

ONNO CRASBORN [ID]

PAULA FIKKERT [ID]

Author affiliations can be found in the back matter of this article

Abstract

Sign language lexicons incorporate phonological specifications. Evidence from emerging sign languages suggests that phonological structure emerges gradually in a new language. In this study, we investigate variation in the form of signs across 20 deaf adult signers of Kata Kolok, a sign language that emerged spontaneously in a Balinese village community. Combining methods previously used for sign comparisons, we introduce a new numeric measure of variation. Our nuanced yet comprehensive approach to form variation integrates three levels (iconic motivation, surface realisation, feature differences) and allows for refinement through weighting the variation score by token and signer frequency. We demonstrate that variation in the form of signs appears in different degrees at different levels. Token frequency in a given dataset greatly affects how much variation can surface, suggesting caution in interpreting previous findings. Different sign variants have different scopes of use among the signing population, with some more widely used than others. Both frequency weightings (token and signer) identify dominant sign variants, i.e., sign forms that are produced frequently or by many signers. We argue that variation does not equal the absence of conventionalisation. Indeed, especially in micro-community sign languages, variation may be key to understanding patterns of language emergence.

1 Introduction

Despite being produced and perceived in distinct modalities, signed and spoken languages parallel on all levels of linguistic structure.

Much like spoken languages, the lexicons of sign languages are shaped by phonological specifications. In 1960, Stokoe (1960) showed that signs in American Sign Language (ASL) are compositional, consisting of the parameters of handshape, movement and location, that differentiate between signs (phonemic contrasts). It was later recognised that signs are composed of finer feature distinctions which can be analysed by a feature geometry model, just like spoken phonemes (Brentari 1998; van der Kooij 2002). Feature sets are most easily determined through lexical contrast. For example, the sign pairs TO-LIVE-IN[1] and HOLIDAY, and ALSO and HOLIDAY represent minimal pairs in Sign Language of the Netherlands (NGT; *Figure 1*). The first pair (TO-LIVE-IN and HOLIDAY) differs in handshape features and the second (HOLIDAY and ALSO) in location features (van der Kooij 2002: 21 f.). Specifically, TO-LIVE-IN uses a handshape with thumb and index touching and HOLIDAY is produced with all fingers extended; HOLIDAY is articulated at the ipsilateral side of the mouth and ALSO at the contralateral side of the chest. Besides these differences, all other features are shared.

The featural organisation of signs allows for comparing pairs of signs in terms of their formational similarity. Minimal pairs can serve as evidence for phonologically distinctive feature values of a language. Sign comparisons have also been widely operationalised to determine form similarities within and across sign language lexicons, in order to assess genealogical relationships (Woodward 1991; 1993; McKee & Kennedy 2000; Currie et al. 2002; Xu 2006; Sasaki 2007; Al-Fityani & Padden 2008), and to correlate formational variation with social factors (Bayley et al. 2000; 2002; Lucas et al. 2002; Schembri et al. 2009; Fenlon et al. 2013; Siu 2016).

While sign phonology has been extensively documented and analysed, most of this research is conducted into sign languages used in large urban deaf communities, here referred to as *macro-community sign languages*[2] (Schembri 2010; Schembri et al. 2018). Among sign languages used within small and tight-knit rural communities, or *micro-community sign languages* (Schembri 2010; Schembri et al. 2018), the status of phonology is unclear. Initial studies of sign language emergence in these settings show that the form of signs varies greatly across signers and iconicity is the strongest force driving sign formation (Sandler et al. 2011). For example, in Al-Sayyid Bedouin Sign Language (ABSL), a young micro-community sign language of a Bedouin village in Israel, signs to refer to a dog used across signers are motivated by the dog's barking, but signers differ greatly in how the barking is encoded (Sandler et al. 2011: 520). The authors also note that minimal pairs (as illustrated in *Figure 1* for NGT) are unattested in ABSL, leading them to conclude that phonological structure has not yet emerged in this language. This raises the fundamental questions as to if, how, and when phonological structure emerges in young sign languages. Crucially, micro-community sign languages are distinct from macro-community sign

| TO-LIVE-IN | HOLIDAY | ALSO |

Figure 1 Minimal pairs of signs in NGT. Images retrieved from Global Signbank (Crasborn et al. 2019).

---

1   Following the convention, translations equivalents of signs are represented as glosses in small caps: GLOSS.

2   The discussion about terminology especially in the field of sign language emergence is extensive and ongoing (see for example de Vos & Pfau 2015; or Hou 2016). As laid out in de Vos & Pfau (2015), many determining factors are confounded and depending on the context of discussion, different and multiple labels may be used to classify a language without being mutually exclusive. For the purpose of this study, we adopt the terminology first suggested by Schembri (2010) within the framework of sociolinguistic typology (Schembri et al. 2018). This terminology is chosen without taking position on the relevant theoretical background and for the sake of simplicity and clarity. Note that these terms are not always used by the authors themselves in previous literature (e.g. Sandler at al. (2011) and Meir et al. (2010) refer to ABSL as village sign language and as emerging sign language and Kisch (2008) describes the community as shared signing community).

languages with respect to ecological niche, emergence setting and age, leaving multiple viable hypotheses regarding the causes and mechanisms underlying the attested phenomena (de Vos & Pfau 2015). Moreover, the methodological approach to studying variation has focused on the level of an iconic motivation in micro-community sign languages (e.g. Richie et al. 2014; Hou 2016; 2018; Horton 2018; Horton & Riggle 2019; Neveu 2019; Reed 2019; Mudd, Lutzenberger, et al. 2020) and on the sub-lexical level in macro-community sign languages (e.g. Bayley et al. 2000; 2002; Lucas et al. 2002; Fenlon et al. 2013; Siu 2016).

This study charts the formational variation in Kata Kolok (KK), another micro-community sign language that emerged six generations ago in a village community in rural Bali, Indonesia. We examine the extent of variation in the form of signs by analysing responses across 20 deaf KK signers to a picture elicitation task. We combine different techniques that have been used for sign comparisons and introduce a newly developed measure of variation. This method yields a numeric score and incorporates variation across i) iconic motivations, ii) surface realisations, iii) feature differences. A gradient measure is able to unify variation stemming from multiple sources, each of which may vary to differing degrees within the same sign. We argue that especially in studies discussing sign language emergence, both token distributions within the dataset and distribution of variants across participants are fundamental to capturing variation adequately.

2 Background
2.1 Iconicity

Iconicity, defined as the structured mapping between meaning and form, acts as an organisational force across all natural languages (Perniss et al. 2010; Dingemanse et al. 2015). Taub (2001) describes iconicity in sign languages as a process of analogue building between semantic and phonological representations. Certain aspects from or related to a sensory image are first selected, then schematised, and finally linguistically encoded by mapping them onto the articulators, i.e., the signer's hands and body. More specifically, overlap in experiences and sociocultural and linguistic background of language users impacts the construal and the interpretation of iconicity (Wilcox 2004; Occhino 2017; Occhino et al. 2017).

One concept may result in many different iconic mappings. For example, the signs for BIRD in ASL and Turkish Sign Language (TİD) entail iconic mappings based on different sensory images and consequently, different schematisation and encoding (*Figure 2*). The ASL sign selects the sensory image 'beak' and maps the signer's hand to the bird's beak, articulating it at the mouth to reflect shared structural and functional traits of a bird's beak and a human's mouth. The TİD sign selects and schematises the wings of the bird, mapping the wings onto the signer's arms, with a flapping motion to represent flying.
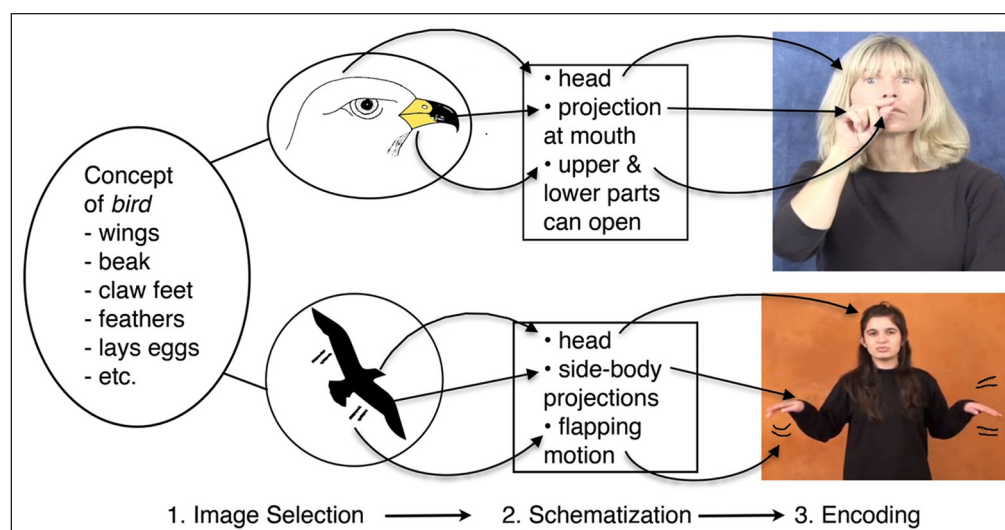

Figure 2 Analogue-building of the sign BIRD in ASL (top) and TİD (bottom) (from Emmorey 2014: 2) [reproduced with permission].

Examining phonology and semantics shows that iconic mappings are compositional. Indeed, for the ASL sign BIRD, iconic mappings can be analysed at the phonological level by examining the

features (i) handshape, (ii) location, and (iii) handshape change. Thus, (i) a handshape with an aperture relation between index finger and thumb relates to the shape of the beak, (ii) the sign is produced at the mouth, reflecting shared properties between the beak and the mouth, (iii) the handshape opens and closes, reflecting the movement of a bird's beak. Both examples provided in *Figure 2* show iconic mappings on multiple formational levels, yielding signs with different iconic motivations. Recent research has further suggested that iconic mappings extend beyond single lexemes; certain mappings are recurrent across the lexicon (van der Kooij & Zwitserlood submitted). Naturally, iconic mappings are situated within the cultural context in which a language is used, and/or emerges; thus, cultural knowledge shapes what features are available for selection and mapping.

To sum up, iconicity is a powerful force organising sign language lexicons. Semantic information related to a concept may be linguistically encoded through analogue-building. Properties of or related to the concept provide the raw materials which then become schematised and mapped onto the signer's articulators, creating iconic signs. This process can lead to different forms across languages and within a sign language.

2.2 Comparisons of signs

Comparisons of signs have long been a staple in both the field of lexicostatistics, where form similarities across different sign languages are used to establish phylogenetic relations (Woodward 1991; 1993; McKee & Kennedy 2000; Currie et al. 2002; Xu 2006; Sasaki 2007; Al-Fityani & Padden 2008) and sociolinguistics, where comparisons of signs within a single language target sociolinguistic variables (Bayley et al. 2000; 2002; Lucas et al. 2002; Schembri et al. 2009; Fenlon et al. 2013; Siu 2016). Studies on sign language emergence, and particularly emerging phonology, have also relied on comparisons of signs produced by signers with varying sociolinguistic profiles from the same linguistic community (Israel 2009; Israel & Sandler 2009; Sandler et al. 2011; Morgan 2015). In the following sections, we provide details about the methodological approaches and findings in each of the fields where sign comparisons are frequently used.

2.2.1 Cross-linguistic comparisons

Within lexicostatistics, pairs of signs for the same concept from different macro-community sign languages are compared on different formational parameters (handshape, location, movement, handedness, other), often classifying signs as *identical, similar*, or *different* (Woodward 1991; 1993; McKee & Kennedy 2000; Currie et al. 2002; Xu 2006; Sasaki 2007; Al-Fityani & Padden 2008). More recent cross-linguistic comparisons are based on the feature level, using a match/non-match criterium (Yu et al. 2018; Börstell, Crasborn & Whynot 2020), or an adapted Levenshtein distance (also known as edit distance between two forms) (Parks 2011; Omardeen 2018). Such comparisons result in a numeric score rather than a categorical outcome, and thus are more appropriate to measure differences across sign language lexicons, allowing to include subtle differences in sign formation.

However, while form similarities in spoken languages have long been taken as evidence for linguistic relatedness, cognates in sign languages may also be due to overlap in iconic mappings stemming from similarities in human experiences that cross-cut cultures. Al-Fityani and Padden (2010) and Börstell and colleagues (2020) show that unrelated sign languages display form similarities. In particular across the basic vocabulary, sign languages show high degrees of formational similarity, possibly due to similar iconic mappings. For example, the sign languages of America, New Zealand and Australia share 19% of the signs from a Swadesh list, including signs like GOOD, BIRD, CAT, CHILD, NARROW, RED, and SUN (McKee & Kennedy 2000). However, integrating iconic motivations into sign comparisons has proven to be a challenge, especially for assessing relatedness, and thus, has not often been attempted. The implementation of a comparison of the iconic motivation before the parametric comparison as suggested by Xu (2006) still overestimates phylogenetic relations (Su & Tai 2009). It is possible that, as suggested by Ebling et al. (2015), a more fine-grained analysis of image-producing techniques and underlying motivation is needed. However, no successful method has been made available at this point.

Summing up, cross-linguistic comparisons quantify form similarities based on parameter or feature overlap (Woodward 1991; 1993; McKee & Kennedy 2000; Currie et al. 2002; Xu 2006; Sasaki 2007; Al-Fityani & Padden 2008; Parks 2011; Omardeen 2018; Börstell et al. 2020). Cross-linguistic similarities may be due to shared iconic motivations or historic relatedness, with iconicity providing a complicating factor in phylogenetic analyses. The few studies that take iconicity into account are limited in their capacity to scale-up and lack a numeric outcome (Xu 2006; Su & Tai 2009; Ebling et al. 2015).

2.2.2 Sociolinguistics

Comparisons of signs within the same language are frequently conducted to study sociolinguistic variation. Early work on signing varieties of Mexico and Costa Rica compares signs from word lists and dictionaries to determine dialectal variation (Bickford 1991; Woodward 1991). More recent studies are more specific about how linguistic and sociolinguistic factors affect sub-lexical variation, in particular age, region, and gender. Using large-scale datasets with signers of diverse demographics, these studies analyse the parametric deviation from a citation form using multivariate analyses (Bayley et al. 2000; 2002; Lucas et al. 2002; Schembri et al. 2009; Fenlon et al. 2013; Siu 2016). For example, Bayley and colleagues (2000) find an effect of age by region in the use of the three variants of the ASL sign DEAF (*Figure 3*) alongside a tendency for the third pictured variant (continuous contact) to occur in compounds. In studies examining variation within a language, sign comparisons are either concerned with a class of signs sharing a specific parameter (e.g. the extended index finger handshape in Bayley et al. 2002; Schembri et al. 2009; or the location parameter in Fenlon et al. 2013; Siu 2016; Lucas et al. 2002) or a class of signs that share the same meaning (e.g. Bayley et al. 2000). Iconicity, however, is not commonly considered, and does not appear as a major factor in the comparisons.
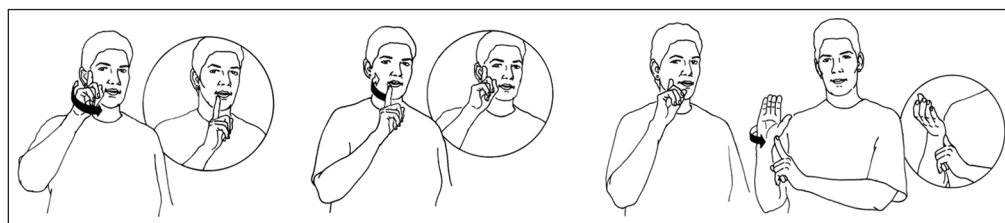

Figure 3 Three sign variants for DEAF in ASL (from left to right): movement ear to chin; movement chin to ear; continuous contact with cheek (in the compound DEAF-CULTURE) (from Bayley et al. 2000: 85) [reproduced with permission].

To conclude, comparative work on signs within a single macro-community sign language does not primarily quantify form similarities or engage with iconicity. Sociolinguistic studies analyse large samples to identify which sociolinguistic variables influence formational variation in terms of deviance from a citation form (Bayley et al. 2000; 2002; Lucas et al. 2002; Schembri et al. 2009; Fenlon et al. 2013; Siu 2016). Nevertheless, the sociolinguistic determinants of the choice of specific variants in interaction remain to be investigated.

2.2.3 Emerging phonology

To examine phonological structure in young sign languages, researchers have employed picture-elicitation tasks, comparing how the same concept is encoded across different signers. The variants produced are then compared across signers to establish the degree of lexical convergence. To date, two studies have examined emerging phonological structure; Sandler and colleagues (Israel 2009; Israel & Sandler 2009; Sandler et al. 2011) examine ABSL, a young sign language of a rural community in Israel, and Morgan (2015; 2017) studies Kenyan Sign Language (KSL), a young sign language that arose in a deaf school in urban Kenya. These studies have suggested that in young sign languages, signers first converge on the same iconic image before aligning the exact phonological form (Sandler et al. 2011; Morgan 2015). As we are interested specifically in sub-lexical variation, studies investigating lexical variation in other small signing communities are not discussed here in detail. Nevertheless, these studies overlap with the present one in some methodological aspects, in particular the role of iconic motivations (Richie et al. 2014; Hou 2016; Horton & Riggle 2019; Neveu 2019; Reed 2019; Reed & Rumsey 2019).

In a cross-linguistic study, Sandler and colleagues compare the systematicity across ten signers of ABSL, Israeli Sign Language, and ASL in the form of signs on a picture-elicitation task (Israel

2009; Israel & Sandler 2009; Sandler et al. 2011). Monomorphemic responses are analysed (these constitute 43% of the data), with the most frequent iconic motivation considered. Comparing responses within the same sign language and then across languages, they find that ABSL, the young micro-community sign language, has the lowest degree of convergence, and ASL, the older macro-community sign language used across the US, the highest. Variation is measured on the phonological feature level, using *mode*, count of the most frequent feature in a set, and *number of variants*, number of features occurring in a set of sign responses with shared iconic motivation. Sandler et al. (2011) conclude that ABSL signers strive for a holistic iconic motivation rather than a compositional sign. However, it is difficult to know how far this outcome is affected by their decision to focus on monomorphemic responses, and to exclude competing variants with different iconic motivations.

Morgan (2015) also uses a picture-based elicitation task to examine the emerging lexicon of KSL. Within the KSL lexicon, the degree of conventionalisation was found to vary from sign to sign, leading Morgan (2015) to hypothesise that signers first converge on an iconic motivation (or *conceptual target*) before aligning their phonology. Across the 20 deaf KSL signers sampled, items like *salt* elicit uniform lexical responses while others such as *island* elicit highly descriptive and probably idiosyncratic responses, similar to ABSL. Morgan posits that during the process of conventionalisation, signs may either converge on a single iconic motivation (e.g. how the fruit is eaten in GUAVA) or stabilise as a compound (e.g. RED^TINY for beans) before aligning in phonological form (Morgan 2015).

Summing up, sub-lexical variation has often been connected to the age of a sign language and its conventionalisation. As observed in KSL and ABSL, signers appear to conventionalise iconic mappings despite large amounts of feature-level variation. The fact that both of these sign languages show variation on the level of the iconic motivation and the sub-lexical level indicates that variation does not materialise on a single level.

2.3 Kata Kolok
2.3.1 Demographic sketch

The sign language Kata Kolok (KK) emerged spontaneously in a village community in North-Bali, Indonesia, due to sudden and sustained incidences of hereditary deafness (de Vos 2012; Marsaja 2008). Deafness was propagated due to geographical isolation and consanguineous marriage patterns within the labour-intensive community (Friedman et al. 1995). KK has been passed on throughout at least six generations of deaf signers (Friedman et al. 1995; de Vos 2012). At present, 33 deaf signers from generation three through six reside in the village permanently, their ages ranging from three years (generation six) to ~80's (generation three) (Lutzenberger in press).[3] Communal living in family compounds with shared religious, social and cultural practices has led to a high proportion of hearing villagers with signing skills (Marsaja 2008).

The tight-knit community of roughly ~3,000 people is socially and geographically structured into ten clans for which membership is determined by birth (Lutzenberger in press). Deaf people have been born into all ten village clans (Marsaja 2008). Following a patrilineal tradition, women transition to their husband's clan through marriage. Within clans, intergenerational households are the norm. This results in tight family bonds where younger generations care for older generations, and childcare becomes a shared task, involving older generations.

The professional landscape of the villagers has long centred around subsistence farming, raising livestock, day labour, or running small local businesses. Recent advances in technology and mobility continue to affect the community's demographics. There is increasing employment within government-supported jobs such as constructing infrastructure or tourism. As a result, villagers may seek job opportunities outside the village and may even relocate to more densely populated areas. This also creates new opportunities for younger deaf villagers who increasingly attend school, or even grow up, in more urban parts of Bali where a variety of Indonesian Sign Language (BISINDO) is used (Lutzenberger in press).

---

3   The literature on sign language emergence lack consensus on how language age is counted and reported, e.g. biological generations, cohorts of signers, etc. and explicit discussions or explanations for the delineation of a generation (Kisch 2012). In this study, we follow the timeframe laid out in de Vos (2012) and provide rough age categories for different generations. Crucially, generation one in this account constitutes five deaf siblings who had an older deaf uncle.

2.3.2 Sketch of the phonology and the lexicon

KK is a sign language isolate, having developed without influence from other sign languages (Marsaja 2008; Perniss & Zeshan 2008; de Vos 2012). Moreover, the surrounding spoken languages Bahasa Indonesia and Balinese do not seem to have strongly influenced the structure of KK (Marsaja 2008; Perniss & Zeshan 2008; de Vos 2012). Evidence for this is, for example, the virtual absence of mouthing, the conventional pairing of manual signs and imitating a spoken word which is a prominent feature in many macro-community sign languages (Crasborn et al. 2008; Bank 2015). In KK, mouthing has been observed only with limited vocabulary and by specific deaf signers; e.g., the word *kopi* alongside the sign COFFEE or *apa* accompanying the sign WHAT. Increasing contact with BISINDO especially among younger signers (Moriarty 2020) who have also received basic education may trigger increased presence of mouthing and occasional lexical borrowings in specific conversational settings, topics, and interlocutors.

Low conventionalisation is found in core domains of the KK lexicon, paralleling other micro-community sign languages across the world (e.g. Washabaugh 1986; Nyst 2007; Schuit 2014). Both colour and kinship terms show a limited paradigm of lexicalised signs in KK. KK uses the four lexicalised colour signs WHITE, BLACK, RED, and GRUE (de Vos 2011) and the three lexicalised kinship signs MOTHER, FATHER, and OFFSPRING (de Vos 2012).

High variation in the KK lexicon is partly influenced by social factors. Mudd et al. (2020) analyse the first target sign in response to a picture elicitation task of 36 common concepts from 20 deaf and 26 hearing KK signers to determine the effect of social factors on sign variation. Using measures for lexical distance and neighbourhood density, they find that gender and hearing status, but not other social variables, such as generation, may predict the use of specific signs (Mudd, Lutzenberger, et al. 2020). This contrasts with studies on macro-community sign languages that often find that age and region strongly predict the choice of sign variants (e.g. Bayley et al. 2002; Stamp et al. 2014).

Neither the lexicon nor the phonology of KK have been studied in detail. Marsaja (2008) summarises basic building blocks of KK in terms of handshapes, locations, and movement features attested in his corpus. However, their status as phonetic or phonologically contrastive features remains unclear. De Vos (2012) revisits Marsaja's classification and suggests some modifications in terms of frequency or discrepancy of a few handshapes. Crucially, KK uses an extended signing space, reflected in some sign locations that are infrequent or even unattested in other sign languages, e.g. the hip or the teeth (Marsaja 2008; de Vos 2012). Lutzenberger (2018) finds major cross-linguistic differences in the phonological characteristics of name signs, a particular group of signs attributed to individuals, between KK and NGT with respect to the use of locations, nonmanuals, and specific handshapes. Similar to preliminary characteristics of KK features and the atypical use of space (Marsaja 2008; de Vos 2012), these findings indicate typologically distinct patterns in KK's phonology.

To sum up, KK shows typological differences to other sign languages in terms of various aspects of its phonology and lexicon, and high degrees of variation in the lexicon.

2.4 Present study

Studies investigating spoken languages and macro-community sign languages often treat variation as a rich resource to answer questions about linguistic diversity, inter-speaker variability, and linguistic landscapes of communities. However, studies on sign language emergence typically focus on convergence between signers rather than variation. In this study, we aim to reach a better understanding of the phonological properties of the lexicon in KK through examining variation. To that end, we ask the following question: what is the variation in the form of signs found across 20 deaf KK signers in response to the same picture prompts as in Mudd et al. (2020)? We combine methods from lexicostatistics and previous work on emerging phonology in order to analyse the present state of the lexicon in KK. Systematic feature-based comparisons used in lexicostatistics have not previously been adapted to data from micro-community sign languages, and common analyses used in sociolinguistics are unsuitable for micro-community sign languages as it is hard to define a standard variant, as explained by Mudd et al. (2020). The measure introduced in this study combines techniques of all those methods, taking into account all relevant signs in a response from a large-scale sample of 60% of the deaf adult KK signers.

3 Methodology
3.1 Participants

The sample in this study comprises the same 20 deaf signers (11 female) of Mudd et al. (2020), all of whom permanently reside in the village and use KK as their primary mode of communication. Participants are sampled from generation three (ages ~65–80) through five (ages ~18–35) (see *Table 1*), as there are no deaf generation two (ages ~80+) signers alive and deaf generation six (ages ~2–5) signers are young children. We also sampled in such a way to maximise diversity in the socio-demographic profiles of the signers.[4]

Code	G	KI	MR	SA	SM	KE	KR	MG	NG	PA	PU	RA	SK	MU	RI	JU	P1	P2	SB	SY
Generation	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5	5
Gender	M	F	F	M	M	F	M	F	M	M	F	F	F	F	F	M	F	F	M	M

3.2 Stimuli

Due to the low literacy rate among our participants, we used picture stimuli. Participants were shown 36 pictures of common objects, spanning the semantic domains *food, animals, colours, praying, miscellaneous* (see Appendix A for a full list). Pictures were either taken in the field during previous field trips, or found on the internet (for materials see Lutzenberger et al. 2018). Stimuli selection was informed by the authors' linguistic and cultural knowledge of the community. In addition, as the task was originally administered to both deaf and hearing participants (Mudd, Lutzenberger, et al. 2020), stimuli were selected to match the knowledge of hearing people with varying degrees of signing fluency. In order to get a broad sample of signs with different degrees of conventionalisation, stimuli were selected based on the expected lexical variation for each item (high, medium, low). These classifications drew on insights from the KK Corpus, consulting local deaf research assistants and language knowledge of the first author, a fluent KK signer.

3.3 Procedure

Data was collected during a field trip in November, 2018. Recording took place in an empty room, and the task was administered by a local deaf research assistant. Before each session, the task was explained to the participants and consent was obtained through signature or thumbprint. The research assistant, who was instructed to act as interlocutor, sat opposite the participant and independently navigated the participant through the stimuli one-by-one on a laptop. The entire session was videotaped using a Canon Legria HF G26 camera at 25 frames per second.

The present set of stimuli was embedded into a larger set of picture- and object-based elicitations. All participants had previously participated in this kind of elicitation and were thus fairly familiar with the task. For this reason, instructions provided to the participant were minimal and mostly delivered during the informed consent by using fictive examples while explaining the procedure. The research assistant invited the participant to respond to a picture through either directing the attention to the screen through eye gaze alone, or a short, signed construction that translates into 'What is this?'.[5]

3.4 Coding

The recorded responses were annotated using ELAN (Crasborn & Sloetjes 2008; *ELAN [Computer software]* 2020) and glossed using the KK dataset in the lexical database Global Signbank (Crasborn et al. 2018; Lutzenberger 2020). All coding was done by the first author, including annotations for activity of both hands of the signer, facial expressions, the stimulus item, any signing of the research assistant, the response type (descriptive, naming, example, etc.), and chunks directly related to the picture stimulus (*Figure 4*). In cases where the research assistant named the stimulus before the participant, all subsequent signs were marked as primed

---

4   For more information on the participants see Mudd et al. (2020).

5   The construction used by the deaf research assistant often included a backwards head tilt and a brow raise which are common markers of questions, and was sometimes combined with a point towards the laptop PT_{laptop} or a sequence of a pointing and a generic question sign PT_{laptop} *WHAT* PT_{laptop}.

(pGLOSS). Immediate and exact repetitions, i.e., next following sign produced by the participant is the same surface realisation as produced by the research assistant, were marked as xGLOSS and excluded from the analyses.
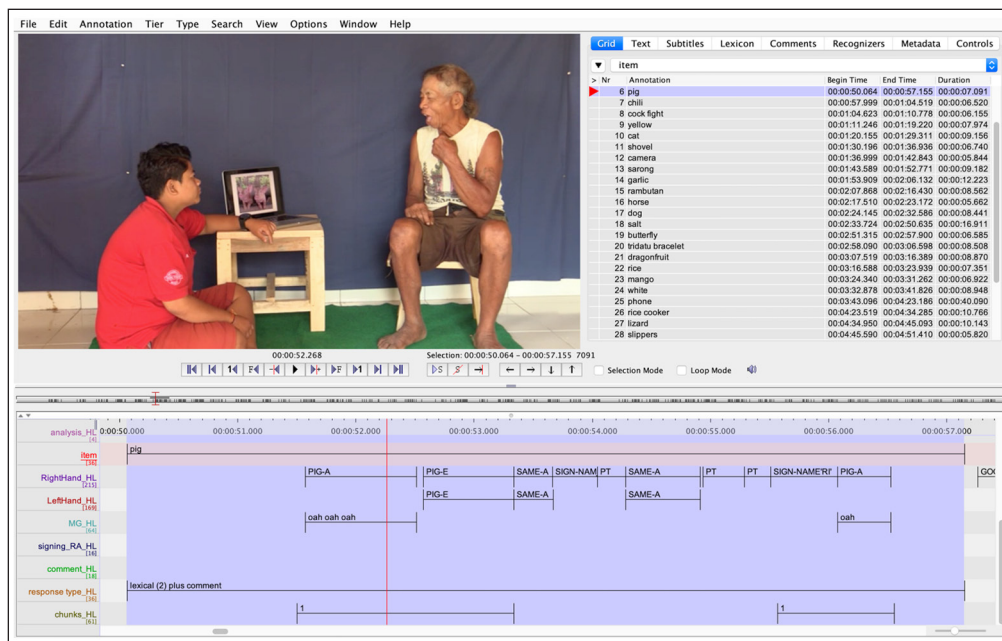
Every sign produced in response to a picture stimulus was coded for form in order to preserve minimal formational differences (phonological or phonetic). The Global Signbank dataset was edited alongside the annotation process, creating a new entry whenever there was no entry for a specific sign variant yet. Sign variants were grouped by shared iconic motivation using numbers, e.g., PIG-1 and PIG-2. Different realisations of the same iconic motivation are marked with capital letters e.g., COW-1A, COW-1B, COW-1C (*Figure 5*). The iconic motivation coding included finer-grained distinctions such as between walking and galloping legs of a horse, similar to what Ebling et al. (2015) refer to as *imaging techniques*. Similar to studies previously mentioned (Richie et al. 2014; Hou 2016; 2018; Horton 2018; Horton & Riggle 2019; Neveu 2019; Reed 2019; Reed & Rumsey 2019), this study uses iconic motivation as a grouping factor for sign variants. Since an analysis and in-depth discussion of iconic patterns and associated mappings is beyond the scope of this paper, we provide a list of iconic motivations for signs in this study in Appendix B and refer to Mudd et al. (2020) for further discussion of patterned iconicity in the same dataset. In order to provide a measure for coding reliability, the first author re-coded the iconic motivation, 11% of the items (N = 4; first two and last two items), as well as feature coding of surface realisations of 10% of the produced iconic motivations (randomly sampled). Intra-coder reliability scores are reported in percentage of overlap, and Cohen's Kappa for iconic motivation, chunks of target responses, and feature coding: iconic motivation (95% overlap; Kappa = 0.94), chunks of target responses (95% overlap; Kappa = 0.909), gloss of surface realisations (97% overlap), and feature coding (94.6% overlap; Kappa = 0.943). Videos and ELAN transcriptions are archived in The Language Archive (Lutzenberger et al. 2018).

3.5 Analyses

Unlike previous work (Israel 2009; Israel & Sandler 2009; Mudd, Lutzenberger, et al. 2020), we analyse all sign tokens naming or directly describing the picture and exclude only those parts

of answers that provide additional information. Responses were classified as off-target in case the stimulus was not recognised (correctly), and when signers provided lengthy explanations, personal narratives, descriptions of details and elaborations without explicitly referring to the target. In the following sections, we describe the analyses in detail. First, we explain the general idea and workings behind how variation may be measured. Then, we discuss how frequency can affect the proposed measuring and how it can be weighted to gain a more nuanced measure.

3.5.1 Measure of variation

In order to address the question as to how to quantify formational variation in our dataset, we developed a new measure of variation. This measure integrates three levels inherent to a sign (iconic motivation; surface realisation; feature differences; *Figure 6*), and results in a numeric and gradient score, the *variation index*.
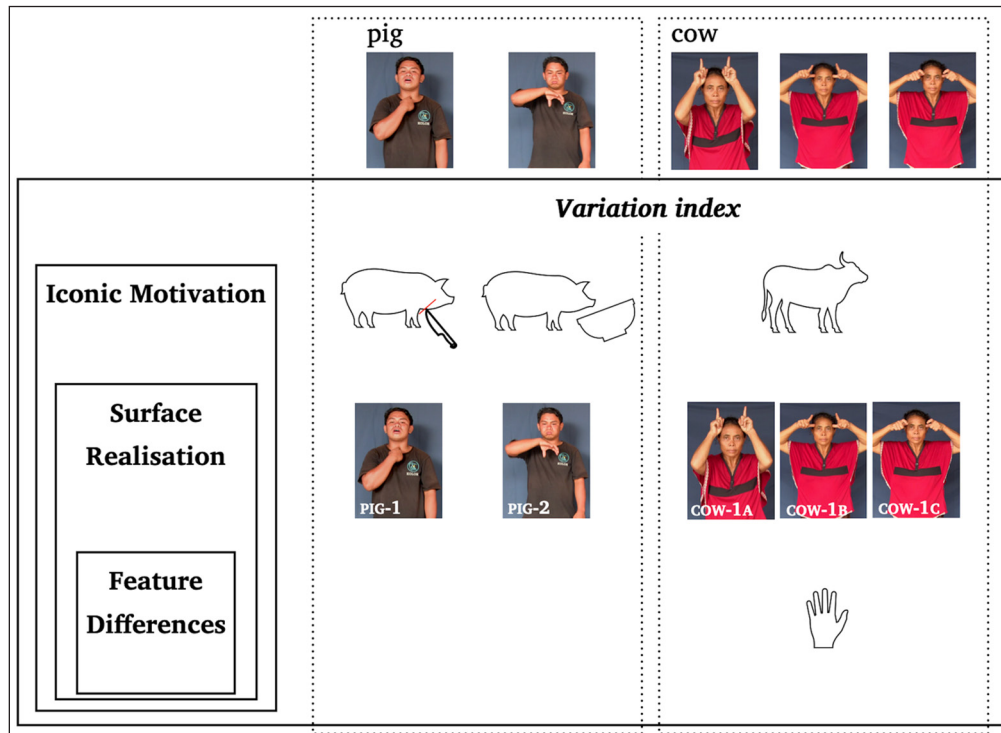

Figure 6 Nested structure of the different levels of the measure.

Iconic Motivation (IM) functions as a grouping criterion for sign variants. As pointed out by Israel (2009), it does not make sense to compare the form of signs that do not share the same iconic motivation as no form similarity is expected. Returning to the example of sign variants for *pig* and *cow* provided in *Figure 5*, we can identify different iconic mappings: the cow's horns in COW-1, killing a pig in PIG-1, and how a pig eats in PIG-2. Each is scored as 1 on the level of IM for the *variation index*. IM can be used independently as an accumulative count of unique iconic motivations occurring in response to a given stimulus. In *Figure 6*, accumulative IM yields a score of 2 for *pig* as responses include two iconic motivations, namely PIG-1 and PIG-2, and a score of 1 for the item *cow*.

Surface Realisation (SR) is a count of the different surface realisations with the same iconic motivation. In *Figure 5*, both variants for *pig*, PIG-1 and PIG-2, have a single surface realisation, resulting in each a score of 1 for SR. Among the COW-1 signs however, COW-1A, COW-1B, and COW-1C are three different surface realisations of the same iconic motivation. The SR for COW-1 is thus 3. Note that the relationship between IM and SR is minimally one-to-one; each iconic motivation occurs with a minimum of one SR.

Feature Differences (FD) explore the locus of variation within surface realisations of each iconic motivation. The feature coding is based on Global Signbank, comparing signs based on 14 different features:

- Handedness
- Hand Configuration Strong Hand
- Hand Configuration Weak Hand
- Handshape Change

- Relation between Articulators
- Location
- Relative Orientation Movement
- Relative Orientation Location
- Orientation Change
- Contact Type
- Movement Shape
- Movement Direction
- Repeated Movement
- Alternating Movement.

FD counts the number of features that differ across the attested surface realisations. In the case of the three COW-1 variants pictured in *Figure 5*, all differences concern the configuration of the hands; specifically, the curvature of the fingers and thumb extension of the strong hand and the weak hand. The FD in COW-1 is therefore 2.

Overall, this measure provides the opportunity to compare different levels as well as different sign variants. Within one item, we can compare variation across different iconic motivations based on the following formula:

$$Variation = \frac{\left(IM + SR + \frac{\left(\frac{FD}{14} \right)}{SR} \right)}{3}$$

A *variation index* is calculated per iconic motivation, hence IM is the constant 1; each *variation index* is normalised by dividing feature differences by the constant 14 for the 14 available features, and the number of surface realisations SR; the entire formula is divided by the three levels (IM, SR, FD). For example, COW-1 received a score of 1 for IM, a score of 3 on SR, and a score of 2 on FD. Each of these levels of variation are weighted equally, hence $(1 + 3 + (2/14)/3))/3 = 1.35$. Another fictious sign variant COW-2 with a different iconic motivation and a single surface realisation result in a *variation index* of 0.67 (see *Table 2*). The difference in variation between the two COW-variants is reflected in their *variation indices*: COW-1 shows a higher score, i.e., more variation, than COW-2. Note that a *variation index* of 0.67 is the minimal value of any sign variant with a single surface realisation because of the one-to-one relationship between iconic motivation and surface realisation. We will refer to this baseline as no variation here after and elaborate in Section 6 on some implications of the fact that the absence of variation may be caused by a hapax or extreme uniformity.

Item		Iconic motivation (IM)	IM	Surface Realisations (SR)	SR	Feature Differences (FD)	FD	*Variation index*
cow	COW-1A	COW-1	1	COW-1A	3	Hand configuration strong hand	2	**1.35**
	COW-1B	COW-1		COW-1B		Hand configuration weak hand		
	COW-1C	COW-1		COW-1C				
	COW-2A	COW-2	1	COW-2A	1	N/A	0	0.67
pig	PIG-1A	PIG-1	1	PIG-1A	1	N/A	0	0.67
	PIG-2A	PIG-2	1	PIG-2A	1	N/A	0	0.67

Table 2 Example calculation of *variation indices*.

Variation indices may be calculated per iconic motivation and account for the formal variation within an item at both the level of surface realisations and feature differences. Note that this measure focuses on the forms that occurred in the data. It does not take into account the linguistic environment of specific forms, any aspects of the participant/social factors, or the frequency of a sign variant. The following section discusses how frequency can be woven into the measure.

3.5.2 Weighted variation index

As a next step, we further develop the *variation index* in order to increase the ecological validity of the comparison among different sign variants. First, we combine the *variation index* with token frequency in the dataset. Second, we combine the *variation index* with the number of signers producing a specific variant, i.e., scope of use across the population. In other words, we apply weight by frequency, once token-based and once signer-based.

Token-based weighting aims to relativise the *variation index* in terms of reflecting high frequency or idiosyncratic sign variants. In the scope of this study, token frequency refers to sign variants produced within an item or pooled across all items. A *token-weighted variation index* is calculated as the product of the proportion of each sign variant and the unweighted *variation index* as explained in the section here above:

$$Variation_{token-weighted} = Variation * \frac{freq\, sign}{total\, signs}$$

First, we calculate *token-weighted variation indices* within items. We obtain the proportions for each iconic motivation as the number of tokens of an iconic motivation/surface realisation within an item divided by the total number of tokens of all iconic motivations/surface realisations of the particular item. For a fictious example, COW-1 has been produced at 20/30 tokens and COW-2 at 10/30 tokens in *cow*. Thus, COW-1 accounts for 67% and COW-2 for 33% of the 30 sign tokens. The *token-weighted variation index* for COW-1 then is 0.9 (1.35*0.67) and for COW-2 0.2 (0.67*0.33). Comparing the *token-weighted variation indices* more accurately captures the idea that COW-1 shows more variation while also being more frequent than COW-2.

Second, we move beyond the item by deriving proportional values for each sign variant from the entire dataset (tokens of sign variant/total tokens). Following the same approach, proportional values for iconic motivations or surface realisations are multiplied with an updated unweighted *variation index*. Updating the unweighted *variation index* is necessary as some sign variants may feature in responses to multiple picture stimuli, e.g., sign variants for RED are frequently produced in response to the picture stimulus *dragon fruit* (pitaya) as well as the colour *red*. Now, sign variants can be compared freely to each other, allowing us to fully capture the variation in the dataset.

However, token frequency does not account for a single signer producing a certain sign variant several times. We gain a more nuanced insight into how widespread specific sign variants are across our participant pool when applying weight through attested variants per signer. Here, we count how many of the 20 signers produce a specific sign variant in each item and then multiply the fraction with the unweighted *variation index*:

$$Variation_{signer-weighted} = Variation * \frac{N\, signer}{20}$$

4 Results
4.1 Descriptive results

The data yielded a total of 1,739 relevant sign variants (151 iconic motivations) to refer to 35 stimuli that entered the analyses. Due to confusability in the picture stimulus we excluded responses to *salt* altogether; the picture showed two different packages of salt, leading to misinterpretations of the picture and some signers creating contrast and others providing general descriptions. In all other items, off-target responses ($n = 24$) and immediate and exact repetition of signs prompted by the research assistant ($n = 21$) were excluded. Note that we deal with the form of signs and their token frequency in the data obtained in this experiment only. Characteristics of the participants, e.g., individual verbosity or correlations of idiosyncratic variants are not under investigation here. Initial analyses of selected social variables have been addressed in Mudd et al. (2020).

In line with Mudd et al. (2020), we observe variation even when considering only the deaf participants. Different stimulus items elicit different numbers of iconic motivations, which, in turn, differ greatly in the number of surface realisations and their feature differences. The number of iconic motivations documented across the 35 included items ranges from one single iconic motivation in *dog, chicken,* and *cow* to as many as nine different iconic motivations in *rice*

cooker (see *Figure 7*).[6] This broad range may indicate different degrees of conventionalisation in sign variants. Equally possible is a more general problem inherent to the nature of picture elicitations, namely differing degrees of precision when naming or describing the picture stimuli, or both.
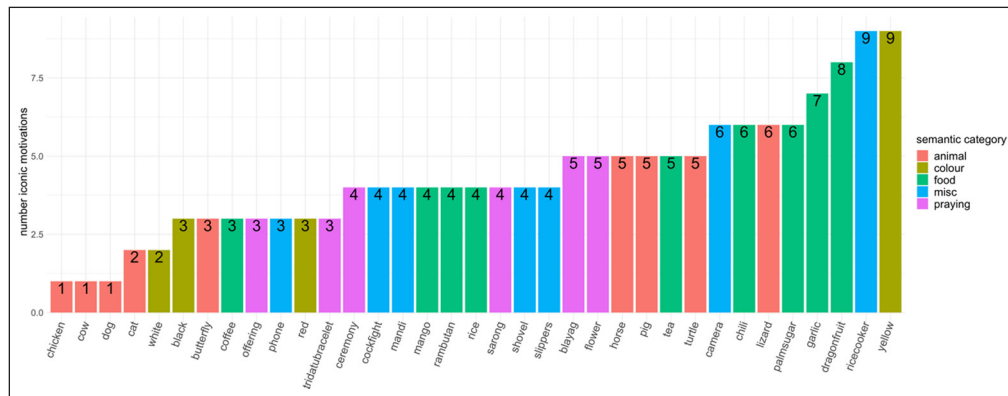
Across all items, iconic motivations with a single surface realisation are frequent (68/151, 45%). Conversely, 55% of the data includes iconic motivations with more than one surface realisation. Specifically, iconic motivations with two (34/151, 22.5%), three (27/151, 17.9%), four (10/151, 6.6%), five (4/151, 2.7%), six (5/151, 3.3%), seven (2/151, 1.3%), and even with eight (1/151, 0.7%) surface realisations are attested. On the level of feature differences, variants of the same iconic motivation vary between 0 to 12 features. As discussed in Section 3.5.1, comparisons of formal aspects across different iconic motivations are not meaningful as no similarity is expected.

4.2 Variation index

Variation indices are calculated for each iconic motivation in each item. *Figure 8* plots all signs in the analysis and their respective *variation indices* by item. The majority of signs shows a *variation index* between 1 and 1.5 (*mean* = 1.43; *sd* = 0.63; *range* = 0.67–3.03). Signs near the y-axis show no variation (*variation index* 0.67), mostly due to a single surface realisation. Signs with a *variation index* higher than 2 are infrequent.
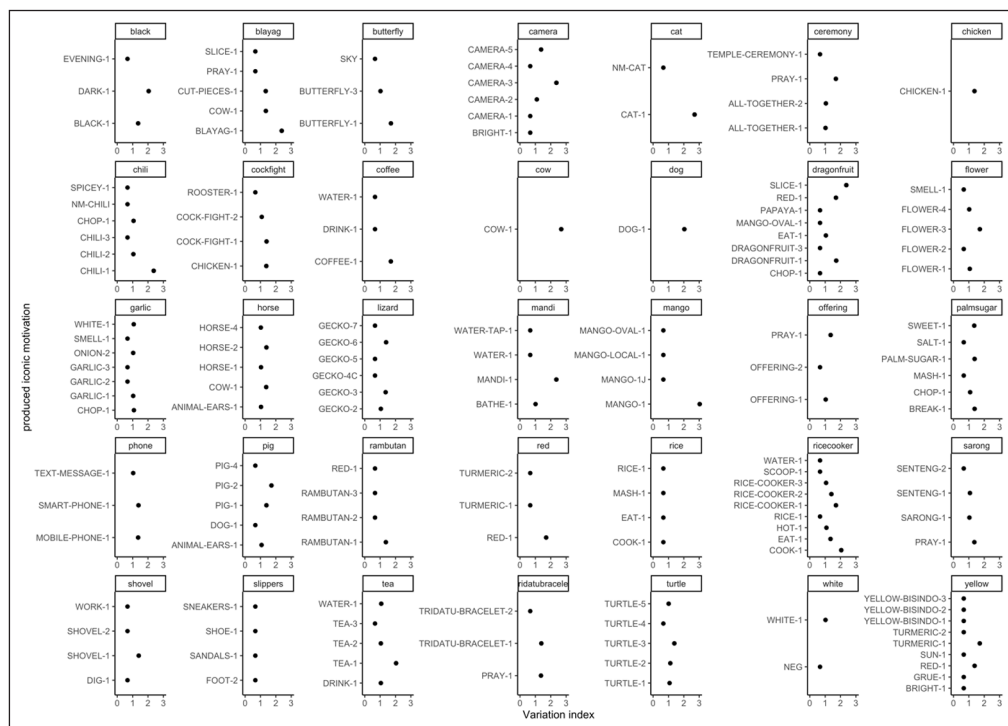
6   Note that the number of iconic motivations elicited per item differs from the one reported in Mudd and colleagues (2020) as this study is based on the full response rather than a single target sign per signer for each stimulus.

Variation occurs to different degrees on different levels. Stimuli that are represented by a single iconic motivation such as *chicken, dog,* and *cow* show different *variation indices* caused by variation on the surface and feature level (*Figures 8* and *9*). CHICKEN-1 has three different surface realisations of the same iconic motivation while there are five for DOG-1 and seven for COW-1 (*Figure 9*). The three surface realisations of CHICKEN-1 vary in three features, the five surface realisations of DOG-1 are a result of differences in six features, and the seven surface realisations of COW-1 differ in four features. Taken together, these differences reflect in different *variation indices* for CHICKEN-1 (1.37), DOG-1 (2.03), and COW-1 (2.68). These examples clearly show that items eliciting a single iconic motivation do not necessarily yield similar degrees of variation on the other levels.

a) CHICKEN-1   b) DOG-1   c) COW-1

Figure 9 Variation across three items eliciting one iconic motivation: *chicken, dog,* and *cow.*

For items eliciting multiple iconic motivations such as *camera, variation indices* do not distribute evenly across all iconic motivations as sign tokens split into more iconic motivations (*Figure 8*). Three of the six iconic motivations in response to the picture stimulus *camera*, BRIGHT-1, CAMERA-1, and CAMERA-4, occur with a single surface realisation, resulting in no variation (*variation index* 0.67). The other three iconic motivations show different *variation indices*: CAMERA-2 has two realisations that differ in seven features (*variation index* 1.10), CAMERA-3 has six realisations with nine feature differences (*variation index* 2.37), and CAMERA-5 has three variants varying in five features (*variation index* 1.37) (*Figure 10*). Once again, this shows that (i) variation emerges on different levels that may vary independently from each other (i.e., fewer surface realisations do not always mean fewer feature differences), and (ii) some iconic motivations show no variation.
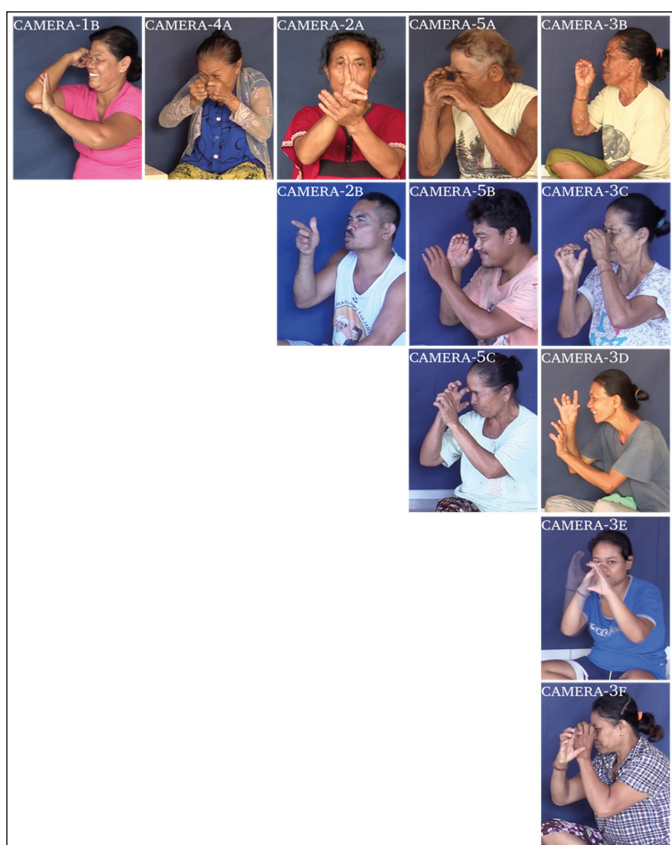

Figure 10 Variation across selected variants elicited to the prompt *camera.*

The *variation indices* reported until now quantify variation in terms of unique iconic motivations. However, the more responses feature a certain iconic motivation, the more opportunity it has to show variation. Sign variants that are produced only once can only have a single surface realisation, resulting in no variation, while sign variants that occur × times could have maximally × different surface realisations and could differ in up to 14 features. Token frequency shapes how much variation can be attested in a given dataset, and indeed, *variation indices* correlate positively with frequency ($r = 0.69$, $p < .001$): sign variants that have been produced more often tend to have higher *variation indices*.

4.3 Weighted variation

In the following, we report *weighted variation indices*, first taking into account token frequency, and then the number of signers producing specific sign variants as laid out in Section 3.5.2.

The *token-weighted variation index* within an item shows different effects on stimuli eliciting a single iconic motivation or multiple ones. For items such as *chicken, dog,* and *cow* that elicited a single iconic motivation, weighting results in identical scores as the weighting factor is 1: the *token-weighted variation index* remains at 1.36 for CHICKEN-1, at 2.03 for DOG-1 and at 2.68 for COW-1 (*Figure 11*). For items that elicited more iconic motivations such as *camera*, the *token-weighted variation index* highlights differences: BRIGHT-1 (token-weighted 0.04; unweighted 0.67), CAMERA-1 (token-weighted 0.09; unweighted 0.67), CAMERA-2 (weighted 0.12; unweighted 1.10), CAMERA-3 (token-weighted 1.0; unweighted 2.37), CAMERA-4 (token-weighted 0.05; unweighted 0.67) and CAMERA-5 (token-weighted 0.29; unweighted 1.37), reflecting the amount of variation in light of how much opportunity a sign variant has to vary. CAMERA-3 remains the variant with the highest *token-weighted variation index* among the iconic motivations in this item, yet the variation is now proportional to the high number of signs corresponding to this iconic motivation (*Figure 11*). This facilitates comparisons of variation across iconic motivations from the same item. In short, a *token-weighted variation index* approximates the actual productions in the data better than the unweighted *variation index* and thus improves how the scores reflect the observed variation.
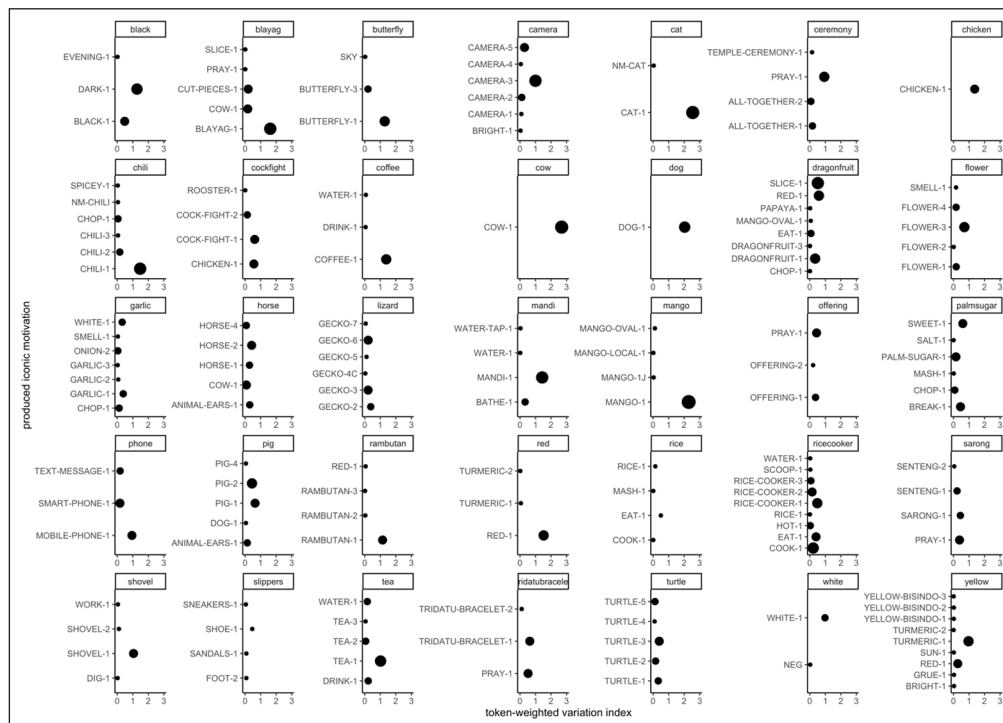

Figure 11 *Token-weighted variation* for iconic motivations produced in each item, points adjusted by frequency.

Treating the data as a mini-corpus, i.e., no grouping by item, the *token-weighted variation index* identifies a tendency for dominant variants. A large part of the data clusters with low(er) *token-weighted variation indices*, indicating no dominant variants (see Appendix C for all datapoints with variation), yet dominant surface realisations become apparent through a considerably higher *token-weighted variation index*. *Figure 12* provides only a snippet of the data for the sake of

readability, showing surface realisations elicited in three selected items with different degrees of variation: *tea, dog,* and *camera*. TEA-variants cluster together, showing a wider distribution of *token-weighted variation indices*; no clearly dominant variant becomes apparent. In contrast, DOG-1A ($n = 24$; token-weighted 0.03) shows higher *token-weighted variation index* than DOG-1B ($n = 3$; token-weighted 0.004), DOG-1C ($n = 3$; token-weighted 0.004), DOG-1D ($n = 1$; token-weighted 0.001), or DOG-1E ($n = 3$; token-weighted 0.004), due to considerably higher frequency applied to the same unweighted *variation index* (*Figure 12*). Similarly, CAMERA-3A ($n = 24$; token-weighted 0.01) has the highest *token-weighted variation index* out of all CAMERA-variants as it occurred most frequently in the data ($n = 7$). Note that the power of frequency is particularly visible in iconic motivations that occur in response to more than one stimulus, leading to a very high overall frequency (e.g., RED-1 was produced in 68 tokens as compared to 24 DOG-1 tokens and 16 CAMERA-3 tokens). However, we must not forget that a considerable portion of the data shows no variation.
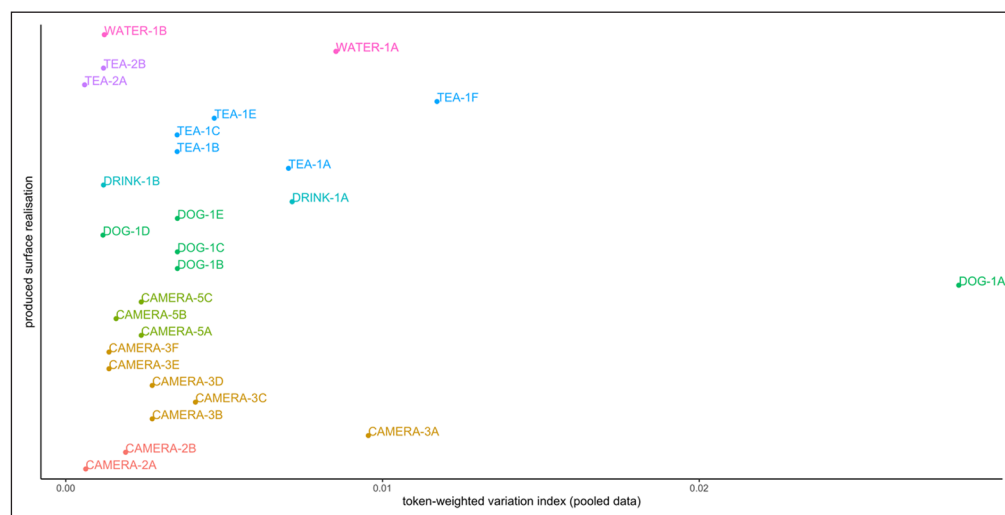

Figure 12 Examples of *token-weighted variation index* for surface realisations, pooled across all data and colour-coded by iconic motivation. Plot includes only sign variants with variation (>1 surface realisation).

The *signer-weighted variation index* corroborates dominant variants. Whereas weight by token frequency only identifies which variants are produced often, weight by signer identifies how widespread variants are across participants. For ease of readability, *Figure 13* illustrates the same selected examples as in *Figure 12* (find full graph with all data with variation in Appendix D). *Figure 13* demonstrates that TEA-1 (brewing loose leaves) is more widespread across participants than TEA-2 (steeping tea bag). Indeed, we may even pinpoint specific dominant surface realisations. Within TEA-1 (unweighted 2.04), two surface realisations are produced by 12/20 signers: TEA-1A (signer-weighted 0.51) and TEA-1F (signer-weighted 0.71), only differing in handedness (TEA-1A is two-handed; TEA-1F is one-handed). Maybe even more striking is the example of DOG-1 variants (unweighted 2.03): DOG-1A (signer-weighted 1.42) was produced by 14/20 signers, DOG-1B (signer-weighted 0.20) and DOG-1C (signer-weighted 0.20) by two and DOG-1D (signer-weighted 0.10) and DOG-1E (signer-weighted 0.10) by one participant. Without doubt, DOG-1A represents the most widespread variant and the high unweighted *variation index* stems from variants produced by only few participants. These examples are clear evidence of (i) patterned variation that often includes particularly widespread surface realisations, and (ii) variation as a result of few signers producing non-dominant variants, either due to idiosyncratic characteristics (potentially caused by individual socio-demographic profiles), or preference of another iconic motivations.

To sum up, both frequency-based weightings improve the ecological validity of the measure; a *token-weighted variation index* demonstrates how tightly variation and frequency are intertwined, a *signer-weighted variation index* identifies the dispersion of sign variants both in terms of iconic motivations and surface realisations across the population. In other words, *Figures 12* and *13* as well as the graphs of the full data (Appendix C and Appendix D) provide two different ways of looking at the same dataset (i.e. the effect of, first, the number of produced tokens, and second, the number of signers using a particular variant). Nevertheless, both point towards the same dominant variants, i.e., iconic motivations and/or surface realisations that are (i) more frequent than others (token-weighted) and (ii) more widespread across participants (signer-weighted).

Given that we allow for one signer producing multiple signs, the *signer-weighted variation index* more clearly displays dominant variants than the *frequency-weighted variation index*.
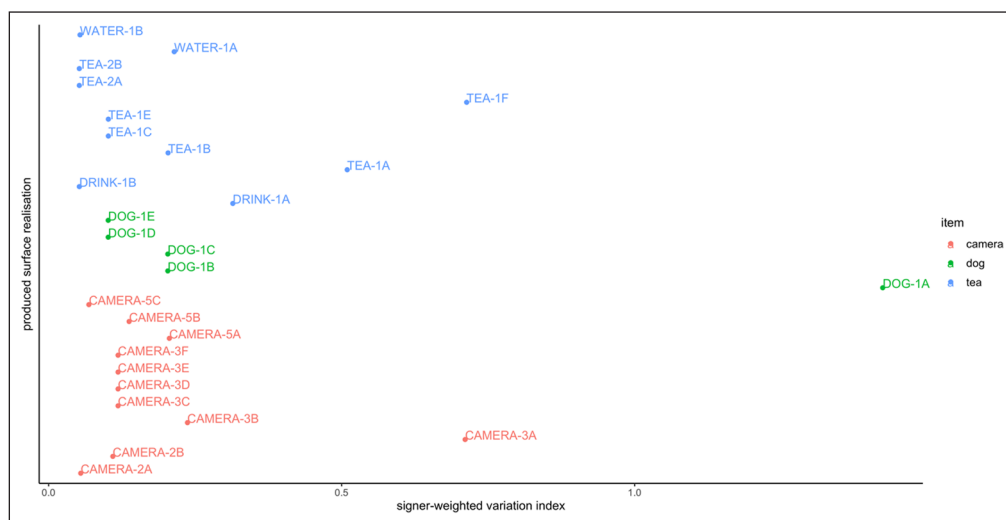
Figure 13 Examples of *signer-weighted variation index* for surface realisations, colour-coded for item. Plot includes only sign variants with *variation* (>1 surface realisation).

5 Measuring variation

Studies in sign language emergence focus on the iconic motivation as developing systems may not yet exhibit phonological structure. The following section is dedicated to laying out how the *variation index* compares to and improves existing methods. To this end, we provide a case study of three items from our data comparing our method to the measure of Israel (2009) and Israel and Sandler (2009): mode and number of variants. As a reminder, Israel and Sandler narrow their analysis to the most frequent iconic motivation provided in responses to a picture stimulus; they then do a feature analysis of this subset of signs calculating mode and number of variants for each feature class. Note that the mode is calculated relative to the number of signers who produce this iconic motivation, leading to different items having differently sized sets. Israel and Sandler use the number of surface realisations as a measure of sign-level variation.

For this case study, we re-use the three stimuli from *Figures 12* and *13* as they elicited different degrees of variation: *tea, dog,* and *camera*. Following Israel and Sandler's method of analysing one-sign responses, we selected the first target sign in each response when the response contained multiple. For each of the three items, we identified the most frequent iconic motivation (TEA-1, DOG-1, CAMERA-3) and then calculated the mode and the number of variants for each feature as described in Israel (2009) (*Figure 14A*). The set sizes differ greatly: DOG-1 was produced by 19 signers, TEA-1 by 16 signers, and CAMERA-3 by nine signers.
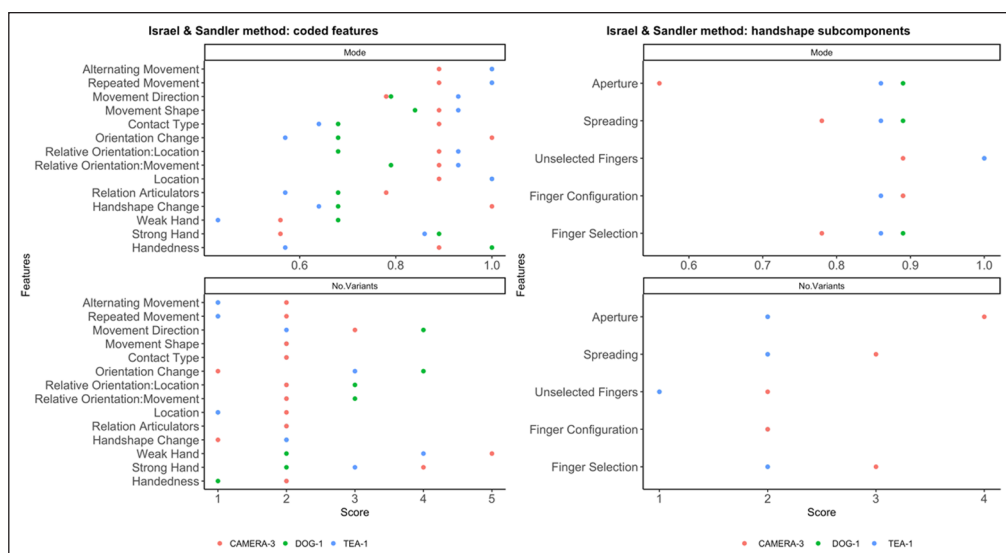

Figure 14 Visualisation of mode and number of variant measures for the selected items, using all coded features in the current study and some subcomponents of handshape.

Figure 14 illustrates that DOG-1, TEA-1 and CAMERA-3 differ in the variation they exhibit in different feature classes. A low number of variants often coincides with high mode (i.e., low variation); for example, TEA-1 has only one variant in Location, Repeated Movement and Alternating Movement and accordingly shows maximal mode in these features while CAMERA-3's four variants in the handshape feature Aperture co-occur with a mode of 0.56 in this feature. Moreover, DOG-1 is an example that nearly all signers produced, i.e., it represents the example with the largest set, and displays similar amounts of variation in different handshape subcomponents of the Strong Hand (2.23 variants and mode of 0.89): 1.9 variants in Finger Selection, 1.77 variants in Aperture, 1.66 in Spreading and 1.56 in Finger Configuration, and a mode of 0.87 for Finger Configuration, 0.85 for Spreading and 0.84 for both Finger Selection and Aperture. Sign-level variation, captured as the number of surface realisations in Israel (2009), is diverse in this case study: five variants were produced for DOG-1, five for TEA-1, and six for CAMERA-3. Hence, on this measure, DOG-1 and TEA-1 are less variable than CAMERA-3.

Figure 13 exemplifies the same surface realisations using the *signer-weighted variation*. We chose the *signer-weighted variation index* as a baseline for comparison to Israel and Sandler's method as the analysed iconic motivation is identified by the number of signers. In line with this method, DOG-1, TEA-1 and CAMERA-3 are the dominant iconic motivations using the *variation index*. Nevertheless, *Figure 13* demonstrates that in many cases, more than one iconic motivation is frequent, e.g., TEA-1 is used by 16 signers and DRINK-1 by seven signers in *tea* and CAMERA-3 was produced by nine signers and CAMERA-5 by six signers in *camera*. This may be linked to including multiple signs per response; the knowledge of signers is not restricted to a single variant, and indeed, DRINK-1 sometimes features the same response as TEA-1. In cases like *camera*, it may be less clear what variant should be considered the most frequent one; CAMERA-3 is only marginally more widespread than CAMERA-5, which suggests that analysing more than one iconic motivation is necessary to appropriately represent the variation. In addition to identifying the most widely shared iconic motivations, we find that specific surface realisations are more frequent than others with the same iconic motivation: DOG-1A (14/20 signers), TEA-1F (7/20 signers), and CAMERA-3A (6/20 signers). These surface realisations have higher *signer-weighted variation indices* (*Figure 13*) because they are shared across more signers. In short, while zeroing in on the most frequent iconic motivation is one way of reducing "noise", the *variation index* shows that preserving variation through factoring in token-frequency or signer-frequency later on may be helpful to uncover structure in variation.

While both the *variation index* and Israel and Sandler's method point to different degrees of variation within a sign, one of the major differences between the methods lies in the measure itself. Israel and Sandler's method presents us with many connected measures rather than one unified measure. Their studies aim at a cross-linguistic comparison of phonological stability on the language-level which might explain why they opted for this method. However, features do not occur in isolation and we therefore argue that separate feature measures need to be re-incorporated into the context of the sign. The *variation index* aims for a more encompassing measure that unifies different levels of a sign and takes into account contextual information, here token- and signer-frequency. In contrast to Israel and Sandler's measures, the *variation index* does not integrate particular feature values but instead tallies the mismatching features. Accounting for feature value differences could be incorporated into the *variation index* as an extra level in the future, potentially even by adapting Israel and Sandler's mode.

Israel and Sandler's method is driven by maximal convergence between signers while the *variation index* is based on charting variation, and thus capturing structure in variation also in less frequent sign variants. There are three main advantages of the *variation index*: (i) in line with many other studies, the measures of Israel and Sandler (2009) are based on one-sign-per-response type of data (e.g. Sandler et al. 2011; Hartzell et al. 2019; Mudd, Lutzenberger, et al. 2020). As multi-sign responses are commonly reported for lexical elicitation data from micro-community sign languages (e.g. Morgan 2015; Hartzell et al. 2019; Mudd, Lutzenberger, et al. 2020), the *variation index* accommodates responses with multiple signs. (ii) Mode and number of variants focus on the most frequent iconic motivation, leading to eliminating much of the variation in the very first step. By examining all produced iconic motivations, the *variation index* allows for participants knowing and providing synonyms or multiple variants in an elicitation task. (iii) Israel and Sandler's and our own data show that there is a lot of variation as to whether signers produce the same iconic motivation in response to an item. For some

items, signers highly aligned on iconic motivation (e.g. *dog, chicken, cow*) while for others they are very dispersed (e.g. *rice cooker*). Rather than selecting only one iconic motivation (which may be only marginally more frequent), the *variation index* factors in token-frequency and participants as a weighting factor later on and thus preserves variation. This enables us to uncover conventions also in less frequent iconic motivations. Although the *variation index* may not (yet) integrate exact feature value measures, it accounts better for the ecological realities of a signing community. Integrating all levels of a sign yields not only a more comprehensive but also a more nuanced account of variation that allows us to objectively assess and compare variation within and across emerging and established systems. In addition, the automatic comparison makes this method easy to be scaled up to large datasets.

Nevertheless, there are also some shortcomings of the *variation index*. The three levels considered are all interrelated which may make the *variation index* somewhat counter-intuitive. While iconic motivation and surface realisations stand in a one-to-one relationship to each other, this is not the case for surface realisations and feature differences. Surface realisations that differ from each other but share the same iconic motivation are grouped together. Form variation arises within an iconic motivation as minimally one surface realisation. Variation among these surface realisations is driven by differing features, resulting in more (or less) surface realisations than feature differences. Despite the different relations to each other, all levels contribute to the overall score where a higher value represents more variation. The major contributor of variation in a variant can, however, only be localised in a separate step; each level can provide a different piece of the puzzle, e.g., the number of produced iconic motivations per item provides information about how many mappings are used to refer to a given concept. Altogether, the *variation index* builds on the complexity of interrelated levels to provide a comprehensive approach to variation.

As explained previously, no variation means that an iconic motivation has been produced with a single surface realisation. Due to the one-to-one relationship between iconic motivation and surface realisation, this yields the *variation index* of 0.67. For the majority of cases, no variation is due to the lack of tokens; hapaxes i.e., single tokens of (idiosyncratic) sign variants cannot show variation. A limited number of signs, however, show no variation due to extreme uniformity across all tokens. SHOE-1A ($n_{token} = 29$; $n_{signer} = 18$), OFFERING-2A ($n_{token} = 19$; $n_{signer} = 10$), TEMPLE-CEREMONY-1A ($n_{token} = 13$; $n_{signer} = 8$), RICE-1A ($n_{token} = 12$; $n_{signer} = 8$), MANGO-OVAL-1A ($n_{token} = 12$; $n_{signer} = 4$), GECKO-5A ($n_{token} = 10$; $n_{signer} = 5$), TRIDATU-BRACELET-2A ($n_{token} = 9$; $n_{signer} = 4$) have been produced with a high(er) frequency but with a single surface form. Thus, albeit less frequent, a *variation index* of 0.67 may also result from the lack of type variation. Developing the *variation index* further could aim for normalising the scores to make them more easily interpretable and potentially even circumvent this ambiguity.

6 Discussion

This study has introduced and applied a new way of measuring variation in sign formation across a signing community numerically, taking into account three interrelated levels inherent to every sign: 1) iconic motivation, 2) surface realisation, 3) feature differences. As such, it builds on previous work which primarily focuses on the feature level (Israel 2009; Israel & Sandler 2009; Parks 2011; Sandler et al. 2011; Morgan 2015; Omardeen 2018; Börstell et al. 2020). These *variation indices* yield gradient outcome measures for sign variants, allowing us to capture that, across the 20 deaf signers sampled, different sign variants exhibit various degrees of variation on different levels. Moreover, this study shows the impact of frequency on what variation can be attested, cautioning the generalizations that can be made from a limited dataset. We have suggested two types of frequency-based weightings of the *variation index* to increase the ecological validity of the measure; weighting by token frequency identifies sign variants of high usage and weighting by signer identifies sign variants that are particularly widespread across the population. Both identify dominant and non-dominant variants, demonstrating a variation continuum.

6.1 Variation and frequency

This study demonstrates that limited datasets are at risk of overinflating variation; signs that are produced frequently have more opportunity to vary while a single token of a sign variant

cannot vary from anything. Indeed, most studies on the emergence of phonology and the lexicon in micro-community sign languages are based on picture elicitations instead of corpus data with robust frequency information (e.g. Israel 2009; Israel & Sandler 2009; Sandler et al. 2011; Richie et al. 2014; Morgan 2015; Horton 2018; Hou 2018; Reed 2019; Hartzell et al. 2019). Variation as measured in this study can be applied to different kinds of data, ideally drawing on both elicited and spontaneous corpus data from different languages or even home sign data that can be compared directly.

Previous analyses on emerging phonology have focused on commonalities across signers, by singling out the most frequent iconic motivation and/or features among responses to a stimulus for analysing variation (Israel 2009; Israel & Sandler 2009; Sandler et al. 2011). For example, Sandler and colleagues (Israel 2009; Israel & Sandler 2009; Sandler et al. 2011) view the most frequent iconic motivation as (most) conventionalised. On these grounds, they examine variation by zooming in on the iconic motivation of barking for *dog* in ABSL, and examine handshape variation within that subset. Our results suggest that such methods might overinflate estimates of variation, given our observation that high frequency of an iconic motivation does not necessarily equal low variation on other levels, and that high-variation variants do not all vary to the same extent and in the same aspects. Especially in studies with limited datasets i.e. elicited data, it is fundamental to understand and acknowledge the correlation between frequency and variation to avoid overinterpretation.

In corpus-based research on spoken languages it has been established that high frequency words change more quickly than low frequency words (Frisch 1996; Bybee & Hopper 2001; Bybee 2010). We found that highly frequent iconic motivations often go hand in hand with high *variation indices*. As indicated previously, this may indeed be related to the increased opportunity for variation to surface. This phenomenon may also be linked to forces of language change, with high frequency signs demonstrating a locus of rapid language change while less frequent sign variants may be less variable. Nevertheless, our study deals with frequency as the tokens within this limited dataset of elicited productions rather than corpus-based frequencies. It is unclear how the type of language task influences the obtained frequency distributions. To corroborate whether or not the variation patterns in this data set are related to token-frequency or reflect language change, we would need to expand the dataset.

6.2 High variation and synonyms

Micro-community sign languages have often been described to exhibit a high degree of variation (e.g. Washabaugh 1986; Meir et al. 2012). Multiple explanations have been suggested to account for this, including the lack of pressure to converge on linguistic symbols due to high common ground (de Vos 2011; Meir et al. 2012).

Morgan (2015) observes that iconic motivations themselves are a locus of variation in KSL. We observe the same in KK; few items in our study elicit a single iconic motivation, e.g., *dog*, while the majority elicit multiple iconic motivations, e.g., *camera* or *pig*. Of the elicited iconic motivations, some are both more frequent and widely shared across signers, e.g., PIG-1 with 25 tokens across 17 signers compared to PIG-2 with 14 tokens across eight signers and ANIMAL-EARS-1 with seven tokens by four signers. Morgan (2015) attributes the presence of multiple iconic motivations to ongoing convergence across signers and Mudd et al. (2020) argue that high familiarity with a concept may reduce variation. On top of this, greater prominence of the stimulus and/or the iconic mapping in one's surrounding might also trigger (individual) preferences for certain iconic mappings and thereby stimulate the selection and persistence of different iconic motivations. For example, pigs are commonly killed by men while feeding pigs might be considered more (but not exclusively) a female task. To test and disentangle these hypotheses empirically, iconicity ratings would have to be collected, potentially alongside a measure of visual prominence.

Research on patterned iconicity suggests that specific types of objects tend to device specific iconic strategies, i.e. signs for tools often relate to handling or manipulating the tool while signs for food items are often linked to size, shape and manipulation (Padden et al. 2013; 2015; Hwang et al. 2017; Hou 2018). In line with this, sign variants elicited for the items *camera* and *pig* show different strategies: most iconic motivations elicited for *camera* relate to holding, handling, or manipulating a video camera whereas iconic motivations in *pig* map to

different aspects around a pig, namely handling/manipulating (killing), embodiment of the animal (feeding), and appearance. While it is possible that sign variants are indeed the result of selection and convergence as suggested in Morgan (2015), it is also possible that qualitatively different mappings such as in the PIG-variants affect the preservation of multiple different iconic motivations in KK and other micro-community sign languages.

Different ideas have been put forth to explain high variation in micro-community sign languages. Previously suggested by de Vos (2011), Meir and colleagues (2012) and Meir and Sandler (2019) and corroborated by a computational model by Thompson et al. (2019), the high degree of shared knowledge and the limited number of community members may allow for high variation. Both make it possible to tolerate idiosyncrasy, i.e., remembering idiosyncratic variants. The high overlap in experiences among signers of small communities may enhance the availability of a large range of possible iconic mappings and hereby accommodate for a large number of (idiosyncratic) sign variants. Recently, Tkachman and Hudson Kam (2020) have argued that rather than community size, tight kinship relations and early signing exposure across deaf signers accounts for the high lexical variation in young micro-community sign languages. While the present study adds that dominant sign variants are shared by many participants and produced often, the question of how to account for the considerable variation of non-dominant variants still remains.

Nevertheless, pressures leading to reducing the number of synonyms with different iconic motivations are unclear. As explained in Section 2.1, concepts provide many properties for potential iconic mappings that may be linguistically encoded. As a result, sign language lexicons may include synonyms with different iconic motivations. For example, three variants of DOG in NGT are based on three different iconic motivations (*Figure 15*): i) the dog's paws, ii) holding something in the mouth as dogs often do, iii) calling a dog by patting one's thigh.

Figure 15 Three sign variants with different iconic motivation for DOG in NGT. Images retrieved from Global Signbank (Crasborn et al. 2019).

DOG-A DOG-B DOG-C

If the reduction of iconic motivations is indeed a first step in conventionalisation as suggested by Morgan (2015: 14 f.), the abundance of synonyms with or without shared iconic motivation in macro-community sign languages points to an obvious lack of conventionalisation. In BSL, 22 conventionalised variants of PURPLE[7] have been found as a result of regional variation (Stamp et al. 2014). In contrast to work on sign language emergence, this variation in the BSL or the NGT lexicon is argued to be rooted in and maintained by various sociolinguistic factors; e.g., age, region, educational background. Altogether, it is unclear how different pressures resulting in the reduction of variation on the level of iconic motivations interact, and whether they differ fundamentally in different signing communities.

6.3 Variation and conventionalisation

Different fields approach variation in different ways: studies on macro-community sign languages embrace variation as sociolinguistic phenomenon, while, in research on sign language emergence, variation across signers is generally taken as a lack of conventionalisation. Where

---

7 BSL Signbank lists 17 different variants: *https://bslsignbank.ucl.ac.uk/dictionary/words/purple-1.html*.

Sandler et al. (2011) argue that the extreme variation in ABSL is explained by the lack of a phonological system, we would like to propose that a more in-depth approach may help to uncover processes and mechanisms underlying variation and conventionalisation.

This study shows that a gradient measure exposes structured variation. Sign variants for DOG in ABSL are presented as an example of extreme variation in Sandler et al. (2011). All DOG-variants share the iconic motivation 'barking' but all ten ABSL participants produce different surface forms, leading to the claim that ABSL signs are driven by holistic, iconic prototypes without combinatorial structure (Sandler et al. 2011: 520). Among our KK signers, *dog* elicited a single iconic motivation, also 'barking', with substantial variation in surface realisations and feature differences: DOG-variants in KK yield feature-level variation on six features distributed over five surface realisations. The *signer-weighted variation index* reflects that 70% of the KK participants produced the same surface realisation DOG-1A. In contrast to ABSL, the variation in DOG-1 results from six participants who produce different surface realisations. Thus, although *dog* elicits variation in both ABSL and KK, our approach gives more insight into the underlying structure of variation. In the case of DOG-1, we can localise the variation to 30% of the signers, suggesting that the driver is likely to be individual participant effects rather than the lack of a phonological system.[8] We plan to explore the effect of social factors on sub-lexical variation in a future study by combining the current method with the method used in Mudd et al. (2020).

Sign language emergence scenarios often describe a development from i) no language to ii) a communication system with high variation and little structure, in which variation decreases as structure increases until iii) reaching structural benchmarks set by research on macro-community sign languages (Meir & Sandler 2019). While this route may be justified for some linguistic aspects, for example grammatical elements (e.g. Pfau & Steinbach 2011; Senghas & Coppola 2011; Johnston et al. 2015; Pfau 2015; but see Safar 2020), the fundamental idea is nurtured by reducing variation to optimally low variation (without taking into account the ecological niche of the particular language). Macro-community sign languages, however, may previously have escaped this pressure and are, now characterised as established, analysed on different grounds. For example, levelling, i.e. the reduction of variation, is attributed to language emergence in Nicaraguan Sign Language for spatially modified verbs (Senghas 2003) while it has been explained as sociolinguistic variation in the BSL lexicon, an older macro-community sign language (Stamp et al. 2014). In macro-community sign languages, variation is increasingly perceived as a sign of linguistic diversity and richness, while variation in micro-community sign languages retains a negative connotation of immaturity (Moriarty Harrelson 2017; 2019; Kusters & Sahasrabudhe 2018; Braithwaite 2020; Hou & Kusters 2020; Kusters et al. 2020).

This tension may arise from the focus on convergence rather than variation in the literature on sign language emergence (Israel 2009; Israel & Sandler 2009; Morgan 2015; Meir & Sandler 2019). Conventionalisation and variation are often understood as "opposing forces" (Meir & Sandler 2019: 9), with the reduction of variation signaling conventionalisation. While we do not intend to question the general idea that decreasing variation increases conventionalisation, we would like to discuss two main problems with a reduction-based definition that are more broadly linked to a discussion about language emergence, language change, and language ideologies (Moriarty Harrelson 2017; 2019; Kusters & Sahasrabudhe 2018; Braithwaite 2020; Hou & Kusters 2020; Kusters et al. 2020): First, a reduction-based definition implies an ideal state of maximal convergence and minimal variation, i.e. a single variant, that any (emerging) language strives for. However, studies increasingly document substantial (sociolinguistic) variation on different levels of description across spoken languages (Bybee 2006; 2010) and well-researched macro-community sign languages (e.g. Stamp et al. 2014; Börstell & Östling 2016; Schembri et al. 2018). It is thus unclear why minimal variation is expected to be a hallmark of an "established language". Linked to this is the second issue, namely the paradox of determining a start and an end point of a process. Although conventionalisation is generally described as a process

---

[8]   In this paper, we are not taking position on whether or not KK exhibits duality of patterning or whether differing features are phonologically contrastive in KK for two reasons: 1) Perception experiments are needed to corroborate whether feature differences are phonologically contrastive or instances of phonetic variation; this study concerns production data and only very limited insights about perception is available for KK in this respect; 2) Van der Hulst & van der Kooij (2021) have argued that grammatical phonological rules are less common among signed than among spoken languages, and that phonology and morphology in sign languages are tightly intertwined which questions whether duality of patterning actually is a helpful concept to apply to sign languages.

(Burling 1999; Schmid 2020), signs tend to be discussed in categorical terms (lexicalised vs non-lexicalised, conventionalised vs. non-conventionalised). This generates the expectation of an ideal end state that can (and should) be reached, i.e. selection of one conventionalised variant and dismisses the fact that variation may be structured; like in other domains, the increasing trend to examine gradience in language use (e.g. Cormier et al. 2012; Ferrara & Halvorsen 2017; Lepic 2019) may be beneficial to understanding both variation and conventionalisation.

Rather than viewing conventionalisation and variation as opposing forces, we may want to move on to understanding them as two sides of the same coin. Lepic (2019) argues that understanding signs as a forced choice between two categories (lexicalised vs. non-lexicalised) instead of on a lexicalisation continuum prevents us from understanding the degree to which mental representations are established. Similarly, internalising conventionalisation (and variation) as continua may help us gain a deeper understanding of how they shape language emergence, language change, and language use. The findings of this paper stress that discussing conventionalisation in a binary manner can be misleading: unlike states, processes require gradient measures that are complex, sensitive to frequency and able to capture different degrees of variation on different aspects of a sign (see also Meir & Sandler 2019; Tkachman & Hudson Kam 2020). Understanding conventionalisation and variation as tightly related continua may allow us to acknowledge and deal with them more appropriately in different types of signing communities, paving the way for valid comparisons across languages that differ greatly in their socio-cultural and ecological niche.

7 Conclusion

This paper has developed and applied a new measure of variation in the forms of signs. The comprehensive approach to variation suggested here relies heavily on integrating three interrelated levels as they may show different degrees of variation. In other words, there is not always a logical trade-off between more and less variation among different levels within a sign; low variation on one level does not equal low variation on another level. Indeed, we propose to consider the possibility that variation does not necessarily equal to the absence of conventionalisation. We also suggest that, particularly in micro-community sign languages, footprints of individual language users may be especially prominent. Measuring variation as suggested here allows to calculate and weight *variation indices* for individual iconic motivations and surface realisations, which can then be utilised to answer further questions about the effect of social factors, or phonological characteristics. In this study, all signs produced in this dataset were treated as individual signs. As single sign responses are rare, *weighted variation indices* may be devised to address the issue of chains of signs, potentially identifying compounds and collocations in a more efficient way. Moreover, weight by signer touches upon social factors that may play a fundamental role in variation across small communities. In a future study, we plan to address the contribution of social factors in order to explore the source and context of sign variants, helping to further disentangle the high collinearity of social factors in micro-community sign languages such as KK. Our study teaches us to be cautious about overinterpreting data from restricted datasets and provides a new tool to examine and even compare variation within diverse sign language lexicons. This sets the ground for comprehensive comparisons of variation across micro- and macro-community sign languages with minimal methodological differences, bringing us closer to embrace variation irrespective of the ecological niche in which they emerged.

Abbreviations

ABSL = Al-Sayyid Bedouin Sign Language, ASL = American Sign Language, BISINDO = Indonesian Sign Language, BSL = British Sign Language, KK = Kata Kolok, KSL = Kenyan Sign Language, NGT = Sign Language of the Netherlands, TİD = Turkish Sign Language.

Acknowledgements

Competing interests

The authors have no competing interests to declare.

Author affiliations

Hannah Lutzenberger (iD) *orcid.org/0000-0003-0574-4060*
Center for Language Studies, Radboud University, Erasmusplein 1, 6525 HT Nijmegen, NL
International Max Planck Research School

Connie de Vos (iD) *orcid.org/0000-0002-4800-4313*
Tilburg Center for Cognition and Communication, Warandelaan 2, 5037 AB Tilburg, NL

Onno Crasborn (iD) *orcid.org/0000-0003-4985-9097*
Center for Language Studies, Radboud University, Erasmusplein 1, 6525 HT Nijmegen, NL

Paula Fikkert (iD) *orcid.org/0000-0001-7317-2100*
Center for Language Studies, Radboud University, Erasmusplein 1, 6525 HT Nijmegen, NL

References

Al-Fityani, Kinda & Carol Padden. 2008. A lexical comparison of sign languages in the Arab World. In Ronice Müller de Quadros (ed.), *Sign Languages: spinning and unraveling the past, present and future*, 2–14. Florianopolis, Brazil, December 2006: Editora Arara Azul. Petrópolis/RJ. Brazil. Retrieved from *http://www.editora-arara-azul.com.br/EstudosSurdos.php.*

Al-Fityani, Kinda & Carol Padden. 2010. Sign languages in the Arab world. In Diane Brentari (ed.), *Sign Languages*, 1st ed., 433–450. Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9780511712203.020*

Bank, Richard. 2015. *The ubiquity of mouthings in NGT: a corpus study*. Nijmegen: Radboud University dissertation. DOI: *https://doi.org/10.1075/sll.18.2.05ban*

Bayley, Robert, Ceil Lucas & Mary Rose. 2000. Variation in American Sign Language: The case of DEAF. *Journal of Sociolinguistics* 4(1). 81–107. DOI: *https://doi.org/10.1111/1467-9481.00104*

Bayley, Robert, Ceil Lucas & Mary Rose. 2002. Phonological variation in American Sign Language: The case of 1 handshape. *Language Variation and Change* 14(1). 19–53. DOI: *https://doi.org/10.1017/S0954394502141020*

Bickford, J. Albert. 1991. Lexical variation in Mexican Sign Language. *Sign Language Studies* 1072(1). 241–276. DOI: *https://doi.org/10.1353/sls.1991.0010*

Börstell, Carl, Onno Crasborn & Lori Whynot. 2020. Measuring lexical similarity across sign languages in Global Signbank. In Mayumi Bono, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie Hochgesang, Jette Kristoffersen & Johanna Mesch (eds.), *Proceedings of the 9th Workshop on the Representation of Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives [Language Resources and Evaluation Conference (LREC)]*, 21–26. Marseille: European Language Resources Association (ELRA).

Börstell, Carl & Robert Östling. 2016. Visualizing lects in a sign language corpus. In Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie Hochgesang, Jette Kristoffersen & Johanna Mesch (eds.), *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining [Language Resources and Evaluation Conference (LREC)]*, 13–18. Portorož: European Language Resources Association (ELRA).

Braithwaite, Ben. 2020. Ideologies of linguistic research on small sign languages in the global South: A Caribbean perspective. *Language & Communication* 74. 182–194. DOI: *https://doi.org/10.1016/j.langcom.2020.06.009*

Brentari, Diane. 1998. *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press. DOI: *https://doi.org/10.7551/mitpress/5644.001.0001*

Burling, Robbins. 1999. Motivation, conventionalization, and arbitrariness. In Barbara King (ed.), *The origins of language: What nonhuman primates can tell us*, 307–350. Santa Fe: School of American Research Press.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733. DOI: *https://doi.org/10.1353/lan.2006.0186*

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9780511750526*

Bybee, Joan & Paul J. Hopper. 2001. Introduction to frequency and the emergence of linguistic structure. In Joan Bybee & Paul J. Hopper (eds.), *Typological Studies in Language*, Vol. 45, 1. Amsterdam: John Benjamins Publishing Company. DOI: *https://doi.org/10.1075/tsl.45.01byb*

Cormier, Kearsy, David Quinto-Pozos, Zed Sevcikova & Adam Schembri. 2012. Lexicalisation and de-lexicalisation processes in sign languages: Comparing depicting constructions and viewpoint gestures. *Language & Communication* 32(4). 329–348. DOI: *https://doi.org/10.1016/j.langcom.2012.09.004*

Crasborn, Onno, Els van der Kooij, Dafydd Waters, Benice Woll & Johanna Mesch. 2008. Frequency distribution and spreading behavior of different types of mouth actions in three sign languages. *Sign Language & Linguistics* 11(1). 45–67. DOI: *https://doi.org/10.1075/sl*

Crasborn, Onno & Han Sloetjes. 2008. Enhanced ELAN functionality for sign language corpora. *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, 39–43.

Crasborn, Onno, Inge Zwitserlood, Els van der Kooij & Anique Schüller. 2018. Global Signbank manual. DOI: *https://doi.org/10.13140/rg.2.2.36603.82729*

Crasborn, Onno, Richard Bank, Inge Zwiterlood, Els van der Kooij, Ellen Ormel, Johan Ros, Anique Schüller, Anne de Meijer, Merel van Zuilen, Ellen Nauta, Frouke van Winsum, & Max Vonk. 2019. *NGT dataset in Global Signbank*. Nijmegen: Radboud University, Center for Language Studies. Retrieved from DOI: *https://doi.org/10.13140/RG.2.1.2839.1446*

Currie, Anne-Marie, Richard Meier & Keith Walters. 2002. A cross-linguistic examination of the lexicon of four signed languages. In Richard Meier, Kearsy Cormier & David Quinto-Pozos (eds.), *Modality and structure in signed and spoken languages*, 1st ed., 224–236. Cambridge: Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9780511486777.011*

de Vos, Connie. 2011. Kata Kolok color terms and the emergence of lexical signs in rural signing communities. *The Senses and Society* 6(1). 68–76. DOI: *https://doi.org/10.2752/174589311X12893982233795*

de Vos, Connie. 2012. *Sign-spatiality in Kata Kolok: How a village sign language of Bali inscribes its signing space*. Nijmegen: Radboud University dissertation. DOI: *https://doi.org/10.1075/sll.16.2.08vos*

de Vos, Connie & Roland Pfau. 2015. Sign language typology: The contribution of rural sign languages. *Annual Review of Linguistics* 1(1). 265–288. DOI: *https://doi.org/10.1146/annurev-linguist-030514-124958*

Dingemanse, Mark, Damián E. Blasi, Gary Lupyan, Morten H. Christiansen & Padraic Monaghan. 2015. Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences* 19(10). 603–615. DOI: *https://doi.org/10.1016/j.tics.2015.07.013*

Ebling, Sarah, Reiner Konrad, Penny Boyes Braem & Gabriele Langer. 2015. Factors to consider when making lexical comparisons of sign languages: Notes from an ongoing comparison of German Sign Language and Swiss German Sign Language. *Sign Language Studies* 16(1). 30–56. DOI: *https://doi.org/10.1353/sls.2015.0024*

ELAN [Computer software]. 2020. (Version 5.9). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from *https://archive.mpi.nl/tla/elan.*

Emmorey, Karen. 2014. Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B* 369(1651). 20130301. DOI: *https://doi.org/10.1098/rstb.2013.0301*

Fenlon, Jordan, Adam Schembri, Ramas Rentelis & Kearsy Cormier. 2013. Variation in handshape and orientation in British Sign Language: The case of the '1' hand configuration. *Language & Communication* 33(1). 69–91. DOI: *https://doi.org/10.1016/j.langcom.2012.09.001*

Ferrara, Lindsay & Rolf Piene Halvorsen. 2017. Depicting and describing meanings with iconic signs in Norwegian Sign Language. *Gesture* 16(3). 371–395. DOI: *https://doi.org/10.1075/gest.00001.fer*

Friedman, Thomas B., Yong Liang, James L. Weber, John T. Hinnant, Thomas D. Barber, Sunaryana Winata, I. Nyoman Arhya & James H. Asher. 1995. A gene for congenital, recessive deafness DFNB3 maps to the pericentromeric region of chromosome 17. *Nature Genetics* 9(1). 86–91. DOI: *https://doi.org/10.1038/ng0195-86*

Frisch, Stefan. 1996. *Similarity and frequency in phonology*. Evanston, IL: Northwestern University dissertation.

Hartzell, Ethan, Rabia Ergin, Leyla Kursat & Ray Jackendoff. 2019, September. Lexical variation in Central Taurus Sign Language. *Presented at the 13th Theoretical Issues in Sign Language Research Conference*, Hamburg, Germany.

Horton, Laura. 2018. *Conventionalization of shared homesign systems in Guatemala: Social, lexical, and morphophonological dimensions*. Chicago: University of Chicago dissertation. DOI: *https://doi.org/10.6082/uchicago.1321*

Horton, Laura & Jason Riggle. 2019. Converging signs: Quantifying similarity and difference in emerging sign systems. *Poster presented at the Theoretical Issues in Sign Language Research (TISLR13)*, Hamburg, Germany.

Hou, Lynn. 2016. *'Making hands': Family sign languages in the San Juan Quiahije community*. Texas: UT Austin dissertation. Retrieved from *http://hdl.handle.net/2152/46879*

Hou, Lynn. 2018. Iconic patterns in San Juan Quiahije Chatino Sign Language. *Sign Language Studies* 18(4). 570–611. DOI: *https://doi.org/10.1353/sls.2018.0017*

Hou, Lynn & Annelies Kusters. 2020. Linguistic ethnography of signed languages. In Karin Tusting (ed.), *The Routledge handbook of linguistic ethnography*, 340–355. London: Routledge, Taylor & Francis Group. DOI: *https://doi.org/10.4324/9781315675824-25*

Hwang, So-One, Nozomi Tomita, Hope Morgan, Rabia Ergin, Deniz İlkbaşaran, Sharon Seegers, Ryan Lepic & Carol Padden. 2017. Of the body and the hands: patterned iconicity for semantic categories. *Language and Cognition* 9(4). 573–602. DOI: *https://doi.org/10.1017/langcog.2016.28*

Israel, Assaf. 2009. *Sublexical variation in three sign languages*. Haifa: University of Haifa dissertation.

Israel, Assaf & Wendy Sandler. 2009. Phonological category resolution: A study of handshapes in younger and older sign languages. *Cadernos de Saude* 2. 13–28.

Johnston, Trevor, Donovan Cresdee, Adam Schembri & Bencie Woll. 2015. Finish variation and grammaticalization in a signed language: How far down this well-trodden pathway is Auslan (Australian Sign Language)? *Language Variation and Change* 27(1). 117–155. DOI: *https://doi.org/10.1017/S0954394514000209*

Kisch, Shifra. 2008. "Deaf Discourse": The social construction of deafness in a Bedouin community. *Medical Anthropology* 27(3). 283–313. DOI: *https://doi.org/10.1080/01459740802222807*

Kisch, Shifra. 2012. Demarcating generations of signers in the dynamic sociolinguistic landscape of a shared sign-language: The case of the Al-Sayyid Bedouin. In Ulrike Zeshan & Connie de Vos (eds.), *Sign Languages in Village Communities: Anthropological and Linguistic Insights*, 1st ed. 4. 87–126. Berlin: De Gruyter Mouton. DOI: *https://doi.org/10.1515/9781614511496.87*

Kusters, Annelies, Mara Green, Erin Moriarty & Kristin Snoddon. 2020. *Sign language ideologies: Practices and politics. Sign Language Ideologies in Practice*, 3–22. De Gruyter Mouton. DOI: *https://doi.org/10.1515/9781501510090-001*

Kusters, Annelies & Sujit Sahasrabudhe. 2018. Language ideologies on the difference between gesture and sign. *Language & Communication* 60. 44–63. DOI: *https://doi.org/10.1016/j.langcom.2018.01.008*

Lepic, Ryan. 2019. A usage-based alternative to "lexicalization" in sign language linguistics. *Glossa: A Journal of General Linguistics* 4(1). DOI: *https://doi.org/10.5334/gjgl.840*

Lucas, Ceil, Robert Bayley, Mary Rose & Alyssa Wulf. 2002. Location variation in American Sign Language. *Sign Language Studies* 2(4). 407–440. DOI: *https://doi.org/10.1353/sls.2002.0020*

Lutzenberger, Hannah. in press. Threat or natural fluctuation? Revisiting language vitality of Kata Kolok, the sign language of a village in Bali. In *State of the art of indigenous languages in research: A collection of selected research papers*, 12. Paris: UNESCO.

Lutzenberger, Hannah. 2018. Manual and nonmanual features of name signs in Kata Kolok and Sign Language of the Netherlands. *Sign Language Studies* 18(4). 546–569. DOI: *https://doi.org/10.1353/sls.2018.0016*

Lutzenberger, Hannah. 2020. *Kata Kolok Dataset*. Radboud University Nijmegen. Retrieved from *https://signbank.science.ru.nl/datasets/Kata%20Kolok.*

Lutzenberger, Hannah, Katie Mudd & Ni Made Sumarni. 2018. *Variation Project [Dataset]*. Nijmegen, The Netherlands: The Language Archive. Retrieved from *https://hdl.handle.net/1839/e3df69b8-91f8-4245-96b6-6728d4cf4ad6.*

Marsaja, I. Gede. 2008. *Desa kolok: a deaf village and its sign language in Bali, Indonesia*. Nijmegen: Ishara Press.

McKee, David & Graeme Kennedy. 2000. A lexical comparison of signs from American, Australian, British, and New Zealand sign languages. In Karen Emmorey & Harlan Lane (eds.), *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima*, 1st ed., 49–76. Mahwah, NJ: Lawrence Erlbaum Associates. DOI: *https://doi.org/10.4324/9781410604972*

Meir, Irit, Assaf Israel, Wendy Sandler, Carol Padden & Mark Aronoff. 2012. The influence of community on language structure: Evidence from two young sign languages. *Linguistic Variation* 12(2). 247–291. DOI: *https://doi.org/10.1075/lv.12.2.04mei*

Meir, Irit & Wendy Sandler. 2019. Variation and conventionalization in language emergence: The case of two young sign language of Israel. In Edit Doron, Malka Rappaport Hovav, Yael Reshef & Moshe Taube (eds.), *Linguistik Aktuell/Linguistics Today* 256. 337–363. Amsterdam: John Benjamins Publishing Company. DOI: *https://doi.org/10.1075/la.256.13mei*

Meir, Irit, Wendy Sandler, Carol Padden & Mark Aronoff. 2010. Emerging sign languages. In Marc Marschak & Patricia Elizabeth Spencer (eds.), *The Oxford Handbook of Deaf Studies, Language, and Education*, Vol. 2. Oxford University Press. DOI: *https://doi.org/10.1093/oxfordhb/9780195390032.013.0018*

Morgan, Hope. 2015. When does a word emerge? Lexical and phonological variation in a young sign language. *San Diego Linguistic Papers* 5. 2–17. Retrieved from *https://escholarship.org/uc/item/0dg64753.*

Morgan, Hope. 2017. *The phonology of Kenyan Sign Language (southwestern dialect)*. UC San Diego dissertation. Retrieved from *https://escholarship.org/uc/item/9bp3h8t4.*

Moriarty, Erin. 2020. "Sign to me, not the children": Ideologies of language contamination at a deaf tourist site in Bali. *Language & Communication* 74. 195–203. DOI: *https://doi.org/10.1016/j.langcom.2020.06.002*

Moriarty Harrelson, Erin. 2017. Authenticating ownership: Claims to deaf ontologies in the Global South. In Annelies Kusters, Maartje De Meulder & Dai O'Brien (eds.), *Innovations in Deaf Studies*, 361–384. Oxford: Oxford University Press.

Moriarty Harrelson, Erin. 2019. Deaf people with "no language": Mobility and flexible accumulation in languaging practices of deaf people in Cambodia. *Applied Linguistics Review* 10(1). 55–72. DOI: *https://doi.org/10.1515/applirev-2017-0081*

Mudd, Katie, Connie de Vos & Bart de Boer. 2020. A preliminary study of the iconic preferences of Kata Kolok signers. *Presented at the Cognitive and Cultural Influences on Language Emergence (CCILE)*, Istanbul, Turkey [conference cancelled due to COVID-19].

Mudd, Katie, Hannah Lutzenberger, Connie de Vos, Paula Fikkert, Onno Crasborn & Bart de Boer. 2020. The effect of sociolinguistic factors on variation in the Kata Kolok lexicon. *Asia-Pacific Language Variation* 6(1). 53–88. DOI: *https://doi.org/10.1075/aplv.19009.mud*

Neveu, Grace Kathleen. 2019. *Lexical conventionalization and the emergence of grammatical devices in a second generation homesign system in Peru.* Austin: University of Texas at Austin dissertation. DOI: *https://doi.org/10.26153/tsw/3230*

Nyst, Victoria. 2007. *A descriptive analysis of Adamorobe Sign Language (Ghana).* Utrecht: LOT dissertation.

Occhino, Corrine. 2017. An introduction to Embodied Cognitive Phonology: Claw-5 handshape distribution in ASL and Libras. *Complutense Journal of English Studies* 25. 69–103. DOI: *https://doi.org/10.5209/CJES.57198*

Occhino, Corrine, Benjamin Anible, Erin Wilkinson & Jill P. Morford. 2017. Iconicity is in the eye of the beholder: How language experience affects perceived iconicity. *Gesture* 16(1). 100–126. DOI: *https://doi.org/10.1075/gest.16.1.04occ*

Omardeen, Rehana. 2018. *The contribution of lexical overlap to perceived iconicity in foreign signs.* Nijmegen, The Netherlands: Radboud University dissertation.

Padden, Carol A., Irit Meir, So-One Hwang, Ryan Lepic, Sharon Seegers & Tory Sampson. 2013. Patterned iconicity in sign language lexicons. *Gesture* 13(3). 287–308. DOI: *https://doi.org/10.1075/gest.13.3.03pad*

Padden, Carol A., So-One Hwang, Ryan Lepic & Sharon Seegers. 2015. Tools for language: Patterned iconicity in sign language nouns and verbs. *Topics in Cognitive Science* 7(1). 81–94. DOI: *https://doi.org/10.1111/tops.12121*

Parks, Jason. 2011. *Sign language word list comparisons: toward a replicable coding and scoring methodology.* Grand Forks, North Dakota: University of North Dakota dissertation.

Perniss, Pamela, Robin L. Thompson & Gabriella Vigliocco. 2010. Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology* 1. 1–15. DOI: *https://doi.org/10.3389/fpsyg.2010.00227*

Perniss, Pamela & Ulrike Zeshan. 2008. Possessive and existential constructions in Kata Kolok. In Pamela Perniss & Ulrike Zeshan (eds.), *Possessive and existential constructions in sign languages*, 125–150. Nijmegen: Ishara Press.

Pfau, Roland. 2015. The grammaticalization of headshakes: From head movement to negative head. In Andrew D. M. Smith, Graeme Trousdale & Richard Waltereit (eds.), *Studies in Language Companion Series* 166. 9–50. Amsterdam: John Benjamins Publishing Company. DOI: *https://doi.org/10.1075/slcs.166.02pfa*

Pfau, Roland & Markus Steinbach. 2011. Grammaticalization in sign languages. In *The Oxford Handbook of Grammaticalization*, 1–15. Oxford: Oxford University Press. DOI: *https://doi.org/10.1093/oxfordhb/9780199586783.013.0056*

Reed, Lauren. 2019. *Sign languages of western highlands, Papua New Guinea, and their challenges for sign language typology.* Canberra: The Australian National University dissertation. DOI: *https://doi.org/10.25911/5d69027b5d5f8*

Reed, Lauren & Alan Rumsey. 2019. Kailge Sign Language: A network-based sign language and its significance for sign language typology. *Presented at the Theoretical Issues in Sign Language Research (TISLR13), University of Hamburg.*

Richie, Russell, Charles Yang & Marie Coppola. 2014. Modeling the emergence of lexicons in homesign systems. *Topics in Cognitive Science* 6(1). 183–195. DOI: *https://doi.org/10.1111/tops.12076*

Safar, Josefina. 2020. *A comparative study of Yucatec Maya Sign Languages.* Stockholm, Sweden: Stockholm University dissertation. Retrieved from *http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-185092.*

Sandler, Wendy, Mark Aronoff, Irit Meir & Carol Padden. 2011. The gradual emergence of phonological form in a new language. *Natural Language & Linguistic Theory* 29(2). 503–543. DOI: *https://doi.org/10.1007/s11049-011-9128-2*

Sasaki, Daisuke. 2007. Comparing the lexicon of Japanese Sign Language and Taiwan Sign Language: A preliminary study focusing on the difference in the handshape parameter. In David Quinto-Pozos (ed.), *Sign Languages in Contact* 13. 123–152. Washington, DC: Gallaudet University Press.

Schembri, Adam. 2010. Documenting sign languages. In Peter K. Austin (ed.), *Language Documentation and Description 7: Lectures in Language Documentation and Description*, 105–143. London: School of Oriental and African Studies. Retrieved from *http://www.elpublishing.org/PID/083.*

Schembri, Adam, David McKee, Rachel McKee, Sara Pivac, Trevor Johnston & Della Goswell. 2009. Phonological variation and change in Australian and New Zealand Sign Languages: The location variable. *Language Variation and Change* 21(2). 193–231. DOI: *https://doi.org/10.1017/S0954394509990081*

Schembri, Adam, Jordan Fenlon, Kearsy Cormier & Trevor Johnston. 2018. Sociolinguistic typology and sign languages. *Frontiers in Psychology* 9. 200. DOI: *https://doi.org/10.3389/fpsyg.2018.00200*

Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment.* Oxford: Oxford University Press. Retrieved from *https://hdl.handle.net/11245/1.404611.*

Schuit, Joke. 2014. *Sign of the arctic typological aspects of Inuit Sign Language*. Ede: University of Amsterdam dissertation. DOI: *https://doi.org/10.1075/sll.17.2.10sch*

Senghas, Ann. 2003. Intergenerational influence and ontogenetic development in the emergence of spatial grammar in Nicaraguan Sign Language. *Cognitive Development* 18(4). 511–531. DOI: *https://doi.org/10.1016/j.cogdev.2003.09.006*

Senghas, Ann & Marie Coppola. 2011. Getting to the point: How a simple gesture became a linguistic element in Nicaraguan signing. In Gaurav Mathur & Donna Jo Napoli (eds.), *Deaf around the world: The impact of language.* Oxford: Oxford University Press.

Siu, Wai Yan Rebecca. 2016. Location variation in Hong Kong Sign Language (HKSL). *Asia-Pacific Language Variation* 2(1). 4–47. DOI: *https://doi.org/10.1075/aplv.2.1.01siu*

Stamp, Rose, Adam Schembri, Jordan Fenlon, Ramas Rentelis, Bencie Woll & Kearsy Cormier. 2014. Lexical variation and change in British Sign Language. *PLOS ONE* 9(4). e94053. DOI: *https://doi.org/10.1371/journal.pone.0094053*

Stokoe, William C. 1960. *Sign language structure: An outline of the visual communication systems of the American Deaf* 8. Silver Spring, MD: Linstok Press.

Su, Shiou-fen & James H. Y. Tai. 2009. Lexical comparison of signs from Taiwan, Chinese, Japanese, and American Sign Languages: Taking iconicity into account. *Taiwan Sign Language and Beyond*, 149–176.

Taub, Sarah F. 2001. *Language from the body: Iconicity and metaphor in American Sign Language* 1st ed. Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9780511509629*

Thompson, Bill, Limor Raviv & Simon Kirby. 2019. Complexity can be maintained in small populations: a model of lexical variability in emerging sign languages. *Talk presented at the Interaction and the Evolution of Linguistic Complexity Workshop*, Edinburgh.

Tkachman, Oksana & Carla L. Hudson Kam. 2020. Measuring lexical and structural conventionalization in young sign languages. *Sign Language & Linguistics* 23(1–2). 208–232. DOI: *https://doi.org/10.1075/sll.00049.tka*

van der Hulst, Harry & Els van der Kooij. 2021. Sign language phonology – Theoretical perspectives. In Josep Quer, Roland Pfau & Annika Herrmann (eds.), *The routledge handbook of theoretical and experimental sign language research.* Abingdon, Oxon: Routledge.

van der Kooij, Els. 2002. *Phonological categories in sign language of the Netherlands: the role of phonetic implementation and iconicity*. Utrecht: LOT dissertation.

van der Kooij, Els & Inge Zwitserlood. submitted. Strategies for new word formation in NGT: a case for simultaneous morphology. *Sign Language & Linguistics*.

Washabaugh, William. 1986. *Five fingers for survival*. Ann Arbor, MI: Karoma Publishers, Inc.

Wilcox, Sherman. 2004. Gesture and language: Cross-linguistic and historical data from signed languages. *Gesture* 4(1). 43–73. DOI: *https://doi.org/10.1075/gest.4.1.04wil*

Woodward, James. 1991. Sign language varieties in Costa Rica. *Sign Language Studies* 1073(1). 329–345. DOI: *https://doi.org/10.1353/sls.1991.0022*

Woodward, James. 1993. Lexical evidence for the existence of South Asian and East Asian sign language families. *Journal of Asian Pacific Communication* 4(2). 91–106.

Xu, Wang. 2006. *A comparison of Chinese and Taiwan Sign Languages: Towards a new model for sign language comparison*. Ohio State University dissertation.

Yu, Shi, Carlo Geraci & Natasha Abner. 2018. Sign languages and the online world online dictionaries & lexicostatistics. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* 6.