

CONNIE DE VOS

# Sampling Shared Sign Languages

## Abstract

This article addresses some of the theoretical questions, ethical considerations, and methodological decisions that guided the creation of the Kata Kolok corpus as well as the Kata Kolok child signing corpus. This discussion is relevant to the formation of prospective sign corpora that aim to portray the various sociolinguistic landscapes in which sign languages, whether rural or urban, emerge and evolve.

WHETHER STATED or implied, the emergence of sign languages is often assumed to be directly linked to the establishment of formal deaf education and other cultural environments in which deaf individuals congregate. For the past ten years, more attention has been given to the emergence of signing varieties in both rural and urban areas that demonstrate a high incidence of deafness and where both deaf and hearing community members form social networks that use visual-gestural forms of communication (Nyst 2012). This latter sociolinguistic category of sign language, as used by both deaf and hearing community members, is known as a *shared sign language*. Shared signing communities vary with respect to social factors such as the causes and incidence of deafness, community size, the ratio of deaf and hearing signers, time depth, and the sociocultural construction of deafness (Kisch 2008; Kusters 2010). The linguistic and anthropological documentation and descriptions of shared signing communities are still in their initial stages (Zeshan and de Vos 2012). Nevertheless, initial analyses of these signing varieties suggest that, as

---

Connie de Vos is a research staff member of the Language and Cognition department at the Max Planck Institute for Psycholinguistics in Nijmegen.

linguistic isolates, they contribute considerably to our understanding of typological variation among sign languages (see de Vos and Pfau 2015 for an overview).

Comparative studies have also reported a number of common features among shared sign languages that have not been attested in previously documented urban sign languages. Such characteristics are mostly related to the way in which signing space is inscribed with conventional meaning and include a significant enlargement of the articulatory signing space (Kendon 1980; Nyst 2007; Marsaja 2008; de Vos, *Sign Spatiality in Kata Kolok*, 2012), the canonical use of geographical pointing signs for third-person reference (as opposed to pronominal pointing forms) (Washabaugh 1986, 36), the existence of a celestial timeline (Nyst 2007; de Vos 2012, *Sign Spatiality in Kata Kolok*; Le Guen 2012), and, most prominently, a significantly reduced use (and even the absence) of spatial verb agreement (Sandler et al. 2005; Marsaja 2008, 162; Nyst 2007; Schuit, Baker, and Pfau 2011; de Vos 2012, *Sign Spatiality in Kata Kolok*).

Perhaps more controversial is the potential impact of the large proportion of hearing signers on these sign languages. On the lexical level, limited conceptual overlap has been reported between *Kata Kolok* and its surrounding spoken languages (de Vos 2011). Similarly, spoken Balinese and *Kata Kolok* vary on core typological features such as constituent order and verb morphology (*ibid.*). With respect to Adamorobe Sign Language, however, its large number of L2 hearing signers has led to the use of phonologically lax forms Nyst (2007). In this language, *mouthings* linked to spoken Akan are also attested, whereas such contact-induced phenomena are virtually absent in *Kata Kolok* discourse.

A number of studies have hypothesized that shared sign languages exhibit these peculiar structures because of their unusual social settings. That is to say, the social dynamics among shared signing communities, such as comparatively limited time depth (Sandler et al. 2005), dense social networks (Washabaugh, Woodward, and DeSantis 1978; de Vos 2012, *Sign Spatiality in Kata Kolok*), and a large portion of second-language users (Nyst 2012) may underlie the processes that have led to these structural commonalities and differences. If so, synchronic variation in these communities may reflect such processes.

Therefore, this article argues that, to test these hypotheses, corpora of shared sign languages should be sampled strategically to reflect the linguistic variation attested in them. As such, this work not only follows in the footsteps of other recently created sign language corpora but also considers aspects of corpus creation that may be uniquely associated with shared signing communities (e.g., Crasborn, Zwitserlood and Ros 2008; Johnston 2010).

This article focuses on a rural signing variety that is indigenous to a village community in Bali: Kata Kolok. It explains how the Kata Kolok corpus and the Kata Kolok child signing corpus were constructed to reflect the history and constitution of this shared signing community. Specifically, data collection captured not only the various social contexts in which the language is used but also subsequent generations of signers, the various levels of proficiency of (hearing) signers, and longitudinal child signing data, among other things. This documentation effort has captured the linguistic variation that may have led (and continues to lead) to diachronic change in the community. As a reflection of this particular signing community, and with further investment, these corpora could help us assess the viability of various hypotheses regarding the relationship between the linguistic structures of sign languages and the dynamics of their signing communities. In addition, this article touches on the ethical considerations relevant to these atypical deaf communities and the community-centered documentation strategies they call for.

### Kata Kolok: A Shared Sign Language of Indonesia

Kata Kolok is a sign language used by the deaf and hearing inhabitants of a farming village of north Bali, in the region of Buleleng (Marsaja 2008). Independently of the signing varieties used in other parts of Bali and Indonesia (Marsaja 2008), Kata Kolok emerged in response to a sudden increase in hereditary deafness more than five generations ago (de Vos 2012, Sign Spatiality in Kata Kolok). The gene causing the high incidence of deafness in the community is recessive and results in nonsyndromal, sensorineural hearing loss due to shortened hair cells in the cochlea. While 2.2% of the villagers are congenitally deaf, 17.6% of the hearing villagers also carry the “deaf” version of the

gene (Winata et al. 1995). The signing community currently consists of forty-seven deaf individuals, as well as approximately 1,200 hearing signers who use Kata Kolok with varying degrees of fluency (Marsaja 2008). For the reasons stated earlier, the Balinese refer to Bengkala as *Desa Kolok*, which is Balinese for “deaf village,” and its sign language as *Kata Kolok* (“deaf talk”).

*Kata Kolok* is manifested in all of the major facets of village life (e.g., when deaf and hearing clan members gather to prepare *lawar* [a communal meal of chopped meat with spices], or when friends chat over coffee at food stands throughout the village or casually get together in the village compounds shared by deaf and hearing families). *Kata Kolok* has also been observed during elaborate ceremonies in which a *pandetta* (Hindu priest) becomes possessed by one of the deaf ancestors; it is used by the deaf and hearing women who prepare the offerings for these ceremonies. *Kata Kolok* also features in professional domains (e.g., when the village nurse tends to deaf villagers; when individuals work together to maintain the village irrigation system or build roads; or when bargaining over the price of cattle). In addition, since the initiation of inclusive deaf education in 2007, *Kata Kolok* has been used as a language of instruction by the deaf unit of the village elementary school (Kortschak 2010, 76–89; de Vos and Palfreyman 2012). As such, it is one of the very few rural signing varieties that are used in education (cf. Panda 2012 for Alipur Sign Language). These observations suggest that *Kata Kolok* is a fully functional language at present.

Ongoing research on *Kata Kolok* has already revealed numerous peculiarities in its linguistic structures, and the number of publications on *Kata Kolok* is growing (cf. Marsaja 2008; Perniss and Zeshan 2008; de Vos 2011, 2012 [*Kata Kolok: An Updated Sociolinguistic Profile; Sign Spatiality in Kata Kolok; Kata Kolok Perfective in Child Signing*], 2014, 2015). While these studies have identified only a few instances of what is remarkable about the language, they clearly indicate that *Kata Kolok* is expected to contribute considerably to our understanding of the cross-linguistic diversity among visual-gestural languages. However, under the increasing influence of the Indonesian signing varieties used in other parts of Bali and Indonesia and because

of socioeconomic changes within the community, Kata Kolok is already endangered, according to the adapted version of UNESCO's Endangered Languages Survey (Zeshan et al. 2011; de Vos 2012, *Kata Kolok: An Updated Sociolinguistic Profile*). For this reason, I have documented the language by creating two corpora that reflect the characteristics of this shared signing community: the Kata Kolok corpus and the Kata Kolok child signing corpus. Both corpora are fully archived and are being expanded by the collection of additional data and transcriptions.

#### Stimulus-Based Linguistic Elicitation: Enabling Typological Comparison

Sign linguists have traditionally chosen to focus on deaf native signers as their main informants, and for good reason. Most importantly, age of acquisition is known to influence signing proficiency and processing (see, for instance, Lillo-Martin 2000; MacSweeney et al. 2008). Additionally, in Kata Kolok a different register appears to be used among deaf (and a few very fluent hearing) signers, as opposed to the signing used when communicating with hearing villagers (Marsaja 2008, 78). Given Marsaja's observation, this documentation project focused on capturing the signing used by deaf Kata Kolok signers among themselves. This initial step in the description of the language also ensures optimal comparability to sign language structures as they are described for urban sign languages, as these studies have often elicited data from fluent, deaf, native signers. Comparisons reveal genuine cross-linguistic variation since structural differences are not attributable to the hearing status or the fluency of signers. In order to facilitate such comparisons, the Kata Kolok corpus includes a section based on stimulus materials that have been used to elicit particular constructions in other spoken and signed languages. Table 1 presents an overview of this branch of the corpus. Importantly, these elicited data sets have already proved vital in studying phenomena that occur infrequently in spontaneous language use. These structures include, for instance, pointing signs for color description (Majid and Levinson 2007). All of the stimulus materials can be obtained by registering with the L&C Field Manuals and Stimulus Materials Web site (Levinson and Majid 2010).

TABLE 1. Overview of the Data Set Elicited by Standardized Stimulus Materials.

Subcorpora	Corpus Branch	Content	Quantity
Stimulus-Based, Elicited Signing	language of perception	Majid and Levinson 2007	13 deaf signers
	reciprocals	Evans et al. 2004	4 deaf signers
	space and number	Özyürek, Zwitserlood, and Perniss 2010	seven pairs of deaf signers
	Die Sendung mit der Maus	Perniss 2007, 261–68	13 deaf signers
	Canary Row	from the stimulus archive of the Max Planck Institute for Psycholinguistics	12 deaf signers
	man and tree game	adapted from Pederson et al. 1998	five pairs of deaf signers

### Spontaneous Language Use by Deaf Native Signers of Various Generations

Face-to-face interaction is the core ecological niche for language: Interaction is the point of origin for language emergence, acquisition, and evolution (Levinson and Holler 2014). To understand the linguistic capabilities of our species, we must therefore analyze language in such interactive contexts. For these reasons, it is crucial to include spontaneous language in use when documenting a language. Nevertheless, the current sign corpora rarely focus on collecting spontaneous data, and when they do, they often involve monologues or guided discussions on designated topics (cf. Konrad 2012). The collection of spontaneous data for the Kata Kolok corpus aimed to gather sufficient data to demonstrate significant variety in the level of formality and range of topics (Himmelmann 1998). In order to do so, a number of different signers were recorded in culturally informed settings and without any assigned topics of conversation.

Spontaneous data were grouped into three participant configurations: group conversations, dyadic conversations, and monologues. Group conversations among families with deaf members were recorded, with (mostly deaf) signers sitting in a semicircle on their

TABLE 2. Overview of the Types of Social Interactions Captured in the Corpus

Subcorpora	Corpus Branch	Quantity
spontaneous signing	monologues	23 narratives by 9 deaf signers
	dialogues	13 dialogues among 18 deaf signers
	multiparty conversations	10 recordings at informal gatherings and religious ceremonies

farmland. These data meet the standards for spontaneous interactional data collection (Enfield et al. 2007). That is, these conversations are among the most informal. Signers are friends or close relatives who interact on a daily basis. The topics discussed range from the price of rice, levels of recent rainfall, and upcoming ceremonies, to recent serious accidents, local politics, or gossip. As explained earlier, the data set is of particular relevance to studies of the interactional foundations of language use and language emergence.

Table 2 presents an overview of the spontaneous data set. All of the forty-seven deaf signers living in Bengkala at the time of the fieldwork appear in these recordings. They include the third, fourth, and fifth biological generations of Kata Kolok signers, and differences between these generations might therefore suggest diachronic change. Accordingly, the diverse recordings include two conversations between signers of generation II, seven between signers of generation IV, and one conversation between signers from generation V, as well as recordings of triadic conversations between members of generations III and IV; III, III, and IV; and III, V and V. Note, however, that these generations have become highly integrated on a social level. Multiple individuals have parents from different generations, and peers from the fourth and fifth generations attend school together. In the case of Al-Sayyed Bedouin Sign Language, Kisch (2012) has therefore argued in favor of social network analysis to chart the various cohorts that may reflect intergenerational processes in that particular shared signing community. Similarly, the dispersal of lexical variation in Ban Khor Sign Language has been linked to the geographical location of two deaf families (Nonaka 2004). Although the metadata of the Kata Kolok corpus include details of kinship relations among participants, additional analysis of social interaction patterns in the village and of

focal communication spots is required to chart the diffusion of forms in the community through social time and space.

### Widespread Bimodal Bilingualism in Hearing Signers

The patterns of language acquisition and use by the members of the hearing community are markedly different from those of this group of sign language users of previously documented sign languages, which is reflected by the unusual constitution of Kata Kolok's signing community. First, the vast majority of Kata Kolok signers (96%) are hearing (Marsaja 2008; de Vos 2012, Sign Spatiality in Kata Kolok). These hearing Kata Kolok signers are bimodal bilinguals; in other words, in addition to Kata Kolok, they also use spoken Balinese on a daily basis. In urban signing communities this group of bimodal bilinguals consists primarily of interpreters, hearing parents of deaf children who learn to sign, and hearing individuals with at least one deaf parent (Codas). Although the total number of hearing signers may vary from one urban signing community to another, the proportion will inevitably be considerably lower than the 96% reported for Kata Kolok. In the case of Sign Language of the Netherlands, for instance, the total number of hearing signers is estimated to be one-third of all sign language users (Crasborn 2001). In the case of Bengkulu, hearing sign language users have been categorized according to their fluency, based on signed interviews by a deaf Kata Kolok signer in 2000 (Marsaja 2008). Deaf native signers constitute 4% of the signing community. The survey also identified a category of native hearing signers who have acquired Kata Kolok as (extended) family members of deaf individuals in one of the village compounds comprising deaf persons. This group of balanced bimodal bilinguals accounts for 6% of the community. The remaining 90% of the signing community consists of fluent (36%) and nonfluent (54%) hearing signers. Table 3 displays absolute numbers and percentages reflecting the patterns of language acquisition in the Kata Kolok community.

For the foregoing reasons, the Kata Kolok corpus also includes a systematically generated collection of data from hearing signers and nonfluent signers (see table 4). This innovation is motivated by both linguistic and ethical arguments. As mentioned earlier, the vast majority of Kata Kolok signers are hearing, and more than half of all



TABLE 3. Patterns of Language Acquisition in the Kata Kolok Community (based on Marsaja 2008).

Category of Signer	Number	Percentage
deaf native signers	47	4
hearing native signers	78	6
hearing fluent signers (nonnative)	449	36
hearing nonfluent signers	681	54
total number of signers	1,255	100

Kata Kolok users are nonfluent in the language. Kata Kolok has thus been in intimate linguistic contact with spoken Balinese from its inception. For this reason, the patterns of signing found among Kata Kolok hearing signers may prove to make a valuable contribution to a future understanding of the structures that have emerged in Kata Kolok. Such phenomena may include the simultaneous production of gestures with speech, Balinese calques reflected in their signing, and second-language learning effects as Kata Kolok is typically acquired later in life.

Moreover, given that more than half of the villagers of Bengkulu use the sign language, hearing villagers are stakeholders in the research process, too. By excluding them from the research process, we would be ignoring 96% of Kata Kolok signers. In a way, these hearing signers share in a deaf village identity along with their deaf relatives, friends, neighbors, and colleagues. Deaf signers of urban sign languages have a sense of ownership of their sign languages and are therefore involved in dictionary projects, for instance. Conversely, in the case of Bengkulu, Kata Kolok may better be viewed as a communication tool shared by both deaf and hearing persons. In support of this view, a number of authors have argued that the vitality of shared sign languages is often linked to the attitudes of hearing signers, who can be regarded as guardians of the language (see, for example, Dikyuva 2012; Lanesman and Meir 2012).

### Sign Bilingualism

On a par with other shared sign languages, Kata Kolok is subject to the increasing influence of the sign language used by the wider

TABLE 4. The Kata Kolok Corpus Systematically Includes Hearing Signers.

Corpus Branch	Content	Quantity
deaf-hearing interaction	multiparty interactions of deaf and hearing family members	12 recording sessions
	dialogues between adult deaf and hearing siblings	3 one-hour recordings
Balinese (with cospeech gesture)	interactions among hearing community members who are bimodal bilingual	3 one-hour recordings

Balinese deaf community<sup>1</sup> (de Vos 2012, *Kata Kolok: An Updated Sociolinguistic Profile*). Since the early 2000s, at least two teenagers have attended a boarding school for deaf students in Jimbaran in southern Bali. Because the village's inclusive school for deaf students provides only a primary education, an increasing number of deaf teenagers have also started attending school in Singaraja since 2010. Subsequently, they have acquired the signing variety used by their deaf peers in this environment. As such, this selected group of up to eight signers has become bilingual between Kata Kolok and the signing variety that is used in school. This type of language contact (i.e., between a rural signing variety and a majority sign language) often leads to the decreased prestige of the minority sign language and to a language shift as a result (Nonaka 2004; Lanesman and Meir 2012). From a linguistic point of view it also leads to a severely understudied domain of unimodal contact between two sign languages that are typologically distinct (cf. Adam 2012), and for this reason, the Kata Kolok corpus also includes a few recordings between such sign-language bilinguals.

### The Kata Kolok Child Signing Corpus

Intriguingly, the patterns of first-language acquisition for Kata Kolok deviate from those of urban sign languages around the world but are optimally comparable to the acquisition of spoken languages (de Vos 2012, *Kata Kolok Perfective in Child Signing*). First, all deaf children in Bengkulu receive linguistic input from birth and thus start language acquisition from this moment. Deaf children in Bengkulu also benefit from a rich linguistic input, as they are surrounded by numerous fluent adult signers. Contrastingly, most deaf children in urban societies

are born to hearing parents who have no knowledge of sign language; consequently, few signers have access to comprehensive sign language input from birth. Approximately 5–10% of deaf children acquire sign language from adults who are themselves native signers (Schein and Delk 1974; Kyle and Woll 1985; Neidle et al. 2000), but the number may be even lower in the case of smaller deaf communities (Costello, Fernández, and Landa 2006). Furthermore, even when urban deaf children receive full linguistic input from birth (i.e., from their deaf signing parents), it is not common for many of these children's relatives also to be native signers. Deaf children who grow up in shared signing communities such as Bengkala thus have a comparatively rich linguistic environment not only with their parents but also with extended family members who are fluent signers.

Moreover, shared signing communities can be characterized as having positive attitudes about deafness and sign language use (Marsaja 2008; Kisch 2008; Kusters 2010). In Bengkala, deaf children are therefore able to use their mother tongue from an early age (e.g., when they buy snacks at one of the village food stands, visit the village nurse, interact with deaf and hearing peers, and begin their primary education). Given the parallels between the first-language acquisition of Kata Kolok and the native acquisition of spoken languages, different developmental trajectories are more easily attributable to differences between both natural language modalities rather than to limitations of the linguistic input the deaf children receive (de Vos 2012, *Kata Kolok Perfective in Child Signing*). Furthermore, given Kata Kolok's unique linguistic characteristics, such data enable comparative acquisition work on typologically distinct sign languages for the first time. At any rate, the youngest generation of signers may embody the locus of linguistic change and are therefore of particular relevance to those who are interested in the emergence and evolution of sign languages (cf. Senghas and Coppola 2001; Meir et al. 2010; Brentari and Coppola 2013).

For the reasons stated earlier, a considerable number of video recordings of Kata Kolok comprise child signing data that have been incorporated into the Kata Kolok child signing corpus. These data have been archived according to three main social settings: at school; during interactions between deaf children and their deaf and hearing peers in informal contexts; and at home (i.e., longitudinal data on

TABLE 5. Kata Kolok by Child Signers.

Child Signing	Content	Quantity
longitudinal data	deaf child of deaf parents 2;0–4;2 and 6;4–6;11; and at 8;4	monthly recordings of at least half an hour each
	deaf child of deaf parents 1;11–3;9 and 6;3–6;11; and at 8;5	monthly recordings of at least half an hour each
	hearing child of deaf parents 0;9–1;5	5 recordings
	hearing child of deaf parents 0;2–1;10	11 recordings
	deaf child of hearing parents (home signer) 5;6–5;11	7 recordings
	deaf child of deaf parents 0;5– . . . (data collection is ongoing)	monthly recordings of at least 4 hours within a single week
	hearing child growing up in deaf family compound 2;1– . . . (data collection is ongoing)	monthly recordings of at least 4 hours within a single week

children and their primary caregivers). Table 5 presents an overview of the longitudinal child signing data.

Between 2007 and 2009 and between 2011 and 2012, two focal deaf children (a boy and a girl) were followed. Each child had two deaf parents, deaf grandparents, and deaf older siblings. These first two children were recorded monthly between 2;3 years and 4;11 years of age as well as between 6;3 and 6;11 years of age. They were also recorded at 8;4 and 8;5 years of age. Each recording session lasted an average of half an hour, and occasionally recordings were missed due to technical difficulties at the site. In addition to these two focal children, the Kata Kolok child signing corpus also includes several recordings of hearing infants who were acquiring Kata Kolok from their parents. During data collection, I also discovered a deaf child of hearing parents living on the outskirts of the community; the little girl was kept mostly in the house. Since her parents were not originally from Bengkulu, they did not know how to sign; thus their daughter was considered a home signer (Goldin-Meadow 2003). The girl was invited to join the deaf unit of the village elementary school at the age of 5;5, which was her first opportunity to interact with deaf peers. Her linguistic development was tracked for this initial year.

Between mid-2005 and late-2014 no new deaf infants had been born into the village, but this changed with the marriage of two deaf individuals of the fifth generation, who had a deaf baby girl in late 2014. This child is the first member of the sixth generation of deaf Kata Kolok signers. Monthly video recordings of infant-caregiver interactions began in early 2015, when the child reached 5 months of age. All of the recordings are being conducted by a deaf research assistant who is from the same community and is also a close friend of the family. In addition, she started recording a hearing grandson of deaf grandparents, beginning when the child was 2;2 of age. Inasmuch as this child is growing up in a deaf family compound with numerous deaf peers and adults, he may turn out to be a balanced bimodal bilingual as an adult. Based on past experiences, we are aiming to gather less frequent yet denser data, such that a clear snapshot will be available of what a child's linguistic capacities are at a particular age. Following Stoll and Bickel (2013), current documentation efforts therefore aim to gather—on a monthly basis—four to five hours of data in a single week.

Metadata are crucial to the future functionality of corpora, especially for acquisition corpora, as a child's expression may not always be fully appreciated without complete access to the signing context. There is also a factor of urgency here, as people's memories of discussed events fade. For this reason, the initial activities focused on getting richer metadata, especially regarding the circumstances in which the data were recorded. In addition, the types of social activities shown in these video recordings were identified in the ELAN transcription files and then categorized. In addition, more than twenty-five hours of raw video data on the corpus have now been transcribed, highlighting the two aforementioned deaf children from 2 to 3 years of age. Reliability was ensured by discussing and double-checking ambiguous translations with both the deaf research assistant mentioned earlier and the mothers of both focal children.

### Archiving the Kata Kolok Corpus

The Kata Kolok corpus currently comprises more than eighty-one hours of video footage of 50.5 hours of sessions (some recordings were made with two or more cameras). The Kata Kolok child signing

corpus comprises seventy-five hours of single-camera video recordings. The video data were digitized and later stored at the Language Archive at the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands. This initial step ensured the protection of these unique video materials as this digital archive is backed up at six locations in the Netherlands and Germany (Koenig 2011). Additionally, this system creates stable Handle.Net links to individual video files for the purposes of publication (see, for example, de Vos 2014). The collection of longitudinal child signing data was funded in part by the Hans Rausing Endangered Languages Project and is therefore archived with the Endangered Languages Documentation Programme. The International Institute for Sign Languages and Deaf Studies in Preston, UK, holds copies of all of the video recordings as well.

For each session of Kata Kolok video data a metadata file was produced based on the IMDI format, which is a standardized way of describing multimedia and multimodal language resources in order to make the data searchable once added to a corpus (Broeder and Wittenburg 2006). The anonymized Kata Kolok metadata can be found at the following URL: [http://corpus1.mpi.nl/ds/imdi\\_browser?openpath=MPI576299%23](http://corpus1.mpi.nl/ds/imdi_browser?openpath=MPI576299%23). Each metadata file was enriched by using a sign language-specific profile that was adapted from Crasborn and Hanke (2003). Their profile includes information on a subject's hearing status, the hearing status of close relatives, age of sign language acquisition, use of hearing aids, and level of education. The deaf residents of Bengkala form a homogeneous population in terms of these metadata, as all of them learn to sign from birth, none of them use hearing aids, and only two adults have had more than a few months of formal education. Notably, these metadata include limited details regarding the kinship relations among individuals in the corpus, and they could be significantly improved by additional sections on social networks and village geography (as discussed in the earlier section on spontaneous language use by deaf native signers of various generations).

A selected portion of the data was translated into Indonesian by a research assistant, Ketut Kanta, who was particularly suitable for this work for a number of reasons. He is from Bengkala and is a fluent Kata Kolok signer, highly literate (with a BA in business studies),

and an experienced linguistics fieldwork assistant, having previously worked on Kata Kolok research (see Marsaja 2008). The Indonesian sentences were subsequently translated into English by Febby Meillisa, who worked at the Jakarta Field Station of the Max Plank Institute for Evolutionary Anthropology. In addition to being highly proficient in English, she has the necessary cultural background to make good translations as she lives with her in-laws in Bali. Whenever ambiguities or a lack of clarity occurred in the Indonesian translations, she contacted Ketut Kanta in order to make the correct translation. The translations in Indonesian and English make the corpus accessible to a national and an international audience. However, for proper linguistic analysis, in-depth annotation and coding are required. I produced these detailed sign-by-sign glosses based on the English translations and my knowledge of Kata Kolok. For the current project, I used an annotation protocol that was developed by the Sign Language Typology Group at the Max Planck Institute for Psycholinguistics in Nijmegen (Zeshan 2005). Currently, approximately 4.5 hours of Kata Kolok discourse have been transcribed in detail, as well as more than twenty-five hours of the Kata Kolok child signing corpus.

The data were annotated and coded using ELAN annotation software, which is freely available at <https://tla.mpi.nl/tools/tla-tools/elan/>. Figure 1, which shows a snapshot of this video program, illustrates a section of the Kata Kolok data that have been annotated and coded. The video at the top left is linked to the timeline at the bottom of the image. Several independent tiers provide linguistic annotations. Each signer in the video has five basic tiers. Two tiers are dedicated to manual signs: “Main Gloss” refers to two-handed signs and signs produced by a signer’s dominant hand; the tier labeled “Non-Dominant Hand” indicates signs produced by a signer’s nondominant hand. A third tier (“Non-Manuals”) is used for nonmanual signals, including various facial expressions and body movements. The Kata Kolok corpus has two tiers that provide sentence-level translations, one in Bahasa Indonesia and one in English. The “Comment” tier provides space for remarks about signed content or culturally embedded information.

ELAN enables the researcher to make time-aligned video annotations in multiple tiers, which can be arranged according to the nature of the research questions. The annotations can subsequently be

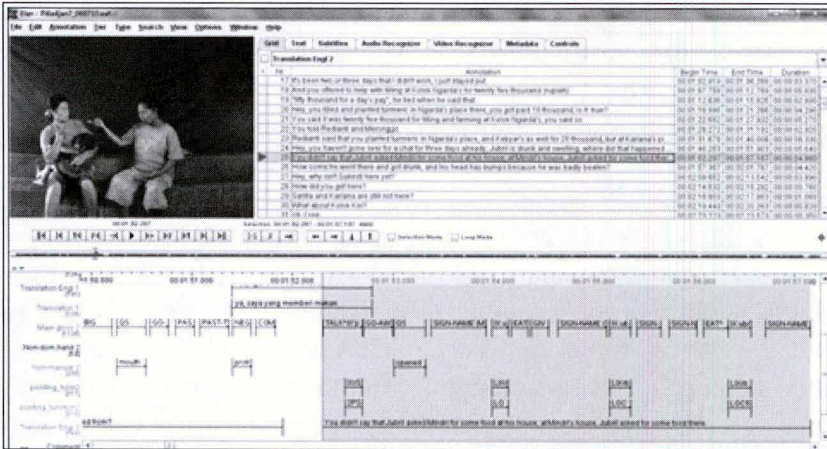


FIGURE 1. ELAN annotation software.

exported to other computer programs for statistical analysis. ELAN also has extensive search functions that can be constrained both by specifying aligned, overlapping, or noncoinciding values on multiple tiers and by providing metadata values that are available in the IMDI files described earlier. By archiving the data in the MPI corpus and by adapting these standardized language archiving tools to our specific purposes, we protect the Kata Kolok corpora and ensure its future compatibility as these formats have been developed to be maintained for the next fifty years.

### Community-Centered Corpus Creation

The methodology discussed in this article is community centered in that it is based on mid- to long-term visits with the deaf and the hearing members of this signing group of people and aims to capture their shared language. From mid-2006 to early 2015, I stayed with them for several months at a time, for a total of eighteen months. Moreover, this article shows how knowledge of the Kata Kolok signing community, as described by Marsaja (2008), has informed the sampling of the Kata Kolok corpus, which contains numerous casual conversations among deaf native signers. In addition, several stimulus materials, which were used to elicit linguistic data from other (sign) languages, were used to elicit Kata Kolok data, too. These methodological decisions



ensure maximal comparability to other (sign) languages and expand the data set for constructions that occur infrequently in spontaneous interactions.

Although the Kata Kolok corpus is comparable in size to the corpora of other sign languages, its composition is different. In contrast to other sign language corpora, which include data from only deaf, native signers, the Kata Kolok corpus also includes data from hearing signers, both fluent and nonfluent (see Konrad 2012). As many as 96% of Kata Kolok signers are in fact bimodal bilingual, and their possible influence can therefore not be disregarded. Additionally, the Kata Kolok child signing corpus is dedicated to first-language acquisition, which appears to be exceptionally rich compared to the context in which most deaf children acquire sign language and is in effect more similar to natural spoken language acquisition. Given Kata Kolok's remarkable linguistic characteristics, the latter corpus allows for typologically informed studies of sign language acquisition for the first time.

I would like to mention a few final points about the future functionality of both corpora. First, the Kata Kolok corpus may be able to serve as a prototype for the documentation of other shared sign languages that display similar sociolinguistic properties. Moreover, some of sampling methods (such as explicitly including nonfluent signers) could also be considered for prospective documentation projects involving other signing varieties, regardless of whether they have emerged in informal or institutional settings. Second, although the data have been processed and archived digitally, along with at least minimal metadata, a large part of the corpus is not yet optimally accessible. Only a small percentage of video data have been fully transcribed, and additional transcription activities are therefore ongoing.

Both the Kata Kolok corpus and the Kata Kolok child signing corpus are searchable with respect to linguistic categories, but researchers from related fields (e.g., social interaction studies, religious studies, deaf studies, studies of the evolution of language) may also find the data to be of interest. In other words, with future investment, these corpora may fulfill their potential as a tool for various research ventures. In sign linguistics, corpus analysis is a crucial tool in identifying both the true linguistic diversity of language in use (e.g., Johnston et al. 2007; de Beuzeville, Johnston, and Schembri 2009) and, as argued

here, the interactional processes that lead to linguistic change within signing communities. At any rate, both the documentation and the description of shared sign languages critically inform (sign) linguistics by identifying, for instance, the domains of sign language typology, cross-modal contact, historical sign linguistics, sign language acquisition, and sign bilingualism.

### Acknowledgments

I would like to thank the deaf and the hearing community members of Bengkulu (Bali), in particular I Ketut Kanta and I Gede Primantara, my research assistants, for their companionship and cooperation during my visits to their village. I would also like to thank Hannah Lutzenberger and Ni Made Dadiastini for their help in reinitiating the longitudinal child signing data collection and our ongoing transcription of the Kata Kolok child signing corpus. The documentation efforts described in this article have been funded by the Max Planck Gesellschaft, the ERC Advanced Grant #269484 INTERACT awarded to Prof. Stephen C. Levinson, and the ELDP Small Grant Longitudinal Documentation of Sign Language Acquisition in a Deaf Village in Bali, awarded to me by the Hans Rausing Endangered Languages Project. Portions of this article have been adapted from de Vos (2012, Sign Spatiality in Kata Kolok).

### Notes

1. By this I mean the Indonesian signing variety that is used in the boarding schools for deaf children in Singaraja and Jimbaran in Bali. At this stage, it is unclear whether this signing variety is in fact part of a single, national Indonesian Sign Language (Palfreyman 2014).

### References

- Adam, R. 2012. Unimodal Bilingualism in the Deaf Community: Contact between Dialects of BSL and ISL in Australia and the United Kingdom. PhD diss., University College London.
- Brentari, D., and M. Coppola. 2013. What Sign Language Creation Teaches Us about Language. *WIREs Cognitive Science* 4: 201–11.
- Broeder, D., and P. Wittenburg. 2006. The IMDI Metadata Framework: Its Current Application and Future Direction. *International Journal of Metadata, Semantics, and Ontologies* 1(2): 119–32.

- Costello, B., J. Fernández, and A. Landa. 2006. The Non(-Existent) Native Signer: Sign Language Research in a Small Deaf Population. Paper presented at the Ninth Theoretical Issues in Sign Language Research conference (TISLR 9), December 6–9, 2006, Florianopolis, Brazil.
- Crasborn, O. 2001. Phonetic Implementation of Phonological Categories in Sign Language of the Netherlands. PhD diss., Utrecht: LOT.
- , and Hanke, T. 2003. Additions to the IMDI Metadata Set for Sign Language Corpora. Radboud University. Retrieved September 17, 2015, from [http://www.ru.nl/publish/pages/522090/signmetadata\\_oct2003.pdf](http://www.ru.nl/publish/pages/522090/signmetadata_oct2003.pdf).
- Crasborn, O., I. Zwitserlood, and J. Ros. 2008. Corpus NGT: An Open-Access Digital Corpus of Movies with Annotations of Sign Language of the Netherlands. Retrieved July 29, 2015, from <http://www.ru.nl/corpusngtuk>.
- De Beuzeville, L., T. Johnston, and A. Schembri. 2009. The Use of Space with Indicating Verbs in Australian Sign Language: A Corpus-Based Investigation. *Sign Language and Linguistics* 12(1): 53–82.
- de Vos, C. 2014. Absolute Spatial Deixis and Proto-Toponyms in Kata Kolok. *NUSA: Linguistic Studies of Languages in and around Indonesia* 56: 3–26.
- . 2012. Kata Kolok: An Updated Sociolinguistic Profile. In *Sign Languages in Village Communities: Anthropological and Linguistic Insights*, ed. U. Zeshan and C. de Vos, 381–86. Berlin: Mouton de Gruyter.
- . 2011. Kata Kolok Color Terms and the Emergence of Lexical Signs in Rural Signing Communities. *Senses and Society* 6(1): 68–76.
- . 2012. The Kata Kolok Perfective in Child Signing: Coordination of Manual and Nonmanual Components. In *Sign Languages in Village Communities: Anthropological and Linguistic Insights*, ed. U. Zeshan and C. de Vos, 127–52. Berlin: Mouton de Gruyter.
- . 2015. The Kata Kolok Pointing System: Morphemization and Syntactic Integration. *Topics in Cognitive Science* 7(1): 150–68.
- . 2012. Sign Spatiality in Kata Kolok: How a Village Sign Language of Bali Inscribe Its Signing Space. PhD diss., Radboud University, Nijmegen.
- , and N. Palfreyman. 2012. Review of *Deaf around the World: The Impact of Language*, ed. G. Mathur and D. J. Napoli. *Journal of Linguistics* 48: 731–35.
- de Vos, C., and R. Pfau. 2015. Sign Language Typology: The Contribution of Rural Sign Languages. *Annual Review of Linguistics* 1: 265–88.
- Dikyuva, H. 2012. Mardin Sign: Language Signing in a “Deaf Family.” In *Sign Languages in Village Communities: Anthropological and Linguistic Insights*, ed. U. Zeshan and C. de Vos, 395–99. Berlin: Mouton de Gruyter.
- Enfield, N. J., S. C. Levinson, J. P. de Ruiter, and T. Stivers. 2007. Building a Corpus of Multimodal Interaction in Your Field Site. In *Field*

- Manual*, vol. 10, ed. A. Majid, 96–99. Nijmegen: Max Planck Institute for Psycholinguistics.
- Evans, N., S. C. Levinson, N. J. Enfield, A. Gaby, and A. Majid. 2004. Reciprocal Constructions and Situation Type. In *Field Manual*, vol. 9, ed. A. Majid, 25–30. Nijmegen: Max Planck Institute for Psycholinguistics.
- Goldin-Meadow, S. 2003. *The Resilience of Language: What Gesture Creation in Deaf Children Can Tell Us about How All Children Learn Language*. New York: Psychology Press.
- Himmelman, N. P. 1998. Documentary and Descriptive Linguistics. *Linguistics* 36(1): 161–96.
- Johnston, T. 2010. From Archive to Corpus: Transcription and Annotation in the Creation of Signed Language Corpora. *International Journal of Corpus Linguistics* 15(1): 106–31.
- , M. Vermeerbergen, A. Schembri, and L. Leeson. 2007. “Real Data Are Messy”: Considering the Cross-Linguistic Analysis of Constituent Ordering in Australian Sign Language (Auslan), Vlaamse Gebarentaal (VGT), and Irish Sign Language (ISL). In *Visible Variation: Comparative Studies on Sign Language Structure*, ed. P. Perniss, R. Pfau, and M. Steinbach, 163–205. Berlin: Mouton de Gruyter.
- Kendon, A. 1980. A Description of a Deaf-Mute Sign Language from the Enga Province of Papua New Guinea with Some Comparative Discussion. Part II: The Semiotic Functioning of Enga Signs. *Semiotica* 32(1/2): 81–117.
- Kisch, S. 2008. “Deaf Discourse”: The Social Construction of Deafness in a Bedouin Community. *Medical Anthropology* 7(3): 283–313.
- . 2012. Demarcating Generations of Signers in the Dynamic Sociolinguistic Landscape of an Emerging Sign Language: The Case of the Al-Sayyid Bedouin. In *Sign Languages in Village Communities*, ed. U. Zeshan and C. de Vos, 87–125. Berlin: Mouton de Gruyter.
- Koenig, A. 2011. *The Language Archive*. Retrieved July 28, 2015, from <https://tla.mpi.nl/>.
- Konrad, R. 2012. *Sign Language Corpora Survey: DGS-Korpus*. Retrieved September 17, 2015, from [http://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt\\_pdf/SL-Corpora-Survey\\_update\\_2012.pdf](http://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/SL-Corpora-Survey_update_2012.pdf).
- Kortschak, I. 2010. Where Everyone Speaks Deaf Talk. *Invisible People: Poverty and Empowerment in Indonesia*. Jakarta: PNPM Mandiri.
- Kusters, A. 2010. Deaf Utopias? Reviewing the Sociocultural Literature on the World’s “Martha’s Vineyard Situations.” *Journal of Deaf Studies and Deaf Education* 15(1): 3–16.
- Kyle, J., and B. Woll. 1985. *Sign Language: The Study of Deaf People and Their Language*. Cambridge: Cambridge University Press.
- Lanesman, S., and S. Meir. 2012. The Survival of Algerian Jewish Sign Language alongside Israeli Sign Language in Israel. In *Endangered Sign*

- Languages in Village Communities: Anthropological and Linguistic Insights*, ed. U. Zeshan and C. de Vos, 153–79. Berlin: Mouton de Gruyter.
- Le Guen, O. 2012. Exploration in the Domain of Time: From Yucatec Maya Time Gestures to Yucatec Maya Sign Language Time Signs. In *Endangered Sign Languages in Village Communities: Anthropological and Linguistic Insights*, ed. U. Zeshan and C. de Vos, 209–49. Berlin: Mouton de Gruyter.
- Levinson, S. C., and J. Holler. 2014. The Origin of Human Multi-Modal Communication. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1651): 2013030.
- Levinson, S. C., and A. Majid. 2010. *L&C Field Manuals and Stimulus Materials*. Retrieved July 28, 2015, from <http://fieldmanuals.mpi.nl/>.
- Lillo-Martin, D. 2000. Early and Late in Language Acquisition: Aspects of the Syntax and Acquisition of Wh-Questions in American Sign Language. In *The Signs of Language Revisited*, ed. K. Emmorey and H. Lane, 401–14. Mahwah, NJ: Erlbaum.
- MacSweeney, M., D. Waters, M. J. Brammer, B. Woll, and U. Goswami. 2008. Phonological Processing in Deaf Signers and the Impact of Age of First Language Acquisition. *NeuroImage* 40(3): 1369–79.
- Majid, A., and S. C. Levinson. 2007. Language of Perception. In *Field Manual*, vol. 10, ed. A. Majid, 22–25. Nijmegen: Max Planck Institute for Psycholinguistics.
- Marsaja, I. G. 2008. *Desa Kolok: A Deaf Village and Its Sign Language in Bali, Indonesia*. Nijmegen: Ishara.
- Meir, I., W. Sandler, C. Padden, and M. Aronoff. 2010. Emerging Sign Languages. *Oxford Handbook of Deaf Studies, Language, and Education*, vol. 2, ed. M. Marschark and P. E. Spencer, 267–80. Oxford: Oxford University Press.
- Neidle, C., J. Kegl, D. MacLaughlin, B. Bahan, and R. G. Lee. 2000. *The Syntax of American Sign Language: Functional Categories and Hierarchical Structure*. Cambridge, MA: MIT Press.
- Nonaka, A. M. 2004. The Forgotten Endangered Languages: Lessons on the Importance of Remembering from Thailand's Ban Khor Sign Language. *Language in Society* 33: 737–67.
- , V. Nyst, and S. Kisch. 2010. The Linguistic Ecology of “Village Sign Languages”: Methodological Pitfalls and Good Practices. Paper presented at the Tenth Theoretical Issues in Sign Language Research conference (TISLR 10), September 30–October 2, 2010, Purdue University, Lafayette.
- Nyst, V. 2007. *A Descriptive Analysis of Adamorobe Sign Language (Ghana)*. Amsterdam: LOT.
- . 2012. Shared Sign Languages. In *The International Handbook of Sign Languages*, ed. M. Pfau, M. Steinbach, and B. Woll, 52–74. Berlin: Mouton de Gruyter.

- Özyürek, A., I. Zwitserlood, and P. Perniss. 2010. Locative Expressions in Signed Languages: A View from Turkish Sign Language (TİD). *Linguistics* 48(5): 1111–45.
- Palfreyman, N. B. 2014. Lexical and Morphosyntactic Variation in Indonesian Sign Language. Ph.D. diss., University of Central Lancashire.
- Panda, S. 2012. Alipur Sign Language: A Sociolinguistic and Cultural Profile. In *Sign Languages in Village Communities: Anthropological and Linguistic Insights*, ed. U. Zeshan and C. de Vos, 353–60. Berlin: Mouton de Gruyter.
- Pederson, E., E. Danziger, D. Wilkins, S. C. Levinson, S. Kita, and G. Senft. 1998. Semantic Typology and Spatial Conceptualization. *Language* 74(3): 557–89.
- Perniss, P. M. 2007. Space and Iconicity in German Sign Language (DGS). PhD diss., Max Planck Institute for Psycholinguistics, Nijmegen.
- , and U. Zeshan. 2008. Possessive and Existential Constructions in Kata Kolok. In *Possessive and Existential Constructions in Sign Languages*, ed. U. Zeshan and P. M. Perniss, 125–50. Sign Language Typology Series. Nijmegen: Ishara.
- Sandler, W., I. Meir, C. A. Padden, and M. Aronoff. 2005. The Emergence of Grammar: Systematic Structure in a New Language. *Proceedings of the National Academy of Sciences* 102(7): 2661–65.
- Schein, J. D., and M. T. Delk. 1974. *The Deaf Population of the United States*. Silver Spring, MD: National Association of the Deaf.
- Schuit, J., A. Baker, and R. Pfau. 2011. Inuit Sign Language: A Contribution to Sign Language Typology. *Linguistics in Amsterdam* 4(1): 1–31.
- Senghas, A., and M. Coppola. 2001. Children Creating Language: How Nicaraguan Sign Language Acquired a Spatial Grammar. *Psychological Science* 12(4): 323–28.
- Stoll, S., and B. Bickel. 2013. Capturing Diversity in Language Acquisition Research. In *Language Typology and Historical Contingency*, ed. B. Bickel, L. A. Grenoble, D. A. Peterson, and A. Timberlake, 195–216. Amsterdam: Benjamins.
- Washabaugh, W. 1986. *Five Fingers for Survival: Sign Language in the Caribbean*. Ann Arbor, MI: Karoma Press.
- Washabaugh, W., J. C. Woodward, and S. DeSantis. 1978. Providence Island Sign: A Context-Dependent Language. *Anthropological Linguistics* 20(3): 96–107.
- Winata, S., I. N. Arhya, S. Moeljopawiro, J. T. Hinnant, Y. Liang, T. B. Friedman, and J. Asher. 1995. Congenital Non-Syndromal Autosomal Recessive Deafness in Bengkulu, an Isolated Balinese Village. *Journal of Medical Genetics* 32(5): 336–43.
- Zeshan, U. 2005. ELAN Transcription Conventions. Unpublished manuscript, Sign Language Typology Research Group, Max Planck Institute for Psycholinguistics.

- , et al. 2011. Adapted Survey: Linguistic Vitality and Diversity of Sign Languages. Questionnaire adapted from UNESCO's Endangered Languages Survey. International Institute for Sign Languages and Deaf Studies, University of Central Lancashire. Retrieved September 17, 2015, from [http://www.uclan.ac.uk/research/explore/projects/assets/UNESCO\\_Questionnaire\\_v5.doc](http://www.uclan.ac.uk/research/explore/projects/assets/UNESCO_Questionnaire_v5.doc).
- , and C. de Vos, eds. 2012. *Sign Languages in Village Communities: Anthropological and Linguistic Insights*. Sign Language Typology 4. Berlin: Mouton de Gruyter.

Copyright of Sign Language Studies is the property of Gallaudet University Press for Sign Language Studies and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.