

Foreword

It is an honor to be asked to write the foreword to this timely and important volume in which the editors and authors—all experts in the field—share their knowledge and experience in creating and working with signed language corpora. I was an early advocate of a corpus-based approach to signed language linguistics, so it is pleasing to see this excellent and comprehensive introduction appear. Students of deaf studies and linguistics, some of whom will be among the next generation of signed language researchers, will find in this volume a huge amount of information to help them plan and execute their corpus-based research programs.

As a researcher from an older generation, and looking back over the 45 years of my involvement in signed language research, I can recall four main issues that stood out for me from the very beginning. First, I believed that empirical language description and testable grammatical claims demanded naturalistic and representative usage-based data, not just introspective judgments reported to or made by the language researcher. Second, it was clear to me that in order to conduct linguistic analysis, researchers needed to find a way to “capture” transient and ephemeral face-to-face signed language utterances, rather than rely on written glosses or transcription systems. Third, I considered the existence and relative percentage of native signers in signing communities to be a key factor in understanding much of the variation in these languages, as well as being relevant in discriminating different types of signed languages found around the world. However, these or other details about the language informants were often not reported when discussing findings. Finally, systematic and extensive recordings of Auslan (Australian Sign Language) were not available. I had a sense of urgency in the documentation of Auslan because it was undergoing rapid change and, for a number of reasons, even appeared to face endangerment within a generation or two. I struggled with these issues from 1978 until the late 1990s, when a productive way forward for signed language linguistics became apparent and feasible; namely, digital multimedia corpora of signed languages.

In 1978, I was in Paris following the lectures of Michel Foucault, Gilles Deleuze, Félix Guattari, and Guy Hocquengem as part of my doctoral research into the sociology of language and the cultural semiotics and representation of sexuality and gender. I was asked by a family friend from Australia to interpret for her during a two-week UN-funded international workshop for deaf actors at the Chateau de Vincennes, in a suburb of Paris. While I interpreted from French and sometimes English into “our sign language” (it was not yet named or identified as “Auslan”) during those two weeks, I met many deaf people from many different countries. I also met Bernard Mottez (a French signed language researcher and advocate), Harry Marcowicz (an American sociolinguist working with Bernard), and Bill Moody (an American Sign Language [ASL] interpreter who had been working with the deaf community in France). They were extremely curious about my background (Coda) and the signed language I was using. It proved to be a life-changing experience because each of them urged me to work with the deaf community in Australia and research the signed language I had grown up with because virtually nothing was known about it. Within weeks, I had decided to change my academic direction and began the shift from sociology to linguistics and the study of signed languages.

Back in Australia, I completed the required studies in linguistics and began to research the literature on signed language linguistics, which was mostly in English and mostly about ASL. I was struck by how confident most of the scholars were in their papers. They did not hesitate to describe what was or was not possible, or grammatically correct, in their signed languages, and to propose exceptionless categorical rules to account for this in their grammars. Despite the fact that so many of the accounts of constructions in ASL and other signed languages resonated with me because almost identical constructions could be identified in Auslan, claims that x-construction was the only grammatical way to utter x-meaning in their signed language did not ring true to me. I was aware of too much variation within the signing community I grew up with, and I was also keenly aware that acts of enactment and depiction, which were relatively frequent in Auslan (and which I had also observed in other signed languages), were rarely mentioned in the literature, or were analyzed in a way that obscured or denied their underlying symbolic motivation. Something was not right.

To me, one explanation for this situation was that most of these early researchers were reporting on the elicited judgments of very small sets of

deaf signers (sometimes consisting of only one or two favored consultants) or, in some cases, they were reporting on their own intuitions. They were schooled in linguistic frameworks that regarded small datasets (elicited or introspected) as reliable and adequate, so most of these researchers did not conduct fieldwork or collect usage-based data as did descriptive and documentary linguists, or linguists studying sociolinguistic variation, anthropological linguistics, comparative linguistics, and linguistic typology. Another explanation was that naturalistic data were much more likely to include the types of enactment and depiction that, I felt, were strangely absent from the early signed linguistics literature, because it relied heavily on invented or elicited examples and grammaticality judgments.

What was not right to me was that there were no accessible, representative, usage-based datasets on which to base our descriptions and generalizations, or against which our claims could be tested. They needed to be created.

The Introduction comprehensively covers the theoretical and methodological reasons for using corpora in language study. Here, in expanding on the four issues I listed above, I hope to underline how the creation of signed language corpora is both a response to and a way of addressing these very issues.

With respect to the first issue (empirical usage-based research), most of the chapters in this volume focus on the collection, creation, and use of signed language corpora to this end. The reader will thus find excellent detailed descriptions dedicated to the design and collection of corpora (sampling, set-up, activities, text types, privacy considerations, etc.); best practices in the annotation of corpora in order to facilitate linguistic analysis (e.g., tokenization, ID-glossing, transcription, grammatical tagging); ways of searching and sorting these annotations within multimedia annotation programs, such as ELAN, or in databases into which they have been exported for further processing; the types of corpora that can be collected (reference, monitor, bilingual, interpreter, acquisition, etc.) and the uses to which they can be put (e.g., descriptive linguistics, teaching, interpreter training, language change studies); and the new and constantly improving language-related technologies (e.g., automatic computer recognition) that have the potential to allow for semi-automated annotation, gesture-based or sign-based human–computer interfaces, and even automated translation from signed to spoken languages. It is timely to see all these topics covered so comprehensively and presented in a single volume.

With respect to the second issue (“capturing” language), until the discovery of technologies for recording vision and sound, one could only do this using some kind of writing system. Writing freezes language and creates texts that can be reflected upon, studied, held up as models, and even idealized. In contrast, there are no everyday writing systems for signed languages, and thus no reference texts have been created for any of them. Researchers compromised by using contextually based glosses in scholarly papers or monographs to represent examples of individual signs or signed utterances. It was thus difficult for most readers to know with certainty exactly which signs were being referred to by each gloss, effectively making proper scientific evaluation of the claims or observations being made extremely difficult. The development of dedicated notation systems for use in transcribing signed languages failed to solve this problem because they have never been widely used in the research community (most are unable to deal adequately with the nonmanual and prosodic aspects of signing), and thus have not been used to create signed language texts of any significant length that could, in turn, constitute a simple corpus.

Ultimately, corpora of signed languages, just like corpora of spoken languages, were created by using the new digital technologies of recording vision and sound, collecting the recordings in archives, and annotating them using multimedia annotation software. There are chapters in this volume that carefully explain best practice in creating time-aligned annotations for signed corpora. Exactly what is annotated in each corpus at any particular stage of analysis varies considerably according to the aims of the researchers and the linguistic framework that informs their practice. The relevant observation I wish to make here is that multimedia annotation software means that signed language utterances need not be ephemeral and transient. Not only can they be captured and played back repeatedly, as any analog recording can, but the digital annotation files can be searched (e.g., a particular sign gloss, or a particular word or phrase used in a translation), and all instances in a file (or all the files in a corpus) almost instantaneously located. In such a corpus, one is not obliged to create a systematic and detailed transcription—as in, one that enables the reader to re-create what was originally signed and thus understand and evaluate the phenomenon one is referring to—in the first instance, just to enable analysis to begin or to be shared. The reason is that the video is always in view. Indeed, electronic versions of scholarly articles and monographs

can now hyperlink a glossed sign or sentence to a corpus timestamp where the video and its annotations can be almost instantly viewed. It is anticipated that this will eventually become the norm for signed language citations in scholarly publications, whenever privacy considerations allow.

With respect to the third issue (native signers), I, for one, have long considered the existence of native signers in deaf signing communities to be one of the defining characteristics of the signed languages that were first identified and described by linguists from the middle of the 20th century. The languages that were initially identified had institutional and familial histories. At the very least, there were users who had learned them as children from other older deaf students or from signing teachers at special residential schools for the deaf. The signed language had thus been learned as a first language by several generations of deaf children during the life of these institutions. In some cases, deaf (and hearing) children had acquired the signed language in families with hereditary deafness from their deaf parents and/or deaf siblings. These children acquired the language natively from birth, and their experience was, for all intents and purposes, comparable to the native speakers of spoken languages, except it was simply not in speech. These signed languages were used in everyday life within deaf families, communities, and schools to fulfill the core communicative functions found in any language. These deaf native signers also attended the residential schools where they became conduits of the language to their peers.

Native signers are central to the description of signed languages as ordinary languages, that is, languages used by deaf people who have experienced no language trauma, such as the delayed acquisition of a first language. Without the existence of intergenerational transmission, one could have argued that these signed languages were actually more like atypical spoken languages, such as pidgins (which appear at specific times and places, have shallow historical depth, and initially have no native speakers), than ordinary spoken languages.

Despite the importance of native signers, nonnative and late learners of signed languages are still very relevant to the description of signed languages. First, the entire signing deaf community needs to be documented to fully understand the dynamics of language use in these communities. To this end, the metadata that is routinely collected and associated with language samples in linguistic corpora allow researchers to keep track of exactly who produced a sample of language, in what context, and to

what end. Thus, it is no surprise that IMDI metadata standards for signed language corpora include descriptors for nativeness and age of acquisition. The standards allow the languages of native and nonnative signers in a single signing community to be compared.

Second, and potentially a problem, other subsequently identified signed language-using deaf communities or social networks around the world have been reported as having very few, if any, native signers associated with them, perhaps because they have much more recent origins and/or are very small. Many lack institutionally based historical depth (formal schooling was never available to deaf children) and/or multi-generational deaf families (hereditary deafness being rare or unknown). Extending a corpus-based approach to these languages will mean that once again metadata can be used to compare the language of native users with nonnative users, but this time between native users of established signed languages and the (nonnative) users of signed languages that do not have any native signers. One may suspect that they differ in much the same way as the language of a home signer is different to that of a signer exposed to signing from birth, a pidgin is different to a natively acquired language, or perhaps a late learner of a second language is different to a native speaker of that language; but, of course, only systematic study, using corpora, can establish the facts.

Finally, with respect to the fourth issue (language endangerment), the very act of creating a linguistic corpus results in an archival resource that can be consulted and then used for language vitalization or revitalization activities in an attempt to reverse decline. If, despite this, the language should become extinct, then at least a record of it will remain. Yes, there are signed language communities that do not feel their language faces endangerment at all. In these cases, a monitor corpus or at least two reference corpora that take a snapshot at two well-separated time intervals, for example, a decade, can be used to track language change in the community and gauge the language's vitality and viability.

The editors and authors of this volume are to be congratulated. They have explained the theoretical and methodological principles that underline corpus-based research and its importance in signed language research, and they have explained the practicalities, including ethical considerations, of creating and using signed language corpora. I look forward to the current and next generation of researchers adding to the small but growing literature based on signed language corpora (most of which is cited in this volume). I especially look forward to research

testing existing accounts of established signed languages, as well as new accounts of recently identified and/or emerging signed languages.

In closing, it is pleasing to note that deaf scholars are the authors or coauthors of the majority of the chapters in the volume, and it is particularly pleasing for me to note that two of the authors (Gabrielle Hodge and Adam Schembri) have previously studied with me as doctoral students in Australia earlier in their careers, and that one of the editors (Julie A. Hochgesang) was a student in a course on corpus linguistics and signed languages that I taught at Gallaudet University in 2009.

Trevor Johnston
Sydney, July 2021