Proceedings of the 10th Workshop on the Representation and Processing of Sign Languages (sign-lang@LREC 2022), pages 173–180
Language Resources and Evaluation Conference (LREC 2022), Marseille, 20-25 June 2022

Signed Language Transcription and the Creation of a Cross-linguistic Comparative Database

Justin M. Power[1], David Quinto-Pozos[1], Danny Law[1]
[1]The University of Texas at Austin
305 E. 23rd Street STOP B5100, Austin, TX 78712, USA
justinmpower@utexas.edu, davidqp@austin.utexas.edu, dannylaw@austin.utexas.edu

Abstract

As the availability of signed language data has rapidly increased, sign scholars have been confronted with the challenge of creating a common framework for the cross-linguistic comparison of the phonological forms of signs. While transcription techniques have played a fundamental role in the creation of cross-linguistic comparative databases for spoken languages, transcription has featured much less prominently in sign research and lexicography. Here we report the experiences of the Sign Change project in using the signed language transcription system HamNoSys to create a comparative database of basic vocabulary for thirteen signed languages. We report the results of a small-scale study, in which we measured (i) the average time required for two trained transcribers to complete a transcription and (ii) the similarity of their independently produced transcriptions. We find that, across the two transcribers, the transcription of one sign required, on average, one minute and a half. We also find that the similarity of transcriptions differed across phonological parameters. We consider the implications of our findings about transcription time and transcription similarity for other projects that plan to incorporate transcription techniques.

Keywords: multi-sign-language resources, transcription, cross-linguistic comparison, HamNoSys

1. Introduction

Over the past two decades, signed language data in various forms have become increasingly available and widely accessible. Numerous online dictionaries are available for individual signed languages; many of these dictionaries have been created by researchers and have been used in scholarly research (e.g., Stumpf et al., 2020; Caselli et al., 2016; Hochgesang et al., 2020). These resources typically include thousands of entries with videos of sign articulations and with other representations of the forms of signs—such as images, transcriptions, and theoretically-informed coding schemes. Several corpus projects have also been started during the past two decades for, inter alia, Australian Sign Language (Johnston & Schembri, 2007), *Deutsche Gebärdensprache* (Prillwitz et al., 2008), British Sign Language (Schembri et al., 2013), and *Russkiy Zhestovyi Yazyk* (Russian Sign Language, Kimmelman et al., 2022). New approaches using recently-developed software, such as OpenPose (Cao et al., 2021), promise to increase the amount and the type of data—as well as the precision of the data—that are available to researchers in the future (e.g., Kimmelman et al., 2020).

Although there has been a marked increase in the amount of signed language data available, cross-linguistic comparative resources are still few in number. Websites such as spreadthesign.com have made it possible to visually compare signs across dozens of languages—though it is unclear whether the data presented in such resources are representative of each respective signing community because the methodologies used to collect data are not always well described. Recent large-scale comparisons of lexical signs (Yu et al., 2018) and of manual alphabets (Power et al., 2020) have used such websites as sources of cross-linguistic data. The network of SignBank dictionaries, which uses a common approach and format (Crasborn et al., 2020; Hochgesang, 2022), offers the possibility of large-scale cross-linguistic comparison in the future (Börstell et al., 2020). In addition to the coding system developed for the Global Signbank, scholars have begun to collaborate on the development of a common phonological coding scheme for resources in American Sign Language (Becker et al., 2020). And, the Dicta-Sign project has created a framework for comparing basic vocabulary and connected signing across four signed languages[1] (Hanke et al., 2010). Despite these welcome advances in the creation of comparable phonological representations, currently, there are extremely few open cross-linguistic comparative databases that include easily comparable representations of the forms of signs.

The challenge of creating a comparative framework for cross-linguistic data confronts scholars of both signed and spoken languages (Forkel et al., 2018). However, scholars of spoken languages have inherited practical orthographies as well as a transcription system, the International Phonetic Alphabet (IPA), which has been used for more than a century to represent hundreds of spoken languages in precise phonetic detail (Albright, 1958); and, the IPA is itself based on alphabetic symbols with millennia-old histories. With the benefit of this inheritance, transcription techniques have played a fundamental role in the creation of cross-linguistic comparative databases for spoken languages and in the possibility of extending those comparisons to new languages (e.g., Greenhill et al., 2008). In contrast, all currently used systems for writing signed languages have short histories, and the transcription of signed language data using one of these systems is not widely practiced (Hochgesang, 2014). In sum, transcription has featured much less prominently in the creation of cross-linguistic resources for signed languages.

Here we report the experiences of the Sign Change project[2] in using transcription techniques to create a cross-linguistic comparative database of basic vocabulary across 13 signed languages. Begun in September 2020, this three-year project aims to make its comparative database open to scholars at the end of the project in mid-2023. To create the comparative database, we have used the transcription system HamNoSys to represent the forms of signs in a uniform system (Hanke, 2004). By using transcription techniques, our aim has been to make the database

[1] British, German, French, and Greek Sign Languages.

[2] See the project website: liberalarts.utexas.edu/lrc/sign-change.

expandable with transcribed data from other signed languages and with data produced within other projects.

In this paper, we report information that concerns two aspects of signed language transcription—namely, transcription time and transcription similarity. That is, first, following a training period in which research assistants learned to transcribe signs in HamNoSys, how much time, on average, is required for those research assistants to complete transcriptions of signs? And, second, once completed, how similar are the transcriptions that were produced by multiple trained transcribers? In exploring these aspects of the transcription process, we aim to inform future projects that use transcription techniques to create cross-linguistic signed language resources.

We report the results of a small-scale study in which we compare two trained transcribers after one year of experience transcribing signs. Thus, although the study is small in scale, it provides a window into the transcription process through the lens of relatively well-trained transcribers. We explore the factors which affected the dimensions of transcription time and similarity, such as the phonological complexity of a sign and the number of symbols systematically required in a HamNoSys transcription. We note a relationship between the number of symbols required in a transcription and the amount of time needed for that transcription. We find that particular phonological parameters of the sign, such as handshapes, were transcribed with a greater degree of similarity than other parameters, such as movement. Finally, we discuss the practical implications of the study for a project that incorporates transcription techniques. We also discuss the theoretical issues raised by the study and propose future research that would compare the transcription process across signed and spoken languages.

2. Methods

2.1 Transcribers

The two transcribers in this study (hereafter transcriber-1 and transcriber-2) were undergraduate research assistants who worked on a part-time basis (up to 10 hours per week). At the time of the study, transcriber-1 had completed three years of a four-year degree program in linguistics, and transcriber-2 was in the final semester of a four-year program in audiology. In their courses of study, both transcribers had been trained to use the International Phonetic Alphabet, and both were second-language adult learners of American Sign Language (ASL) who had taken multiple courses in that language at the undergraduate level. The ASL coursework for each transcriber totaled approximately 225 face-time hours of instruction over an 18-month period; transcriber-1 had also assisted a deaf instructor with first-semester courses for an additional 270 hours of exposure to classroom ASL. Prior to their training in HamNoSys transcription, neither transcriber had been trained in any signed language transcription system, and neither had any knowledge of HamNoSys.

Both transcribers completed an initial training period of approximately one month. The training period, which was conducted by the first author, included weekly hour-long sessions that focused on the transcription of one sign parameter per week (i.e., handshape, orientation, location, movement). For lesson material, we used the *HamNoSys 4 Handshape Chart* (Hanke et el., 2010) and the *HamNoSys 4.0 User Guide* (Smith, 2013). Using video data, the transcribers then practiced transcription techniques during the week following each session. After their initial training period, both transcribers received ongoing training on an as-needed basis, as well as feedback on their transcriptions.

The training program instructed the transcribers to produce relatively narrow transcriptions. A comparison of a relatively broad versus a relatively narrow HamNoSys transcription is given in (1); the transcriptions are aligned to group parameters in a visually distinctive way. The transcription in (1a), which is taken from Hanke (2018), represents a relatively broad transcription. The location of the hand at the beginning of the sign is close to the right side of the forehead; this relationship is indicated in (1a) by the three symbols, which respectively represent the forehead, the right side, and closeness. In the relatively narrow transcription in (1b), there are seven additional symbols—the open and close parenthesis symbols group the five number symbols (1 2 3 4 5), which represent the fingers that are close to the forehead as well as the finger part symbol, which represents the tips of the fingers.

(1) a. ‘Hamburg’ *DGS*
b.

At the time of the study, transcriber-1 had completed 13 months in the project, and transcriber-2 had completed 12 months. By that time, the transcribers had each gained experience transcribing more than 1,000 basic vocabulary signs from multiple sign languages, including ASL, British Sign Language, *Lengua de Señas Mexicana*, *langue des signes française*, *Nederlandse Gebarentaal*, *Langue des signes de Belgique francophone*, and others; and they had experience editing many hundreds of transcriptions that had been completed by other transcribers.

2.2 Signs

For the study, we selected 100 basic vocabulary signs from *Vlaamse Gebarentaal* (VGT, or Flemish Sign Language), a language that, prior to the study, neither transcriber had had experience transcribing. We used the VGT Signbank (Vlaams GebarentaalCentrum, 2018) as a source for basic vocabulary signs in *Vlaamse Gebarentaal*. Table 1 shows the concepts that were included in the study. Sixty-three (63) of the signs in the table were articulated with one hand, and 37 were articulated with two hands. Only one sign had two parts[3]: the sign meaning ‘dull’ (i.e., blunt) is composed of parts meaning ‘not’ and ‘sharp’.

Because entries in the VGT Signbank do not include information about part of speech, in order to organize the 100 concepts by part of speech (as in Table 1) we had to rely on the part of speech of the Flemish translation of each sign. That is, we had to assume that the sign meaning ‘mother’ in *Vlaamse Gebarentaal* shares the same part of speech as the word *moeder* in Flemish—namely, noun. Given problems classifying signs into word classes in some sign languages (Schwager & Zeshan, 2008), this assumption represents a potential limitation of our study.

[3] In our database, each sign part is transcribed in a separate row, such that each row has only one value per parameter.

Part of speech	Concepts
Noun (n=45)	animal, ant, back, belly, bird, child, cloud, day, dog, ear, earth, egg, eye, father, feather, fingernail, fire, fish, flower, fly, fog, foot, fruit, hair, hand, heart, knee, man, meat (2), moon, mother, name (2), night, nose, person, stone, tongue, tooth, tree, water, wife, woman, year
Verb (n=19)	bite, blow, breathe, burn, come, count, cry, cut, die, do, drink, eat, fall, fight, float, flow, fly, know, play
Adjective (n=26)	all, bad (2), big (2), black (2), cold (2), dirty, dry, dull, fat, few, five, four, good, green, heavy, new, old, three (2), two, white, yellow
Adverb (n=7)	far, how, not, what, where, who, yesterday
Pronoun (n=3)	he, I, you

Table 1: Concepts used in the study, organized by part of speech. (2) = two variants of the relevant concept.

In addition to their organization in part of speech categories, the concepts in Table 1 can be grouped into semantic and semiotic categories. For example, five of the signs in our data represent colors: black (2 variants), green, white, and yellow; five signs refer to numerals: five, four, three (2 variants), and two; and thirteen concepts in the noun category refer to a part of the body: back, belly, ear, eye, fingernail, foot, hair, hand, heart, knee, nose, tongue, and tooth. Sixteen of the signs in the table are thought, according to Parkhurst & Parkhurst (2003), to be non-iconic vocabulary items: bad (2 variants), black (2 variants), father, good, green, how, mother, name, new, old, play, what, where, white, who, and year.

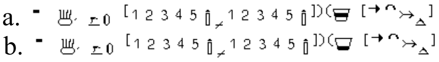
2.3 Transcription

In advance of the study, we prepared two identical copies of an online spreadsheet in which the transcribers were instructed to complete their transcriptions. For each sign, the spreadsheet included columns for the concept in English, a translation of the concept in Flemish (taken from the VGT Signbank), and the link to the VGT Signbank entry page for each sign. In addition, the spreadsheet had separate columns for each sign parameter and for symmetry, for the number of hands used to articulate a sign, and for whether the signer appeared to be right- or left-dominant. Finally, the spreadsheet also had a column in which the transcribers were instructed to record the amount of time (in minutes and seconds) that it took them to complete each sign transcription.

The transcribers only had access to their own copy of the spreadsheet; they were asked not to consult with one another while completing their transcriptions. They used a web-based HamNoSys input tool (http://www.sign-lang.uni-hamburg.de/hamnosys/input/) to compose their transcriptions, which they subsequently copied into their spreadsheet. They were allowed to complete their transcriptions at their own pace during a two-week period: transcriber-1 completed the 100 transcriptions over the course of six days, while transcriber-2 took five days to complete the transcriptions.

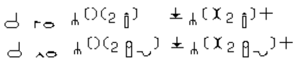
2.4 Comparison Methodology

We used Levenshtein distance to pairwise compare the difference between transcriptions. We normalized the resulting distance scores by dividing each score by the length, ignoring spaces, of the longer string in the relevant pairwise comparison. We then subtracted the normalized scores from 1 in order to have a measure of the similarity of each pair of strings. Consider, for example, the transcriptions in (2), which are taken from the study data.

(2) a.  'big' *VGT*
b.

There is only one difference between the transcriptions in (2a-b)—namely, the location symbols and , which represent different heights in the neutral space in front of the signer; thus, the Levenshtein distance for this comparison is 1. Each transcription consists of 28 symbols (handshape=2 symbols, orientation=2, location=17, movement=6, symmetry=1). The normalized Levenshtein distance between the two strings is thus 1/28≈0.036; and the similarity score is 1-0.036≈0.964.

Consider now the transcriptions in (3). There are 5 differences between the transcriptions in (3a-b): one difference in the orientation parameter (vs.), one difference and one additional symbol in the location parameter (vs.), and again one difference and one additional symbol in the movement parameter (vs.). The length of the longer string (3b) in this comparison is 21. Thus, the normalized Levenshtein distance between the two strings is 5/21≈0.238; and the similarity score is 1-0.238≈0.762.

(3) a.  'nose' *VGT*
b.

We used Python (Van Rossum & Drake, 2009) and the pandas library (McKinney, 2010) to perform the comparisons. For statistical analyses, we used NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), and researchpy (Bryant, 2021). Figures were produced using Matplotlib (Hunter, 2007).

3. Results

In this section, we report the results of our two main analyses—namely, of the time required per transcription and of the similarity of the transcriptions produced by the two transcribers. As part of the timing results, we highlight selected factors that evidently affected the rate at which the transcribers completed their transcriptions.

3.1 Transcription Time

Taken together, the two transcribers averaged 95.2 seconds (*SD*=38.1) per transcription—roughly, one minute and a half. But, they completed their transcriptions at different rates. On average, transcriber-1 (*M*=82.2 seconds, *SD*=26.7) completed each transcription at a significantly faster rate, according to a Welch's t-test, than did transcriber-2 (*M*=108.3 seconds, *SD*=43.1), $t(166.6) = -5.1$, $p < .001$. At present, we have not attempted to count the number of errors in the transcribers' transcriptions or to assess the relationship between errors and speed.

3.1.1 Effect of the Number of HamNoSys Symbols on Transcription Time

Intuitively, the more symbols that are required by HamNoSys for a transcription, the longer it will take on average to complete that transcription. In our data, there is, as expected, a significant positive correlation between the number of symbols used in a transcription and the time it took to complete the transcription, $r(200) = .72$, $p < .001$.

In consequence of this general feature of the transcription process, one-handed signs were transcribed more quickly than two-handed signs: on average, one-handed signs were transcribed in 84.8 seconds (*SD*=31.4), while two-handed signs were transcribed in 113.1 seconds (*SD*=41.8). There is a significant correlation between the number of hands used to articulate a sign (1 or 2) and the amount of time needed per transcription, $r(200) = .36$, $p < .001$. This correlation is expected because there is a systematic difference in the number of symbols required by HamNoSys to transcribe one- versus two-handed signs. HamNoSys minimally requires one extra symbol—namely, a symmetry symbol—to transcribe two-handed versus one-handed signs. And, for any parameter in a two-handed sign that is asymmetrical, at least four extra symbols are required: three meta-symbols to show which is the dominant and which is the non-dominant hand, and at least one symbol to represent the non-dominant parameter.

3.1.2 Effect of Grammatical, Semantic, or Semiotic Features and Transcription Time

There is an evident relationship in our data between the grammatical, semantic, or semiotic features of certain concepts and the average number of symbols that were used to transcribe signs representing those concepts. In consequence, there is a difference in the average amount of time required to complete a transcription based on features of the concept. Table 2 reports the average number of symbols used in a transcription and the average time per transcription, broken down by part of speech and by selected semantic and semiotic categories for the 100 concepts included in the study. The non-iconic signs in Table 2 are those concepts in our data that, according to Parkhurst & Parkhurst (2003), may represent non-iconic concepts in four European signed languages and in regional varieties of *Lengua de Signos Española*; see Section 2.2.

Part of speech	Mean symbols per transcription	Mean time (seconds)
Pronoun (n=6)	12.0 (6.0)	63.2 (16.9)
Adverb (n=14)	14.4 (4.8)	72.4 (13.0)
Verb (n=38)	17.7 (7.1)	89.7 (31.6)
Adjective (n=52)	18.2 (8.6)	96.1 (43.4)
Noun (n=90)	23.6 (10.7)	102.8 (38.1)
Semantic category		
Numeral (n=10)	7.8 (2.5)	59.9 (16.0)
Body part (n=26)	20.3 (7.8)	91.2 (31.3)
Color (n=10)	18.5 (9.5)	100.2 (41.8)
Semiotic category		
Non-iconic (n=18)	20.2 (9.1)	97.6 (34.2)

Table 2: Average transcription time (in seconds) and length of the transcription in selected grammatical, semantic, and semiotic categories; standard deviations are provided in parentheses in the two rightmost columns.

If we take the number of HamNoSys symbols used to transcribe a sign as a rough estimate of the phonological complexity of that sign—that is, if we assume that increased phonological complexity will require, on average, more symbols to transcribe—then the results in Table 2 suggest that phonological complexity is unevenly distributed across certain parts of speech and across certain semantic categories in *Vlaamse Gebarentaal*. For example, pronouns and numeral signs in that language were transcribed using relatively few symbols, whereas nouns, body part signs, and non-iconic signs were transcribed using comparatively more symbols.

In sum, there are at least two factors that affected the time per transcription in this study—namely, the transcriber and whether a sign was one- or two-handed. In addition, our results suggest that the grammatical, semantic, or semiotic features of a concept may affect the phonological complexity of a sign—as measured by the number of symbols required to transcribe the sign—and, thus, the average amount of time required for a transcription.

3.2 Transcription Similarity

Using the comparison methodology outlined in Section 2.4, we measured the similarity of each pair of transcriptions that were produced for the same concept. The average similarity of a pair of full transcriptions was 0.69 (*SD*=0.18) for all 100 pairs. The distribution of similarity scores for all 100 transcription pairs is shown in the histogram in Figure 1.

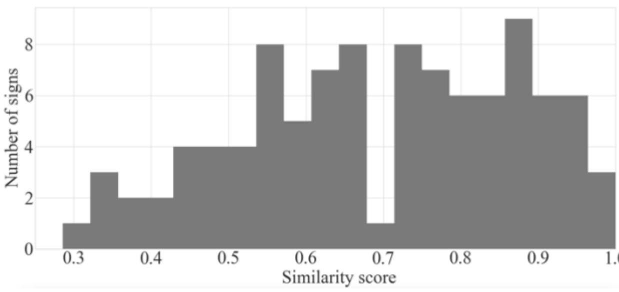


Figure 1. Distribution of similarity scores of full transcriptions for all 100 pairs of signs.

Just two pairs of transcriptions were exactly the same, but fifteen pairs scored 0.9 or higher according to our similarity metric and 84 pairs were at least 0.5 similar.

Similarity was not evenly distributed across all parts of the full transcriptions; that is, some parameters were more similarly transcribed than others. Our approach to organizing transcriptions in a spreadsheet (see Sec. 2.3) allowed us to individually compare the sign parameters (handshape, orientation, location, and movement) and other global aspects of the signs, such as symmetry, the number of hands used to articulate the sign, and hand dominance.

We first report our results pertaining to the last two of these global aspects of the sign. With respect to their coding of the number of hands used to articulate a sign, the two transcribers differed in only one comparison. This difference was apparently a mistake: one transcriber coded the sign as being produced with one hand, but then also used a symmetry symbol in the transcription, which is a type of symbol that is only used to transcribe two-handed signs. Thus, this transcriber likely thought the sign was articulated with two hands, but mistakenly coded the sign as being produced with one hand. The transcribers differed in four comparisons with respect to hand dominance. In two of these four differences, the signs are articulated using two hands; the signs are also symmetrical. Hence it would be challenging (and perhaps impossible) to determine hand dominance solely based on the articulations of these two signs. When viewing other signs in the data set produced by the same signer, it becomes clearer that the signer is likely left-dominant. In the two other comparisons in which the transcribers came to differing conclusions about hand dominance, the signs are articulated with the left hand.

Figure 2 shows the distribution of similarity scores across transcriptions of handshape, orientation, location, movement, and symmetry. Note first that there is a relationship between the number of symbols used to transcribe each parameter and the distribution of scores. For example, an orientation transcription for a one-handed sign requires exactly two symbols: one symbol to represent the orientation of the back of the palm and one symbol to represent the relative orientation of the palm. Hence, using our comparison methodology, the similarity scores can only be 0 (both symbols different), .5 (one symbol different and one identical), and 1.0 (both symbols identical); and the distribution of similarity scores in Figure 2 largely reflects these possible scores.

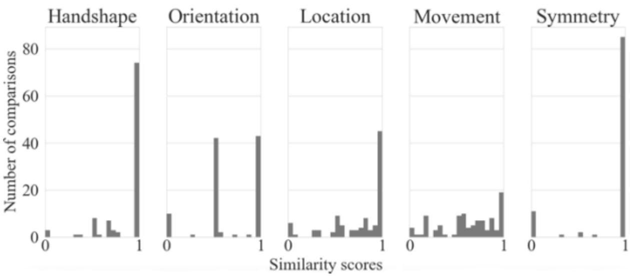


Figure 2: Comparison of the distributions of similarity scores for the transcriptions of four parameters and of symmetry.

In addition to the general relationship in HamNoSys between certain parameters and the number of symbols used to transcribe them, there was also an imbalance in our data in the number of symbols used to transcribe each parameter. For example, whereas handshape transcriptions comprised on average 2.8 symbols (*SD*=1.7), movement transcriptions comprised 8.8 symbols (*SD*=5.7) on average. Intuitively, if more symbols are needed to transcribe a given parameter, then there is a greater number of opportunities for differences across transcriptions and, perhaps, a greater number of differences. However, we found no significant correlation in our data between the average length of a pair of transcriptions and the average similarity score for that pair, $r(100) = -.01$, $p = .90$. In addition, there was no significant correlation between these factors when separately considering each parameter. Among the four main parameters, the strongest correlation was found for movement; but even this correlation is weak and is not significant: $r(100) = .16$, $p = .11$.

On average, transcriptions of handshapes (*M*=.88, *SD*=.24) and of symmetry values (*M*=.87, *SD*=.32) scored highest for similarity, followed by locations (*M*=.76, *SD*=.30), orientations (*M*=.67, *SD*=.33), and movements (*M*=.63, *SD*=.29). Consider the difference in similarity scores between transcriptions of handshapes and of movements. Although, as we have seen, more symbols were used on average in our data to transcribe movements than to transcribe handshapes, there was only a weak correlation between the number of symbols used to transcribe a parameter and the similarity score for that parameter. Hence the difference in the average similarity scores of handshape transcriptions versus scores of movement transcriptions (0.88-0.63=0.25) at best can be only partly explained by the difference in the average number of symbols used to transcribe the two parameters. Thus, these results may suggest that it is comparatively easier, in a sense, to accurately transcribe handshapes than it is to accurately transcribe movements.

4. Discussion

In this section, we consider how the results in Section 3 might inform future projects that incorporate transcription methods. We also discuss questions raised by our results that pertain to the phonological features of signs and the transcription process.

Before highlighting the practical lessons that can be gleaned from our study, it is important to note one preliminary point. First, our study was designed to compare two transcribers and their transcriptions prior to any subsequent editing of those transcriptions. For each sign, the current study resulted in two transcriptions that were not edited by any other individual. Thus, the results of the study differ in two ways from the results that we aim for in our project. Although just one transcriber in our project completes an initial transcription of a sign, that transcription is edited by at least two other members of the transcription team in successive stages. Our aim is to arrive at one best transcription of a sign, rather than multiple, unedited transcriptions of that sign.

4.1 Transcription Time in the Creation of a Comparative Database

Our analyses in Section 3 focused on the amount of time it took for the transcribers to complete transcriptions and on the similarity of their transcriptions. With respect to the time required for transcriptions, our results may reflect a relatively conservative estimate (approx. 1.5 minutes per transcription; see König & Langer, 2009, who report that one minute of *DGS* text requires 135 to 200 transcription minutes, depending on the details included in the transcription). Why is our estimate relatively conservative? As briefly discussed in Section 2.1, our project aims to produce narrow transcriptions in HamNoSys that, for example, in a sign involving contact, provide details about the exact part of the hand (or parts of the hands in two-handed signs) that make contact with the body or with the nondominant hand. Our approach to transcription will tend to require a greater number of symbols per transcription than will an approach that systematically aims at broad transcriptions. And, as we have shown in Section 3.1, there is a relationship in our data between the number of symbols used to transcribe a sign and the time it takes to complete a transcription.

This general finding about the relationship between transcription time and the number of symbols required by HamNoSys has several consequences for any project that incorporates transcription methods. For instance, the balance of one- versus two-handed signs in a dataset will affect the amount of time required to transcribe that dataset using HamNoSys and, perhaps, using other sign transcription systems. In our data, one-handed signs were transcribed on average in under 1.5 minutes, while it took nearly 2 minutes on average to transcribe two-handed signs (84.8 seconds versus 113.1 seconds, see Sec. 3.1). That average difference of 28.3 seconds can result in large differences in the time necessary to complete transcriptions in a project with a large dataset. Compare, for example, one dataset of 2,000 signs with a balance of one- versus two-handed signs that matches our dataset (63% vs. 37%) and a second dataset of 2,000 signs with an equal balance (50% vs. 50%). Based on our results, we estimate that the first dataset would require 29.7 hours for the one-handed signs (1260 signs * 84.8 seconds) and 23.2 hours for the two-handed signs (740 signs * 113.1 seconds). For the second dataset, we estimate 23.6 hours for the one-handed signs (1000 * 84.8 seconds) and 31.4 hours for the two-handed signs (1000 * 113.1 seconds). The difference in time required for the two datasets is thus estimated at 2.1 hours for the initial draft transcriptions (55 hours – 52.9 hours).

Our results also suggest that the type of vocabulary that comprises a dataset can affect the amount of time it will take to transcribe that dataset. In Section 3.1, we showed that signs representing numerals and pronouns—signs that have been thought to be articulatorily simple—required the least amount of time to transcribe on average (approx. 1 minute for both categories); whereas signs representing nouns and colors both required more than 100 seconds on average to transcribe.

Finally, personal characteristics of the transcribers themselves will inevitably affect the time required to transcribe a dataset. In our study, the transcribers were asked not to focus too carefully on their speed—though by asking them to record the time required for each transcription, the study contained an implicit emphasis on speed. Because the study did not include a target transcription, but rather compared the similarity of each pair of transcriptions, we cannot measure the relationship between transcription time and accuracy.

Given our timing results, how long would it take to produce first draft transcriptions for datasets of various sizes that are similar to our dataset (i.e., datasets with similar balances of one- versus two-handed signs and with similar types of vocabulary)?

Number of signs	Time (hours)
1000	26.5
2000	52.9
5000	132.3
10000	264.6

Table 3: Estimated transcription time required for datasets of various sizes.

In our discussion of transcription time, we have not considered the time required for the editing process; nor have we discussed the potential use of avatar technology or of software such as OpenPose (Cao et al., 2019) in both the transcription and editing processes. Although the editing process was not the focus of the current study, based on the experience of our project, that process is at least as time-consuming as the first-draft transcription process.

4.2 Transcription Similarity

Our study compared the similarity of a pair of transcriptions; it did not attempt to directly measure the accuracy of transcriptions in comparison with target transcriptions. For purposes of this discussion, however, we will take the similarity score as a proxy for accuracy; that is, we will assume that higher similarity scores reflect more accurate transcriptions because, while possible, it is nevertheless unlikely that two transcribers would independently make similar mistakes in the transcription of a sign. Hence any agreement in their transcriptions likely reflects transcription accuracy.

As we have seen, there were differences in similarity scores across parameters: for example, handshape transcriptions showed higher similarity scores than did movement transcriptions (0.88 versus 0.63). Why were some parameters transcribed more accurately than others? In Section 3.2, we argued that the difference in the number of symbols used to respectively transcribe these parameters (on average, 2.8 versus 8.8) at best only partly explains the difference in accuracy. Recall that, overall, there was no significant relationship between the average number of symbols in a pair of transcriptions and their similarity score.

There may be multiple factors that influence the relative reliability of transcribing handshape versus transcribing other parameters, including factors concerning the perception of parameter values and factors related to characteristics of the transcription system. Regarding the former, handshapes might be perceived more categorically than orientations, locations, and movements. Efforts to determine if categorical perception exists in signed language have found that handshape shows traditional patterns of categorical perception for native signers of ASL, but the same is not true for location (Emmorey et al., 2003). Although the transcribers in this study were not native signers, they each had had a substantial amount of exposure to ASL before being trained in transcription. That exposure could have influenced their ability to transcribe handshapes in categorical ways.

Alternatively, the transcribers might have perceived other parameters, such as location and movement, in a more gradient fashion. For example, two typical locations in sign articulations are the right side of the forehead and the cheek; but, locations in-between these two typical locations can also serve as valid locations for a sign in discourse (Russell et al., 2011). The variability of location values during signing—as a result of phonetic and sociolinguistic factors, among others—might have been differently perceived by the two transcribers. A similar argument could be made for movement values in signs.

The options available within a transcription system might also have an influence on which symbols are used to transcribe parameter values. In the example of forehead-to-cheek variants that was given above, HamNoSys includes several options for coding different locations (e.g., beside the eyebrows, eyes, and cheek) that could correspond to signer productions. However, there might be other locations, orientations, and movements that do not correspond to signer productions—and vice versa. These mismatches between productions and symbols could result in less reliability across transcribers.

Yet another possible factor in the reliability of handshape transcription could be linked to the focus on handshapes during the learning of a signed language. Sign language curricula typically stress the importance of handshape values; often, handshapes charts feature prominently in such curricula. Signed language games that rely on handshape contrasts (e.g., so-called ABC handshape games in ASL) could serve to make a learner hypersensitive to the handshape parameter, which could lead to increased inter-transcriber reliability of handshape transcriptions.

Finally, handshape symbols in HamNoSys iconically represent the forms of handshapes in a way that is independent of any other articulatory aspect of the sign. That is, the symbol represents a handshape with an extended index finger—whether the handshape is produced in space, in contact with the body, at various heights, or with various orientations. In contrast, orientations are arguably more challenging to transcribe because the second symbol in an orientation transcription (e.g.,) represents the orientation of the palm with respect to the direction of the back of the hand. Thus, orientation symbols are arguably less iconic than handshape symbols, and they may be more challenging to transcribe.[4] In sum, transcription reliability could be influenced by factors that are linked to the properties of sign parameters and to the perception of those parameters. They could also be influenced by factors that are associated with characteristics of signed language learning. And, they may be affected by differences in the iconicity of HamNoSys symbols.

5. Conclusions and Future Research

The results of this study raise intriguing questions about the general process of transcription across signed and spoken languages. Is more time required to transcribe signs or words? Based on the third author's experience transcribing lexical data from many of the spoken languages of Mesoamerica, the transcription of spoken words in IPA requires much less time than the time it took in our study to produce transcriptions in HamNoSys. However, the type of narrow phonetic transcriptions produced by the signed language transcribers in our study are arguably much more fine-grained than the type of phonemic transcriptions that are typically produced for a word list with the basic vocabulary of a spoken language.

A second intriguing comparison with spoken language transcription concerns transcription similarity. In Section 4, we suggested that, because certain parameters such as handshape may be more categorically perceived than other parameters (Emmorey et al., 2003), it may be easier, in a sense, to assign categorical symbols to certain parts of sign articulations. It has been shown that stop consonants in English, for example, are more categorically perceived than vowels (Fry et al., 1962). Could there also be an imbalance in transcriber reliability across differing classes of speech sounds?

6. Acknowledgements

This research has been supported by NSF grant BCS-1941560 "Regularity and Genetic Relatedness in Sign Languages". As always, any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation.

7. References

Albright, R.W. (1958). The International Phonetic Alphabet: Its Backgrounds and Development. Reprinted in *International Journal of American Linguistics*, 24(1).

Becker, A., Catt, D., and Hochgesang, J.A. (2020). Back and Forth between Theory and Application: Shared Phonological Coding Between ASL Signbank and ASL-LEX. In *Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages*, 1–6. Marseille, France. European Language Resources Association (ELRA).

Börstell, C., Crasborn, O., and Whynot, L. (2020). Measuring Lexical Similarity across Sign Languages in Global Signbank. In *Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages*, 21–26. Marseille, France. European Language Resources Association (ELRA).

Bryant, C. (2021). researchpy Documentation, release 0.3.2. researchpy.readthedocs.io/_/downloads/en/latest/pdf/

Cao, Z., Hidalgo, G., Simon, T., Wai, S.-E., and Sheikh, Y. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Computer Society*, 43(1):172–186.

Caselli, N.K., Sehyr, Z.S., Cohen-Goldberg, A.M., and Emmorey, K. (2016). ASL-LEX: A lexical database of American Sign Language. *Behavior Research Methods*, 49:784–801.

Crasborn, O., Zwitserlood, I., van der Kooij, E. and Ormel, E. (2020). *Global Signbank Manual, version 2*. Radboud University, Centre for Language Studies.

Forkel, R., List, J.-M., Greenhill, S.J., Rzymski, C., Bank, S., Cysouw, M., Hammarström, H., Haspelmath, M., Kaiping, G.A., and Gray, R.D. (2018). Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. *Scientific Data*, 5:180205.

Fry, D.B., Abramson, A.S., Elmas, P.D., and Liberman, A.M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5(4):171–189.

Greenhill, S.J., Blust, R., and Gray, R.D. (2008). The Austronesian Basic Vocabulary Database: From Bioinformatics to Lexomics. *Evol. Bioinformatics*, 4:271–283.

Hanke, T. (2004). HamNoSys – Representing Sign Language Data in Language Resources and Language Processing Contexts. In *Proceedings of the Workshop on the Processing of Sign Languages (LREC 2004)*, 1–6. Lisbon, Portugal. European Language Resources Association (ELRA).

Hanke, T. (2010). HamNoSys 4 Handshapes Chart. https://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/HamNoSys_Handshapes.pdf

Hanke, T., König, L., Wagner, S., and Matthes, S. (2010). DGS Corpus & Dicta-Sign: The Hamburg studio setup. In P. Dreeuw, et al., (Eds.), *Proceedings of the Fourth*

[4] Our thanks to one of the reviewers for suggesting that the iconicity of handshape symbols may affect the relative accuracy with which handshapes are transcribed.

Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (LREC 2010), 106–109. Valletta, Malta. European Language Resources Association (ELRA)

Hanke, T. (2018). HamNoSys – Hamburg Notation System for Sign Languages. Hamburg, Germany: Universität Hamburg. https://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/HamNoSys_2018.pdf

Harris, C.R., et al. (2020). Array programming with NumPy. *Nature*, 585:357–362.

Hochgesang, J.A. (2014). Using Design Principles to Consider Representation of the Hand in Some Notation Systems. *Sign Language Studies*, 14(4):488–542.

Hochgesang, J.A. (2022). SLAASh ID Glossing Principles, ASL Signbank and Annotation Conventions, Version 3.2. https://doi.org/10.6084/m9.figshare.12003732.v4.

Hochgesang, J.A., Crasborn, O., Lillo-Martin, D. (2018). Building the ASL Signbank: Lemmatization Principles for ASL. In N. Calzolari, et al., (Eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 69–74. Miyazaki, Japan. European Language Resources Association (ELRA).

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

Johnston, T. and Schembri, A. (2007). Testing language description through language documentation, archiving and corpus creation: The case of indicating verbs in the Auslan Archive Corpus. In P. K. Austin, et al., (Eds.), *Proceedings of Conference on Language Documentation and Linguistic Theory*, 145–154. London: SOAS.

Kimmelman, V., Imashev, A., Mukushev, M., Sandygulova, A. (2020). Eyebrow position in grammatical and emotional expressions in Kazakh-Russian Sign Language: A quantitative study. *PLoS ONE*, 15(6):e0233731.

Kimmelman, V., Komarova, A., Luchkova, L., Vinogradova, V., and Alekseeva, O. (2022). Exploring Networks of Lexical Variation in Russian Sign Language. *Frontiers in Psychology*, 12:740734.

König, S. and Langer, G.. (2009). Signs fiction? Ein Wörterbuch DGS – Deutsch wird entwickelt. *Das Zeichen*, 81:82–89.

McKinney, W. (2010). Data structures for statistical computing in Python. In S. van der Walt and J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference*, 56–61. Austin, TX: SciPy 2010.

Parkhurst, S. and Parkhurst, D. (2003). Lexical Comparison of Signed Languages and the Effects of Iconicity. *Work Papers of the Summer Institute of Linguistics, North Dakota Session*, 47:1–17.

Power, J.M., Grimm, G.W., and List, J.-M. (2020). Evolutionary dynamics in the dispersal of sign languages. *Royal Society Open Science*, 7:191100.

Prillwitz, S., Hanke, T., König, S., Konrad, R., Langer, G., and Schwarz, A. (2008). DGS Corpus Project – Development of a Corpus Based Electronic Dictionary German Sign Language / German. In O. Crasborn, et al. (Eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, 159–164. Marrakesh, Morocco. European Language Resources Association (ELRA).

Russell, K., Wilkinson, E. and Janzen, T. (2011). ASL sign lowering as undershoot: A corpus study. *Laboratory Phonology*, 2(2):403–422.

Schembri, A., Fenlon, J., Rentelis, R., Reynolds, S., and Cormier, K. (2013). Building the British Sign Language Corpus. *Language Documentation & Conservation*, 7:136–154.

Schwager, W. and Zeshan, U. (2008). Word classes in sign languages: Criteria and classifications. *Studies in Language* 32(3):509–545.

Smith, R., (Ed.). (2013). HamNoSys 4.0 User Guide. Institute of Technology Blanchardstown, Ireland. https://robertsmithresearch.files.wordpress.com/2012/10/hamnosys-user-guide-rs-draft-v3-0.pdf.

Stumpf, M.R., Pizzio, A.L., Lucinda, J.O., Quadros, R.M. de, and Crasborn, O. (2020). SignBank da Libras. *Fórum Linguístico*, 17(4):5475–5487.

Van Rossum, G. and Drake, F.L. (2009). *Python 3 reference manual*. Scotts Valley, CA: CreateSpace.

Virtanen, P., et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3):261–272.

Yu, S., Geraci, C., and Abner, N. (2018). Sign Languages and the Online World Online Dictionaries & Lexicostatistics. In N. Calzolari, et al., (Eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 4235–4240. Miyazaki, Japan. European Language Resources Association (ELRA).

8. Language Resource References

Crasborn, O., et al., (Eds.). (2020). *Global Signbank*. Radboud University, Nijmegen. signbank.science.ru.nl.

Hochgesang, J.A., Crasborn, O., and Lillo-Martin, D. (2020). ASL Signbank. New Haven, CT: Haskins Lab, Yale University. https://aslsignbank.haskins.yale.edu/.

Vlaams GebarentaalCentrum. (2018). VGT Signbank. Gent, Belgium. https://vlaamsegebarentaal.be/signbank.