Assessing an automated tool to quantify variation in movement and location: a case study of American Sign Language and Ghanaian Sign Language

Fragkiadakis, M.

Assessing an Automated Tool to Quantify Variation in Movement and Location

Author(s): MANOLIS FRAGKIADAKIS

Source: *Sign Language Studies*, Fall 2022, Vol. 23, No. 1 (Fall 2022), pp. 98-126

Published by: Gallaudet University Press

Stable URL: https://www.jstor.org/stable/10.2307/27186925

MANOLIS FRAGKIADAKIS

Assessing an Automated Tool to Quantify Variation in Movement and Location: A Case Study of American Sign Language and Ghanaian Sign Language

ABSTRACT

Signs in sign languages have been mainly analyzed as composed of three formational elements: hand configuration, location, and movement. Researchers compare and contrast lexical differences and similarities among different signs and languages based on these formal elements. Such measurement requires extensive manual annotation of each feature based on a predefined process and can be time consuming because it is based on abstract representations that usually do not take into account the individual traits of different signers. This study showcases a newly developed tool named DistSign, used here to measure and visualize variation based on the wrist trajectory in the lexica of two sign languages, namely American Sign Language (ASL) and Ghanaian Sign Language (GSL), which are assumed to be historically related (Edward 2014). The tool utilizes the pretrained pose estimation framework OpenPose to track the body joints of different signers. Subsequently, the Dynamic Time Warping (DTW) algorithm, which measures the similarity between two temporal sequences, is used to quantify variation in the paths of the dominant hand's wrist across signs. This enables one to efficiently identify cog-

Manolis Fragkiadakis is a PhD candidate at the Leiden University Centre for Linguistics with a research focus on sign language recognition.

nates across languages, as well as false cognates. The results show that the DistSign tool can recognize cognates with a 60 percent accuracy, using a semiautomated method that utilizes the Levenshtein distance metric as a baseline.

To date, there are more than 200 sign languages documented in the world (Hammarström et al. 2021). However, little is known about the possible genealogical connections between different sign languages, and many assertions on this topic are based purely on historical records of language contact (Börstell, Crasborn, and Whynot 2020). As Abner et al. (2020, 18) point out, "[S]cant research exists on historical change and historical relations among sign languages." Moreover, there has not been a methodical and large-scale study concerned with families of sign languages yet.

Typically, historical-comparative research on sign languages uses lexical comparison and lexicostatistics to measure similarity in parameter values among various sign languages (Woodward 1991; Woodward 1993; Johnston 2003; Bickford 2005; McKee and Kennedy 2000). As Börstell, Crasborn, and Whynot (2020) highlight, these studies measure the amount of lexical overlap on a predefined list of concepts to account for the probability of two languages being related. Lexical similarity is often accounted to form similarity between signs that share the same meaning (Börstell, Crasborn, and Whynot 2020). Each sign is annotated based on its four (sometimes three) basic form parameters: location, handshape, orientation, and movement. If all four parameters between a set of signs are exactly the same, the signs are counted as "identical." Two matching parameters result in "similar" signs; otherwise, they are considered different forms (Börstell, Crasborn, and Whynot 2020). However, there is inconsistency in how scoring criteria are used to account for sign similarity (Parks 2011). The result of this approach, as discussed by Power, Quinto-Pozos, and Law (2021), is that sign languages with an associated history yield a higher percentage of similar signs, but sign languages that do not share a related history might do so too.

Admittedly, the use of lexicostatistics in sign languages raises several issues. As highlighted by Parks (2011), many studies do not share a common set of similarity criteria. As a result, by virtue of using different benchmarks, each comparison may lead to a different result. Moreover, as sign languages do not share an official transcription system, different studies have used disparate annotation systems to encode the sign form parameters. Some of them used SignWriting (Sutton 2009) and HamNoSys (Hanke 2004) while others have used independently developed annotation systems (Abner et al. 2020; Börstell, Crasborn, and Whynot 2020; Yu, Geraci, and Abner 2018). Consequently, it is not possible to directly compare signs encoded in different formats without translating one transcription system to another. Although partial overlap between some of these annotation systems exists, this is not always the case (Power, Quinto-Pozos, and Law 2021). In addition, studies often differ in the total number of signs compared, which signs were compared, and how lexical categories were defined. For example, Yu, Geraci, and Abner (2018) compared 100 signs in twenty-three sign languages while Börstell, Crasborn, and Whynot (2020) compared 301 items in three sign languages.

Furthermore, to measure similarity between sign languages, one has to manually annotate the form parameters for each sign. Such a task can be time consuming and requires a predefined set of values to be used. Some systems, such as the one presented by Yu, Geraci, and Abner (2018), separated the form parameters into more thorough subfeatures. For example, the parameter of movement was expanded into features such as movement direction, movement shape, and movement repetition. Nevertheless, regardless of the detail these features encode, they are still abstract representations of the actual characteristics they represent. Slight differences in the way signs are articulated cannot always be encoded by the transcription systems used. In addition, many signs tend to be annotated inconsistently, resulting in discrepancies in their transcriptions, even though there might be no real difference of articulation present in the video.

This study presents a newly developed tool, named DistSign, to explore variation measurement and similarity calculation between sign languages. By eliminating the need for manual transcription, the tool can work toward an automated analysis of lexical comparison.

The OpenPose pose estimation framework (Cao et al. 2017) was used to extract the location and movement of the dominant and nondominant hands' wrists. Subsequently, using the DTW algorithm, the tool quantified variation in the movement parameter. DTW is designed to compare sequences that evolve through time. Hence, it was used by the tool to compare the trajectories produced by the dominant hand's wrist. Both OpenPose and DTW are further explained on pages 102–4.

There are three reasons why this study used the wrist trajectory as the main feature to assess lexical similarity. First and foremost, as Napoli and Sanders (2022, 4) discuss, "[T]he movement parameter has been reported by signers to be the most salient parameter for recognizing signs." Second, the accuracy with which OpenPose predicts the locations of the fingers in a video can vary widely. As a result, a possible handshape recognition process would be vastly skewed by the mis-prediction of the fingers' locations. Therefore, using these locations without an additional processing step (and possibly a dedicated handshape prediction step), was found not to improve sign recognition accuracy in an earlier study using OpenPose (Fragkiadakis, Nyst, and van der Putten 2020). Finally, in the present study, location is not predicted as a separate value, since the predicted feature, that is, the wrist trajectory, includes both the location and the movement parameters.

Overall, this study explores three applications for the DistSign tool: visualization of the distribution of the wrist trajectory, assessment of whether certain lexical fields exhibit more variation than others, and identification of potential cognates.

To verify the applicability of the DistSign tool, a comparison of two sign languages that share a common history, namely ASL and GSL (Edward 2014), was performed. Using the word list described by Parks (2011), I measured the lexical similarity between these sign languages by looking at signs in ten lexical fields. I examined the results using as a baseline a semiautomated method on manually transcribed data utilizing the Levenshtein-based distance metric (further explained on pages 106–7).

Methods and Data

Data

The initial word list used in this study contained the 241 items utilized in Parks's (2011) study, resembling the word lists by Bickford (1991) and Swadesh (1955). These lexical items belong to various semantic domains and grammatical word classes such as animals, food, physical activities, and adjectives. ASL Signbank (Hochgesang, Crasborn, and Lillo-Martin 2018) and the GSL dictionary app (HANDS!Lab 2020) were used to collect the equivalent video sign entries. However, not all the items in my original list could be found in both lexica and, as a result, I compared 141 signs corresponding to ninety-four different concepts. The additional forty-seven entries were variants of some of the signs.

Pose Estimation

Using the video entries in their raw format would be time and computationally consuming. Moreover, videos contain information that is not always useful for the purpose of this type of analysis, such as the background of the signer. As a result, different techniques have been developed to extract only the information that is relevant for further computations. A popular technique employed is the use of a pose estimation framework to extract the locations of the body, face, and fingers joints. In this study, the pose estimation framework named OpenPose was used. OpenPose is an open source for academic purposes, a real-time 2D pose estimation framework that can detect body, hand, and facial key points. It is a highly accurate framework that has been extensively used in the sign language and gestural domains (Fragkiadakis, Nyst, and van der Putten 2020; Östling, Börstell, and Courtaux 2018; Liang et al. 2019). Its ability to run under different operating systems and architectures as well as its options for visualization and output file generation makes it an ideal tool to process sign language and gestural videos. Its output consists of multiple files, each containing the pixel x, y coordinates of the body (shoulder, elbow, wrist, etc.), hand, and face joints per frame. In a recent study conducted by Liang, et al. (2019), the authors concluded that OpenPose is an accurate and robust framework when tracking the hand trajectories.

In the present study, each video sample for each of the two sign languages compared was processed with OpenPose. As described earlier, outputs consist of a series of files (one for each frame of the video) that contain the x and y pixel coordinates for each predicted body joint as well as a confidence interval referring to how certain that prediction is. Furthermore, an additional cleaning and normalization process was applied by first removing all the predicted joints with low confidence (<0.2). Moreover, the preparation phase during the realization of each sign was removed by applying a threshold. If the height (y pixel coordinate) of the dominant hand's wrist at the initial or ending points was lower than this threshold, the coordinate was removed. This step ensures that the first and last wrist locations of the first sign sequence being compared will match the first and last wrist locations of the second sign sequence being compared, which is required for the DTW algorithm.

As there was variation in the way people were aligned with respect to the video frame, a normalization process was also applied. The scale normalization method is based on previous studies by Fragkiadakis, Nyst, and van der Putten (2020), Celebi et al. (2013), Schneider et al. (2019), and Östling, Börstell, and Courtaux (2018), and it ensures that the coordinates generated by OpenPose are all roughly in the same position with respect to the video frame no matter how far, close, or to the left or to the right each signer appears in the video. Furthermore, an additional step was introduced to account for left-handed signers. When the average velocity of the left hand was greater than that of the right hand, a horizontal flip on the video was applied. This step ensured that comparisons would always involve the dominant hand.

Similarity Measure

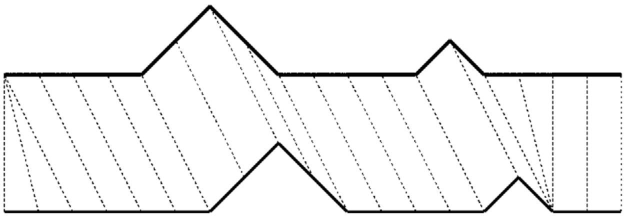
As mentioned, the main goal of this study is to explore how lexical similarity or distance between sign languages can be measured in a quantifiable and automated way. Such functionality should be applicable to any digital sign language dictionary or word list regardless of its size and quality. While deep learning approaches have demonstrated accurate results in classifying and predicting signs from video material (Li et al. 2020), they often require a vast amount of training data. Moreover, such architectures are often trained in specific languages, with the need to be re-trained so as to use another language.

The proposed tool, by virtue of using the DTW algorithm, does not require any training whatsoever and can be applied to any sign language dictionary and word list. Finally, to ensure that its functionality can be applied in a user-friendly and inclusive manner, a Python Jupyter Notebook on Google's Colab environment (Bisong 2019) has been created.[1] In principle, such a step can ensure that anyone with an internet connection can use the tool no matter the computational power at hand.

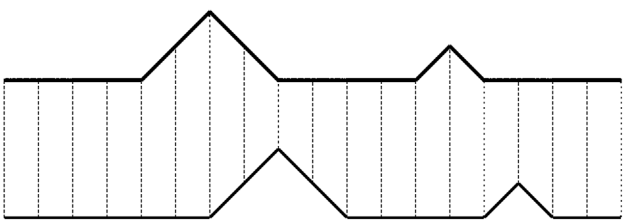
Dynamic Time Warping. DTW is a dynamic programming algorithm designed to compare time series data. It can be used as a distance metric between two series that may vary in speed. It has been widely used in the speech recognition domain (Abdulla, Chow, and Sin 2003; Axelrod and Maison 2004; Myers, Rabiner, and Rosenberg 1980) and is a reliable solution when limited training data are available, as it requires no training whatsoever.

Assume that we have two sequences that evolve through time, similarly to the ones in figure 1. In this study, these could be the wrist paths of two signs being compared. While the black and the gray line sequences in figure 1 look almost identical, the gray one is slightly shifted to the right. Such a pattern is quite common when comparing sequences that move in time, such as signs. If we were to compare these sequences using the Euclidean distance (figure 1b) by simply trying to overlap them, that is, by connecting each of the points of the first and the second sequences and measuring the distance between them, we would end up with a poor match, because the peaks of the sequences are not temporarily aligned. DTW allows to correct that behavior by dynamically warping the path, which allows us to compare each point of a sequence with a point that is most likely a better fit (as in figure 1a). In figure 1a, the peaks are aligned, yielding a good match. DTW will assign 0 if the two sequences or paths are identical. The higher the index, the more different the paths are.

In the gestural and sign language domains, DTW has been extensively used for classification and recognition tasks to match a gesture or a sign against an existing data set. (Ten Holt, Reinders, and Hendriks 2007; Jangyodsuk, Conly, and Athitsos 2014). Recently, DTW combined with the output of OpenPose has been used for the same kind of sign recognition tasks (Ripperda, 2020) using simple RGB camera material (Schneider et al. 2019). Fragkiadakis, Nyst, and



(a) Dynamic Time Warping



(b) Euclidean distance

FIGURE 1. Example of how two time series can be compared using DTW (a) and (b) the Euclidean distance.

van der Putten (2020) compared how different configurations of the OpenPose output can affect prediction accuracy when a query sign is compared to 100 signs in a database. The goal of their study was to develop a reverse search functionality for sign language dictionaries by allowing a user to sign in front of a simple web camera and find possible matching signs. Overall, three different parameter configurations were compared: only the dominant hand's wrist, the trajectories of the fingers, and a combination of these two. Their results showed that using the trajectories of the fingers as well as merging them with the trajectory of the dominant hand's wrist did not yield compelling results. Hence, using just the data from the dominant hand's wrist is sufficiently adequate when it comes to matching sign retrieval.

A possible explanation for this could be that OpenPose does not predict the finger joints accurately and consistently when the elbow is not visibly present and when the source is a very low-quality video. In a later study, Fragkiadakis and van der Putten (2021) have further verified the applicability of DTW in a larger lexicon. They experimented with different body joint combinations (upper body, dominant hand's arm and wrist) to a greater extent. They concluded that DTW performs equally well irrespectively of the body joint configurations selected. Hence, in the present study, only the trajectory of the dominant hand's wrist was used in order to visualize and quantify variation among signs based on the location and movement parameters. To develop such functionality, the Python package created by Tormene et al. (2009) and Giorgino (2009) was utilized.

Cognates. Two signs that share the same form and meaning are considered cognates, or true-friends. In addition, the greater the number of shared cognates, the higher the mutual intelligibility in a cross-signing context is and vice versa (Börstell, Crasborn, and Whynot 2020). An automated way to find cognates in a set of sign languages can potentially be used to efficiently and quickly predict communicative success in such contexts. For instance, Omardeen (2018) demonstrated that the degree of form overlap between a native and a foreign sign, as evaluated by the amount of similar phonological characteristics, can predict a signer's iconicity rating for a foreign sign.

The process developed in this study in order to automatically find cognates between two sign languages is the following: First, I disregarded meaning and let DTW compare each sign from one language to all the signs of the other language. Subsequently, I sorted the results based on the output of this process that indicates how similar two signs are (similarity index). Finally, if the most similar match had the same gloss as the query sign, then the two signs were counted as cognates.

In a preliminary experiment, it was noticed that, on average, the similarity index of the first four most similar signs varied within the range of the second decimal, while the index would increase significantly beyond the fourth most similar sign. When comparing one sign in one language to all the other signs in another language, it is possible that the first four or five most similar matches might vary only minimally in their degree of similarity to the query sign. As a result, if one of the four most similar matches also shared the same meaning as the query sign, I considered it a cognate to the query sign.

Manual Annotation. I evaluated the performance of the automated process by using a semiautomated method that resembles the one described by Börstell, Crasborn, and Whynot (2020) as a benchmark. For the semiautomated process, each sign (and its variants) was manually transcribed in terms of its initial and final handshape as well as the initial and final location parameter, following the Global Signbank's manual (Crasborn et al. 2020). This manual explains how to encode a sign in terms of its phonological features in order to be submitted to the lexical database of Global Signbank. It contains possible encodings for handedness, handshape, location, movement direction, orientation, and the like. Overall, I annotated ninety-six possible values for the handshape parameter. For instance, the handshapes that were similar to the ASL handshape "B" were annotated with that value (i.e., "B"). Variations in the fingers' articulation would result in a different annotation value such as "B_bent." Location was categorized in four groups (head, body, extremities, and neutral space) with multiple values per group (cheek, chin, eye, for the head etc.). Subsequently, using a process that derives from the Levenshtein distance metric that is used to classify spoken languages, I calculated the similarity index between all possible sets of signs. This process calculates whether each parameter that is compared between two signs is different (0) or similar (1) and divides the total score by the number of parameters compared (in this study: 4). If all parameters are the same, then the two signs have a similarity index of 1, and if they also share the same meaning-gloss then they are counted as identical and thus cognates. Finally, this process was compiled into a Python Jupyter Notebook to allow other researchers to quickly and reproducibly calculate the Levenshtein-based distances between sets of sign languages.[2]

Lexical Fields. Lexical variation occurs when two or more forms appear in the same conditions without a change in meaning. For example, in GSL, food terms tend to display considerable variation (Abudu 2019). This, according to Abudu (2019), stems from sociocultural differences and the situation in which the sign is realized.

In order to explore how much GSL vocabulary diverges from ASL vocabulary, several lexical fields were examined. In total, twelve categories were initially considered, based on Parks's (2011) word list. However, for two of the categories, most concepts were missing from the sign databases that I used. Thus only ten fields were compared in this study.

I compared all the ASL signs in each selected field to the corresponding GSL signs by calculating their similarity indexes based on the trajectory of the dominant hand. The signs were then sorted from most similar to most dissimilar. This process allows us to infer which GSL variant might have been derived (thus having low distance) from the equivalent ASL sign.

Additionally, the average similarity rate per field was computed. The purpose of this analysis was to test the hypothesis that some fields in GSL may have undergone more change from ASL than other semantic domains.

Finally, I calculated the average similarity rate among the signs of each field within each language. High distance would mean that the wrist trajectories are quite different among the signs (high intra-field variation), while low distance would mean that the trajectories follow a more similar pattern (low intra-field variation). This step could be used as a proxy to assess the crosslinguistic distance calculated in the previous step. It can provide additional information as to whether low or high distance is a result of degree of similarity in the movement parameter or just an outcome of intra-field variation. The latter could be caused by signs in a particular field exhibiting strong iconicity. This investigation, however, goes beyond the scope of this study.

Distribution and Visualization of the Wrist Trajectory

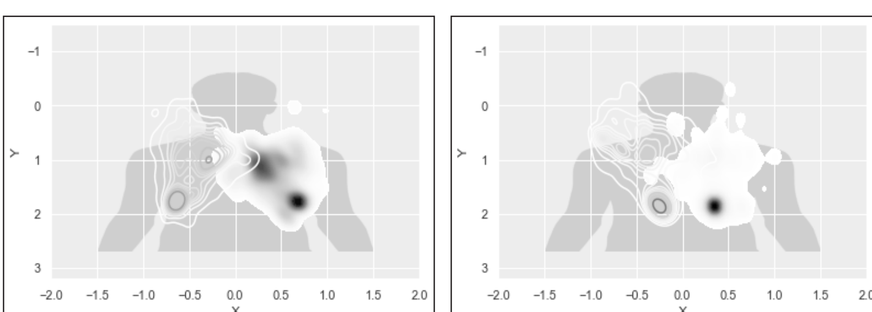
To explore lexical differences between the two languages within each field, visualizations of the average wrist trajectory parameter per semantic field were generated.

By projecting the locations of the wrists for all the signs in both lexica and within each semantic field, the DistSign tool was able to visualize the average location. The normalized coordinates of the wrist trajectory per language and semantic field were projected on a grid on top of a silhouette used as a reference point (see figure 2). Filled contours indicate the locations of the nondominant hand's wrist only. The darker a spot is, the more signs had wrist movement in that area. Signing speed was omitted for this analysis, as it is not essential for the DTW algorithm as described previously. This process is more or less identical to the one described by Östling, Börstell, and Courtaux (2018), who explored iconic patters of sign locations in a crosslinguistic study. In their study, sign locations were calculated based on the hand position, which was extrapolated by adding half the distance between the elbow and the hand to the wrist location in a straight line. However, their method did not take into account any possible wrist flexions, which could result in lightly altered locations. Hence, in the present study, only the dominant hand's wrist was used to extrapolate the location parameter, as the position of the wrist is a direct outcome of the OpenPose framework. This means, however, that most outputs will be shifted compared to the actual hand location. This shift will be mostly downward or to the center.

Wrist Locations

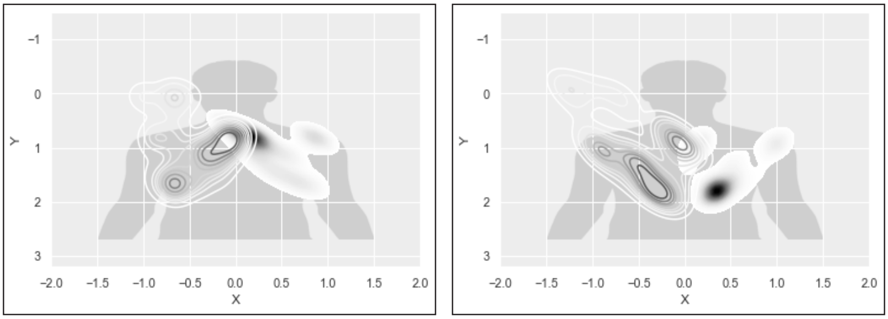
Figure 2 presents the average wrist location for all the signs in the ASL and GSL word lists. The ASL data have more wrist movement with the nondominant hand, as seen by the two dark spots. Moreover, the average wrist location is less central in the torso compared to the GSL wrist locations.

Figures 3 and 4 present wrist location visualizations for the semantic fields of emotions and food, respectively. It is noticeable that for the signs within the domain of emotions, ASL wrist locations concentrate toward the neck, while the equivalent GSL signs are concentrated in the mid torso. Considering that these heat maps reflect



(a) ASL (b) GSL

FIGURE 2. Visualization of the wrist location parameter for all signs in (a) the ASL and (b) GSL lexica. Concentrated wrist movement in an area is shown with darker color. Transparent contours indicate the locations of the dominant hand's wrist, and filled contours indicate the locations of the nondominant hand's wrist.

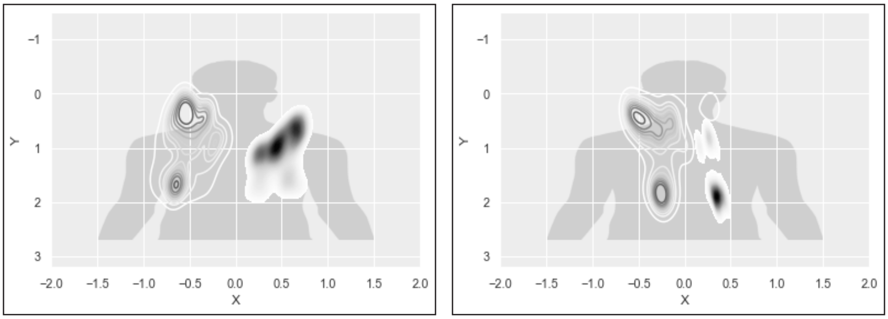


(a) ASL (b) GSL

Figure 3. Visualization of the wrist location parameter for the semantic field of emotions. Concentrated wrist movement in an area is shown with darker color. Transparent contours ◎ indicate the locations of the dominant hand's wrist, and filled contours indicate the locations of the nondominant hand's wrist.

the locations of the wrists, not the hands, one can extrapolate that the actual average hand location for the ASL signs is at the general head area, while for GSL is at the torso and neck areas.

By contrast, signs for food (figure 4) have relatively similar locations for the dominant hand's wrist movement in ASL and GSL, although GSL shows less movements, on average, for the nondominant hand feature.



(a) ASL (b) GSL

Figure 4. Visualization of the wrist location parameter for the semantic field of food. Concentrated wrist movement in an area is shown with darker color. Transparent contours ◎ indicate the locations of the dominant hand's wrist, and filled contours indicate the locations of the nondominant hand's wrist.

Distances Per Lexical Field. As previously mentioned, each GSL and ASL sign was compared for each lexical field. Table 1 presents average crosslanguage distances, as calculated by the DistSign tool using the DTW algorithm, and based on the paths produced by the dominant hand's wrist. The lowest distances are found in the fields of time and food. The highest distances are found in the categories of adjectives,

Table 1. Mean distance and standard deviation per lexical field between the ASL and GSL signs

Lexical Field	Signs	Mean Distance	Standard Deviation
Time	AFTERNOON, DAY, MONDAY, MORNING, NIGHT, SATURDAY, THURSDAY, TUESDAY, WEDNESDAY	9.4	4.1
Food	APPLE, BANANA, CARROT, CORN, GRAPES, ONION, RICE, TOMATO	9.5	4.4
Physical Activities	ASK, BUY, COOK, CATER, COUNT, DANCE, DIE OR DEAD, KILL, LAUGH, LIVE, MEET, PAY, RUN, SELL, SIT, SLEEP, STAND	10.3	5.9
Environment	LEAF, MOON, MOUNTAIN, RIVER, SALT, STAR, TREE, WATER, WIND, WOOD	10.4	5.6
Family and Human	BOY, BROTHER, DAUGHTER, DEAF, HUSBAND, KING, MAN, MOTHER, NAME, SISTER, SON, WIFE, WOMAN	10.4	3.1
Clothing and Household Items	BED, DOOR, HOUSE, KNIFE, SHIRT, SHOE, WINDOW	10.5	3.5
Animal	CAT, CHICKEN, DOG, ELEPHANT, HORSE, LION, MOUSE, RABBIT, SNAKE, SPIDER	11.7	5.1
Adjectives	BAD, BEAUTIFUL, CLEAN, DIRTY, HUNGRY, NEW, OLD, POOR, RICH, STRONG, SWEET, UGLY	13.1	9.8
Work and Occupation	DOCTOR, POLICEMAN, SOLDIER, TEACHER	14.2	5.9
Emotions	ANGRY, DREAM, LOVE, THANK YOU, WELCOME	14.5	6.8

The lower the mean distance, the more similar the signs in that field are in terms of wrist trajectory.

work and occupation, and emotions. Each field contains those signs described in Parks's (2011) study that I could find in both lexica.

Additionally, the standard deviation was calculated per lexical field. Standard deviation signifies how dispersed the data are in relation to the mean average distance described in the previous step. Low standard deviation means that the data are clustered around the mean, and high standard deviation indicates that the data are more spread out.

Table 2 presents the intra-field variation per sign language, which was calculated by comparing each sign to all the other signs in each field and within each language. As can be seen in table 2, the ASL semantic fields with the lowest average distances are time, physical activities, and food. Similarly, the GSL semantic fields with the lowest average distances are time, food and clothing and household items. On average, intra-field variation in GSL is much higher than that in ASL. This means that the GSL signs for each semantic field have less homogeneous wrist trajectories than in ASL. Furthermore, we can observe that the semantic fields that had lower crosslinguistic variation (food, time, and physical activities), as seen in table 1, were the ones that also had lower intra-field variation (table 2).

Table 2. Intra-field variation per sign language

Lexical Field	ASL Variation	GSL Variation
Adjective	12.8	16.6
Animal	12.3	23.3
Emotions	13.4	20.6
Environment	10.3	20.3
Family and Human	9.4	16.3
Food	9.1	12.7
Physical Activities	8.8	18.1
Time	7.8	15.2
Work and Occupation	15	15.8
Clothing and Household Items	10.8	15.7

The higher the variation the less homogeneous the wrists trajectories in that field.

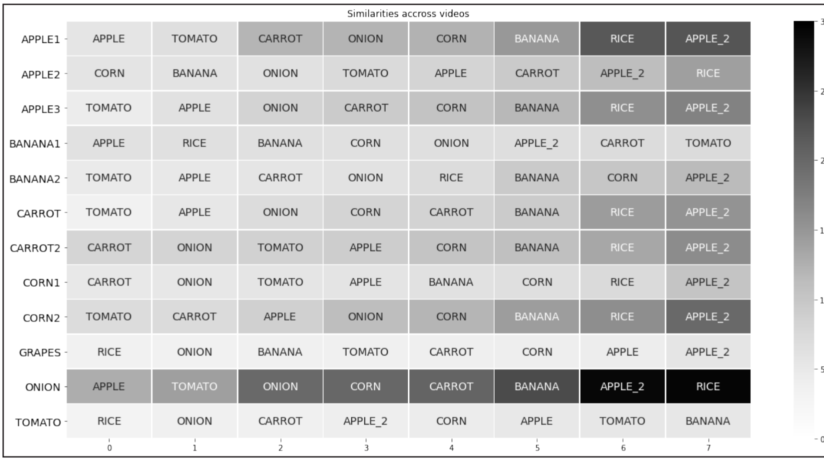


FIGURE 5. Heat map representing the calculated distances for the lexical field of food. Transcripts (indexes) on the first column represent the ASL signs, and each row within the heat map represents the GSL signs that are compared to each ASL variant. The darker the color, the higher the distance.

Aside from calculating mean crosslanguage distances per field, the tool also measured the similarity index for each sign compared against all the signs of the other language for each category. Figure 5 shows how the tool outputs this comparison in the form of a heat map for the lexical field of food. For every ASL variant, marked by a different index (e.g., APPLE1, APPLE2, etc.), the tool quantifies the similarity between that sign and all the GSL signs in that specific field and sorts them from the most to the least similar. This allows us to observe how each sign from one sign language relates to all the other signs from the other language by field. An important output of this process is the automatic identification of which lexical variants are identical and which are similar. For example, in the food category, of all the ASL variants for APPLE, APPLE1 is the most similar to the equivalent GSL sign (figure 5).

Finally, the heat map indicates which signs had a completely different trajectory (and location) than the query sign, resulting in a higher distance index. For example, figure 6 shows that the only GSL signs that are similar in wrist trajectory to the ASL variants for DREAM are the

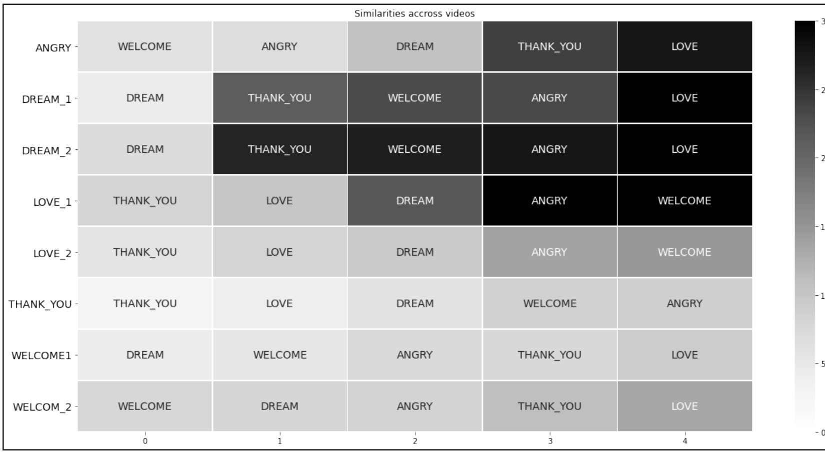


Figure 6. Heat map representing the calculated distances for the semantic field of emotions. Transcripts (indexes) on the first column represent the ASL signs, and each row within the heat map represents the GSL signs that are compared to each ASL variant. The darker the color, the higher the distance.

ones that convey the same concept, while all the other signs within the same lexical category have completely different movements (as indicated by the darker color, i.e., higher variation). On the other hand, some signs with different meaning are identified as similar in form, for example, the ASL sign love_2 and the GSL sign thank_you (figure 7).

(a) The ASL sign love_2[3] (b) The GSL sign thank_you

Figure 7. (a) The ASL sign love_2 and (b) GSL sign thank_you as identified by the DistSign tool for having similar movement of the dominant hand's wrist.

Cognates

This section presents the results of the cognate identification task. Figure 8 displays the glosses for the sign pairs that were identified as cognates. Each line represents whether the ASL/GSL sign pairs corresponding to that gloss were identified as cognates by the DistSign tool (darker color) or by the Levenshtein distance (lighter color). Overall, sixty-four sign pairs were identified as cognates by aggregating the results of both tools.

The DistSign tool identified thirty-nine signs as cognates. Twenty-eight of them were also recognized by the Levenshtein distance metric. However, there were twenty-five sign pairs that were not identified as

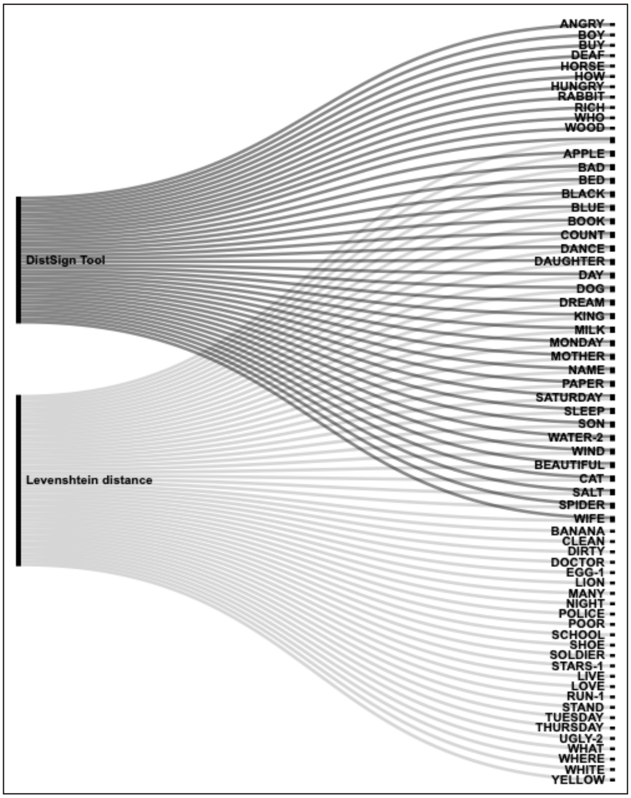


FIGURE 8. Sign pairs found to be cognates by the semiautomated method using the Levenshtein distance (light-gray) and the DistSign tool (dark-gray). The sign pairs identified solely by one technique are found at the edges.

Table 3. Disagreements between the DistSign tool and the semi-automated process using the Levenshtein distance on the recognition of cognates

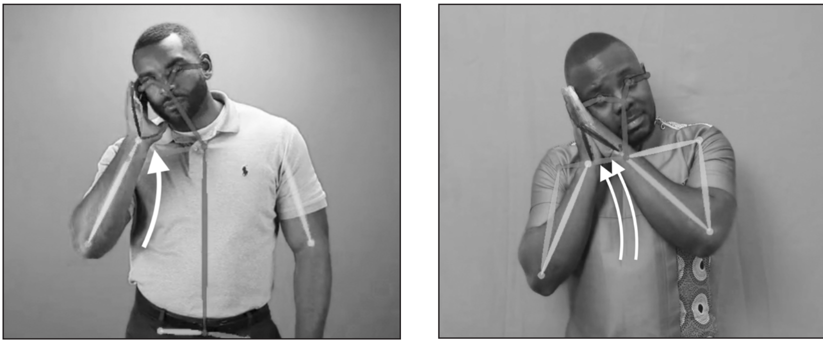
	Reason of Disagreement	Number of Signs
Cognates identified by the Semi-Automated Process only	Different final or initial wrist locations	7
	Different handshape transcription	4
	Total	11
Cognates identified by the DistSign Tool only	Movement repetitions	16
	OpenPose performance	9
	Total	25

cognates by the DistSign tool that were recognized by the Levenshtein distance metric. The accuracy rate of the DistSign tool regarding cognate identification was at around 60 percent, while the accuracy of the semiautomated method using the Levenshtein distance metric was at around 82 percent.

As reported, there were cases where a disagreement between the DistSign tool and the semiautomated process using the Levenshtein distance could be observed. Table 3 presents the number of cognates that were identified by only one tool or the other and the reasons the other tool failed to recognize them as such.

Four cognates were not identified by the semiautomated method using the Levenshtein distance due to the way I annotated the parameters. For example, figure 9 displays a pair of signs that were identified as cognates by the DistSign tool but not by the semiautomated method. The reason why the semiautomated method did not recognize them was that the ASL sign was coded with the "B_curved" final handshape, while the GSL sign was annotated with a "B" handshape. As a result, these two signs "matched" only in three out of the total four parameters required to be classified as cognates.

There were several other instances of disagreement between the DistSign tool and the semiautomated method. First and foremost, the DistSign tool recognizes a set of signs as cognates even if the final or initial wrist locations are slightly different as long as the trajectory is similar. Figure 10 displays the signs for ANGRY as an example of



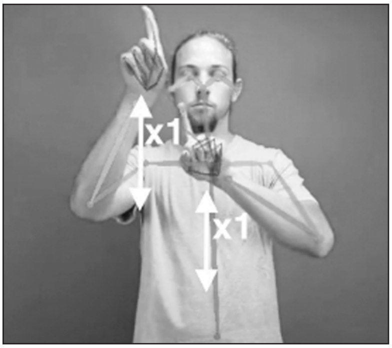
(a) The ASL sign BED (b) The GSL sign BED

FIGURE 9. (a) The ASL sign BED (https://aslsignbank.haskins.yale.edu/dictionary/gloss/384.html) and (b) GSL sign BED.

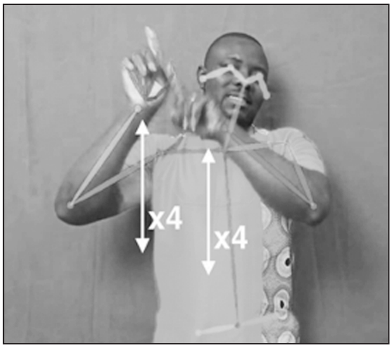
such a set of signs. Note that while the initial realization of the GSL sign is at the lower part of the torso, the overall wrist trajectory is identical to that of the ASL sign. By contrast, these two signs were not recognized as cognates using the semiautomated method with the Levenshtein distance as they match in only two (i.e., initial and final handshape) out of the four manually coded parameters. This example demonstrates one of the core functions of the DistSign tool, that is, the extrapolation of the location parameter based on the wrist trajectory parameter.

(a) The ASL sign ANGRY (b) The GSL sign ANGRY

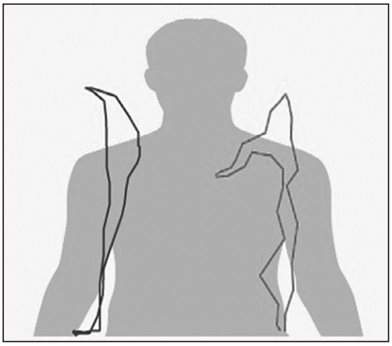
FIGURE 10. (a) The ASL sign ANGRY (https://aslsignbank.haskins.yale.edu/dictionary/gloss/15.html) and (b) GSL sign ANGRY.



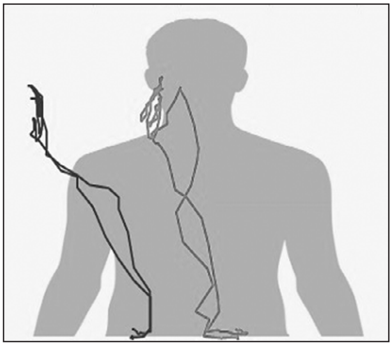
(a) The ASL sign STAR[4]



(b) The GSL sign STARS-1



(c) The trajectories of the wrists for the ASL sign STAR



(d) The trajectories of the wrists for the GSL sign STARS-1

FIGURE 11. The (a) ASL sign STAR and (b) GSL sign STARS-1 as well as the trajectories of the (c) dominant- and (d) nondominant hand wrists.

Most importantly, most of the signs that were identified as cognates using the semiautomated method but not using the DistSign tool were signs that included multiple repetitions of a specific movement. When the number of repetitions varies between a set of signs, the tool fails to correctly match the signs as possible cognates. Figure 11 shows the ASL and GSL signs for STAR(S), which exhibit such a pattern. In the ASL sign, the up and down movement of the dominant hand's wrist is only repeated once, as compared to the GSL sign where the movement is repeated four times. One explanation for this outcome may be that repetitions tend to create slight variations in the movement parameter, which leads DTW to assign a higher variation index. The longer and more complex a sign path is, the more probable it is that slight changes in movement will occur, thus resulting in a higher probability for the tool to fail. While repetitions, in some cases, can be attributed to a specific sign item or to a specific signer, they cannot be perceived as such by the tool.

In addition, it is worth mentioning that a small portion of the instances in which the DistSign tool failed to automatically identify cognates could be attributed to poor performance of OpenPose. Specifically, when signers crossed their arms, OpenPose failed to predict the location of the wrists, resulting in a distorted path.

Discussion

In this study, a method to quantify lexical similarities and differences among sign languages and sets of signs was introduced. By eliminating, or reducing, the need to manually annotate the form parameters, the tool described in this study can be used to automatically retrieve cognates across sign languages and to visualize the wrist location and trajectory parameters. I argue that the DTW algorithm is a well-suited method for quantifying variation among signs and sign languages. The movement and wrist location of the dominant hand, as predicted by OpenPose, have been used by several sign identification studies (Börstell, Crasborn, and Whynot 2020; Liang et al. 2019). However, to the best of my knowledge, this is the first time they have been used for measuring variation in a quantifiable way.

Using data from ASL and GSL word lists as a case study, I assessed the results of the automated tool using as a baseline a semiautomated method on manually transcribed data utilizing the Levenshtein-based distance metric. Overall, thirty-nine sign pairs out of sixty-four that could be potentially identified as cognates were correctly recognized as such by the tool, as compared to the semiautomated method, which recognized fifty-three. While the semiautomated method yielded higher accuracy, the automated tool did not involve any manual annotation, which significantly speeds up the process of cognate identification. Furthermore, it allows for a more large-scale comparison of sign language word lists.

Nevertheless, automated processes should always be validated with manually transcribed data. This is in agreement with the suggestion by Östling, Börstell, and Courtoux (2018) that researchers should manually annotate as much data as possible in order to validate automated methods, which, like in the case of this study, can also be error prone.

In general, an important advantage of the DistSign tool is its ability to serve as a toolkit to support sign language linguistics research. The core functionality of the tool is designed so that it works on a folder (and subfolder) level. As such, one has to structure the data accordingly in order to explore additional hypotheses. For example, one can measure the variation among different signers by separating the sign videos into different folders per signer. This makes it easy to design new experiments for nontechnical users.

There are limitations regarding the general applicability of the tool to videos of low quality. First and foremost, as described on pages 6 –7, the only normalization process applied is with regard to the position of the signer in the video frame. As a result, camera angle or rotational variance of the signer would produce less accurate results. In addition, the findings of this study are only transferable to videos of isolated signs, as opposed to multi-sign expressions, as these will be processed by the tool as one sign. In these cases, I would recommend manually splitting the videos into isolated signs or using the "manual activation classifier" tool described in Fragkiadakis, Nyst, and van der Putten (2021) prior to using the tool. On the other hand, longer utterances should not influence the performance, as DTW is specifically designed to compare time series that vary in speed.

In addition, a major limitation of using OpenPose on sign language video material is that it provides only information for the two-dimensional space. Thus, a substantial source of information is lost, as its output does not contain depth information such as how far a sign has been articulated from the signer's body. For instance, signs articulated in neutral space in front of the signer are confounded with signs articulated on the signer's chest or torso (Östling, Börstell, and Courtoux 2018).

Finally, due care must be exercised in the use of the Levenshtein-based distance metric when applied to sign language manual annotations. Given that the results based on the manual transcriptions were obtained using only four primary parameters (i.e., initial and final handshape, initial and final location), its general use should be further investigated. It was observed that subtle changes, especially in the location parameter, could not be fully rendered in the manual annotation. As a result, signs that exhibit minor but meaningful differences might be mislabeled as identical. This is in agreement with the observation by Greenhill (2011) that distance metrics, such as the Levenshtein distance, omit a large amount of information from the raw data. While a possible solution to this could be a more detailed annotation scheme, it would never reach the detail afforded by the automated method.

In other words, a disadvantage of using categorical phonological features is that many signs tend to be described inconsistently, resulting in differences in their transcriptions. To overcome that, some lexical comparison studies such as the ASJP database (Wichmann, Holman, and Brown 2020), use phonemic forms as a better approximation for the "raw" material for spoken languages. Nevertheless, the findings in this study highlight the usefulness of an automated process to minimize errors due to manual annotation inconsistencies.

Conclusions

In conclusion, this work is the first attempt to quantify lexical variation among sign languages based on the movement parameter using an automated tool. Applying the DTW algorithm on the trajectories produced by the dominant hand's wrist, as recognized by OpenPose, the developed tool compares the wrist trajectories of different signs and visualizes the average wrist location parameter.

Data from the ASL and GSL lexica were used as a case study to assess the automated process of the DistSign tool against a semiautomated method incorporating an adaptation of the Levenshtein distance metric as a baseline. The findings of this study indicate that such an automatic process can be used to compare the vocabulary of two sign languages either comprehensively or by semantic fields. Results show that the tool could accurately predict around 60 percent of sets of signs that had the same form and meaning out of an identified set.

However, this work clearly has some limitations. For example, cognates that contained multiple repetitions of a movement were not recognized as such by the DistSign tool. Furthermore, this study was specifically designed to work with videos of isolated signs and with signers in an upright position. Nevertheless, I believe that this study could be the starting point toward an accurate automatic process to measure similarities among sign languages. The validation of such process, as Östling, Börstell, and Courtoux (2018) suggest, can be further examined with a manually annotated portion of the data compared against the output of the automatic tool.

Future work will concentrate on expanding the functionality of the developed tool to also incorporate the movement of both the dominant and the nondominant hand. Moreover, additional work on reliable handshape recognition can potentially be adapted and embedded into the tool in order to raise its accuracy. On a wider level, research is also needed to determine the use of both the DistSign tool and the Levenshtein distance metric in different sign languages and data sets.

Acknowledgments

I would like to express my special thanks of gratitude to Timothy Hadjah, who provided the manual annotations for the data sets used in this study. I am also grateful for the insightful comments offered by the anonymous peer reviewers. The generosity and expertise of one and all have improved this study in innumerable ways and saved us from many errors; those that inevitably remain are entirely my own responsibility.

Notes

1 https://colab.research.google.com/drive/1bghRecy8jqLpl0NoXq64OalncoA2HmGZ?usp=sharing.

2. https://github.com/ManolisFrag/sign-language-levenshtein-distance.

3. https://aslsignbank.haskins.yale.edu/dictionary/gloss/179.html.

4. https://aslsignbank.haskins.yale.edu/dictionary/gloss/274.html.

References

Abdulla, W. H., D. Chow, and G. Sin. 2003. Cross-Words Reference Template for DTW-Based Speech Recognition Systems. In *TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region,* 1576–79. New York: Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/tencon.2003.1273186.

Abner, N., C. Geraci, S. Yu, J. Lettieri, J. Mertz, and A. Salgat. 2020. Getting the Upper Hand on Sign Language Families: Historical Analysis and Annotation Methods. *FEAST. Formal and Experimental Advances in Sign Language Theory* 3:17–29.

Abudu, F. T. 2019. Lexical Variation in the Ghanaian Sign Language and Its Effects on the Language Development of the Hearing Impaired. Master's thesis, University of Education, Winneba.

Axelrod, S., and B. Maison. 2004. Combination of Hidden Markov Models with Dynamic Time Warping for Speech Recognition. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing,* 1–173. New York: Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/ICASSP.2004.1325950.

Bickford, J. A. 1991. Lexical Variation in Mexican Sign Language. *Sign Language Studies* 72 (1): 241–76.

———. 2005. *The Signed Languages of Eastern Europe*. Dallas: SIL International.

Bisong, E. 2019. Google Colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners,* 59–64. Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-4470-8_7.

Börstell, C., O. A. Crasborn, and L. Whynot. 2020. Measuring Lexical Similarity across Sign Languages in Global Signbank. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives,* 21–26. Paris, France: European Language Resources Association (ELRA).

Cao, Z., T. Simon, S. E. Wei, and Y. Sheikh. 2017. Realtime Multi-Person 2d Pose Estimation Using Part Affinity Fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017,* 1302–10. New York: Institute of Electrical and Electronics Engineers (IEEE).

Celebi, S., A. S. Aydin, T. T. Temiz, and T. Arici. 2013. Gesture Recognition Using Skeleton Data with Weighted Dynamic Time Warping. In *Proceedings of the 8th International Joint Conference on Computer Vision, Imaging, and Computer Graphics Theory and Applications (VISAPP'13*), ed. S. Battiato and J. Braz, 620–25. Setúbal, Portugal: SciTePress-Science and and Technology Publications. https://doi.org/10.5220/0004217606200625.

Clark, J., C. Yallop, and J. Fletcher. 2007. *An Introduction to Phonetics and Phonology.* Hoboken, NJ: Wiley-Blackwell.

Crasborn, O., I. Zwitserlood, E. Van Der Kooij, and E. Ormel. 2020. *Global SignBank Manual, Version 2.* Radboud University. https://doi.org/10.13140/RG.2.2.16205.67045/1.

Edward, M., 2014. The Phonology and the Morphology of the Ghanaian Sign Language. Presented at *ATELIER International Conference.* Abidjan, Cote d'lvoire.

Fragkiadakis, M., V. A. S. Nyst, and P. van der Putten. 2020. Signing as Input for a Dictionary Query: Matching Signs Based on Joint Positions of the Dominant Hand. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives,* 69–74. Paris, France: European Language Resources Association (ELRA).

Fragkiadakis, M., V. A. S. Nyst, and P. van der Putten. 2021. Towards a User-Friendly Tool for Automated Sign Annotation: Identification and Annotation of Time Slots, Number of Hands, and Handshape. *Digital Humanities Quarterly (DHQ)* 15 (1). https://hdl.handle.net/1887/3158649.

Fragkiadakis, M., and P. van der Putten. 2021. Sign and Search: Sign Search Functionality for Sign Language Lexica. In *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*, ed. D. Shterionov, 23–32. East Stroudsburg: Association for Machine Translation in the Americas. https://aclanthology.org/2021.mtsummit-at4ssl.3.

Giorgino, T. 2009. Computing and Visualizing Dynamic Time Warping Alignments in R: The DTW Package. *Journal of Statistical Software* 31 (7): 1–24.

Greenhill, S. J. 2011. Levenshtein Distances Fail to Identify Language Relationships Accurately. *Computational Linguistics* 37 (4): 689–98.

Hadjah, T. M., and A. Okyere. 2021. The History of Ghanaian Sign Language. Presented at the 5th African Sign Languages Workshop at the World Congress of African Linguistics (WOCAL). *Leiden, the Netherlands.* https://doi.org/10.17605/OSF.IO/7JQAW.

Hammarström, H., R. Forkel, M. Haspelmath, and S. Bank. 2021. *Glottolog 4.4.* Leipzig. https://doi.org/10.5281/zenodo.4761960.

HANDS!Lab. 2020. *Ghanaian Sign Language App.* https://play.google.com/store/apps/details?id=com.ljsharp.gsldictionary&hl=en&gl=US.

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

Hochgesang, J., O. A. Crasborn, and D. Lillo-Martin. 2018. Building the ASL Signbank. Lemmatization Principles for ASL. In *Proceedings of the Language Resources and Evaluation Conference 2018*, ed. M. Bono, E. Efthimiou, S. E. Fotinea. Miuazaki, 69–74. Paris, France: European Language Resources Association (ELRA).

Jangyodsuk, P., C. Conly, and V. Athitsos. 2014. Sign Language Recognition Using Dynamic Time Warping and Hand Shape Distance Based on Histogram of Oriented Gradient Features. In *Proceedings of the 7th International Conference on Pervasive Technologies Related to Assistive Environments,* 1–6. New York: Association for Computing Machinery.

Johnston, T. 2003. BSL, Auslan and NZSL: Three Signed Languages or One? In *Theoretical Issues in Sign Language Research Conference (7th: 2000),* ed. A. Baker, B. van den Bogaerde, O. Crasborn, 47–69. Hamburg: Signum.

Li, D., C. Rodriguez, X. Yu, and H. Li. 2020. Word-Level Deep Sign Language Recognition from Video: A New Large-Scale Dataset and Methods Comparison. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision,* 1459–69. New York: Institute of Electrical and Electronics Engineers (IEEE).

Liang, X., E. Kapetanios, B. Woll, and A. Angelopoulou. 2019. Real Time Hand Movement Trajectory Tracking for Enhancing Dementia Screening in Ageing Deaf Signers of British Sign Language. In *Machine Learning and Knowledge Extraction,* ed. A. Holzinger, P. Kieseberg, A. M. Tjoa, and E. Weippl, 377–94. New York: Springer.

McKee, D., and G. Kennedy. 2000. Lexical Comparison of Signs from American, Australian, British and New Zealand Sign Languages. *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima,* 49–76. New Jersey: Lawrence Erlbaum Associates.

Myers, C., L. Rabiner, and A. Rosenberg. 1980. Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28 (6): 623–35. https://doi.org/10.1109/TASSP.1980.1163491.

Napoli, D. J., and N. Sanders. Under review. Iconicity and Biomechanics in the Historical Reconstruction of Sign Languages: A Case Study of the Movement Parameter in the Old LSF Family. http://sanders.phonologist.org/Papers/napoli-sanders-iconicity-and-biomechanics.pdf

Nyst, V. 2007. A Descriptive Analysis of Adamorobe Sign Language (Ghana). PhD diss., Graduate School of Linguistics.

Omardeen, R. 2018. The Contribution of Lexical Overlap to Perceived Iconicity in Foreign Signs. Master's thesis, Radboud University.

Östling, R., C. Börstell, and S. Courtaux. 2018. Visual Iconicity across Sign Languages: Large-Scale Automated Video Analysis of Iconic Articulators and Locations. *Frontiers in Psychology* 9: 725.

Parks, J. 2011. *Sign Language Word List Comparisons: Toward a Replicable Coding and Scoring Methodology*. PhD diss., University of North Dakota.

Power, J. M., D. Quinto-Pozos, and D. Law. 2021. Methods and Models in Historical Comparative Research on Signed Languages. Presented at the 43rd Annual Conference of the German Linguistic Society (DGfS). Freiburg, Germany. https://repositories.lib.utexas.edu/handle/2152/96451

Ripperda, J., L. Drijvers, and J. Holler. 2020. Speeding Up the Detection of NonIconic and Iconic Gestures (SPUDNIG): A Toolkit for the Automatic Detection of Hand Movements and Gestures in Video Data. *Behavior Research Methods* 52 (4): 1783–94.

Schneider, P., R. Memmesheimer, I. Kramer, and D. Paulus. 2019. Gesture Recognition in RGB Videos Using Human Body Keypoints and Dynamic Time Warping. In *Robot World Cup*, ed. S. Chalup, T. Niemueller, J. Suthakorn, MA. Williams, 281–93. Cham, Switzerland: Springer .

Sutton, V. 2009. *SignWriting Reference Manual 1: SignWriting Basics*. California: SignWriting Press. https://www.academia.edu/download/60126225/sw0821_SignWriting_Basics_Instruction_Manual_Sutton20190726-19024-xq6bq4.pdf

Swadesh, M. 1955. Towards Greater Accuracy in Lexicostatistic Dating. *International Journal of American Linguistics* 21 (2): 121–37. Indiana: Indiana University.

Ten Holt, G. A., M. J. T. Reinders, and E. A. Hendriks. 2007. Multi-Dimensional Dynamic Time Warping for Gesture Recognition. In *Thirteenth Annual Conference of the Advanced School for Computing and Imaging,* ed. F.W. Jansen, G.E.O. Pierre, C.J. Veenman, J.W.J. Heijnsdijk, Heijen, the Netherlands, 23-32. Delft: Advanced School for Computing and Imaging.

Tormene, P., G. Toni, S. Quaglini, and M. Stefanelli. 2009. Matching Incomplete Time Series with Dynamic Time Warping: An Algorithm and an Application to Post-Stroke Rehabilitation. *Artificial Intelligence in Medicine* 45 (1): 11–34.

Wichmann, S., E. W. Holman, and C. H. Brown, eds. 2020. The ASJP Database (version 19). https://asjp.clld.org/.

Woodward, J. 1991. Sign Language Varieties in Costa Rica. *Sign Language Studies* 73 (1): 329–45.

———. 1993. The Relationship of Sign Language Varieties in India, Pakistan, and Nepal. *Sign Language Studies* 78 (1): 15–22.

Yu, S., C. Geraci, and N. Abner. 2018. Sign Languages and the Online World Online Dictionaries and Lexicostatistics. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation,* 4235–40. Paris, France: European Language Resources Association (ELRA).