Evaluating the Alignment of Utterances in the Swedish Sign Language Corpus

Carl Börstell
University of Bergen
Bergen, Norway
carl.borstell@uib.no

Abstract

The Swedish Sign Language (STS) Corpus mainly contains segmentations on the lexical level (i.e. signs), which makes it difficult to extract information at clause- or utterance-like levels. In this paper, I evaluate three different methods of segmenting the data into larger units: prosodic, syntactic and translation-based *utterance units*. The results show that none of the utterance units have particularly high accuracy in their alignment with the others, illustrating the challenges facing researchers who are looking to extract meaningful units above the lexical level. In a second step, I extract articulation information from the corpus videos using computer vision methods, but find no clear alignment of articulatory features of the hands and head with the boundaries of the utterance units.

Keywords: sign language, corpus, segmentation, clause, utterance, alignment, prosody, computer vision

1. Introduction

Today, there is an increasing number of corpora of sign languages in the world (Fenlon and Hochgesang, 2022; Kopf et al., 2022, 2023). Technical approaches can benefit from these resources, as well as facilitate their future expansion (see Morgan et al., 2022). Substantial information can be extracted even from very basic annotations, such as simple lexical level annotations – i.e. segmentations and annotations of each individual sign produced – which tend to be the initial steps of sign language corpora annotation work (Johnston, 2014). While such annotations can provide important insights into, e.g., lexical frequency, collocations and duration (Börstell, 2022b), it is more challenging to use lexical annotations alone to investigate grammatical constructions. This is mainly due to the fact that many sign language corpora lack any form of syntactic segmentation of the signing. One exception is the Auslan Corpus, which features so-called *clause-like units* that internally also have annotations for grammatical functions, enabling more detailed investigations into the syntactic organization of the language (Johnston, 2019). From the perspective of Conversation Analysis, Bono et al. (2020) annotated various layers of linguistic information – e.g., pragmatic, syntactic and phonetic – to segment a corpus of Japanese Sign Language (JSL) dialogues into utterance units based on those combined layers, facilitating research on the interactional aspects of sign language communication.

In this paper, I look at the Swedish Sign Language (STS; *svenskt teckenspråk*) Corpus (Mesch et al., 2012), which does not feature any clause- or utterance-unit segmentations on the whole. However, a small subset of the corpus has previously been annotated for syntactic relations (Östling et al., 2017), which can be used to infer clause or sentence units for that specific subset. Prosodically motivated segmentation of the corpus has been piloted as well, but was deemed inefficient as a method (Börstell et al., 2014). Without dedicated segmentations above the lexical level, research that required sentence-based segmentations has instead used the translation tier segmentations as an approximation of sentence units (Sjons, 2013; Östling et al., 2015). To date, there has been no evaluation of how past approaches to sentence- or utterance-unit segmentation/approximation align with one another. The goal of this paper is thus to evaluate the equivalence across approximations of utterance units in the STS Corpus, namely those based on available or inferred prosodic, syntactic and translation segmentations.

2. Background

The Swedish Sign Language (STS; *svenskt teckenspråk*) Corpus (Mesch et al., 2012) has been available for research since 2011, and has since been published as an online interface (Öqvist et al., 2020). The STS Corpus has mainly been annotated for sign glosses and idiomatic translations into written Swedish (Mesch et al., 2012; Mesch and Wallin, 2015), but has later been enriched with word class annotations (Östling et al., 2015). Smaller subsets have in addition been annotated for other properties such as backchannel responses (Mesch, 2016), mouthings (Mesch et al., 2021) and syntactic segmentations and relations (Börstell et al., 2016). However, there is no comprehensive type of segmentations beyond the original sign and translation tier annotations. In Börstell et al. (2016), we attempted a basic syntactic annotation of the STS Corpus, which involved segmenting clause-

Proceedings of the 11th Workshop on the Representation and Processing of Sign Languages, pages 36–45

like units on the basis of a combination of syntactic, semantic and prosodic properties of the signing. The definition centered around predicate-type signs as the core, and expressing a single idea within a single prosodic unit, definitions that were further used in later cross-linguistic research (Börstell et al., 2019). In Börstell et al. (2016), the first step was identifying and segmenting a syntactic unit, followed by annotating their internal relations for each sign. This proved to be quite time-consuming, and it involves simultaneous bottom-up and top-down approaches. That is, you need a segmentation to know which signs can relate to each other, but the signs that relate to each other also define the segmentation itself. In several other studies, utterances were inferred on the basis of the translation tier segmentations – i.e. the span of the Swedish translations across signs were used as approximate *utterance units* (defined here as a unit of segmentation corresponding to a level above the sign) – cf. Bono et al. (2020). For example, this was used in approaches to automatically word class tag the STS Corpus (Sjons, 2013; Östling et al., 2015). The translations are, however, not segmented systematically based on the signed articulation, but rather conversational content. In fact, translation annotation was mainly done independently of the sign gloss annotations, based on what could be conveniently expressed in written Swedish. Furthermore, translation segments do not always even correspond to a full sentence in neither Swedish nor STS, as many of them are partial sentences or fragments.

In Börstell et al. (2014), we experimented with ways of segmenting units based on visual prosodic cues, and whether these would correspond to syntactic units. A number of deaf signers were recruited to segment a subset of the STS Corpus based on visual prosodic cues alone, and these were compared to a syntactic segmentation made on the same subset. The results showed a lot of variation in the prosodic segmentations, and whereas some major prosodic breaks aligned across participants, it was deemed less reliable and inefficient as a method for segmenting the corpus data for syntactic purposes. Instead, the work from Börstell et al. (2016) was expanded on later in Östling et al. (2017), when we submitted a subset of the STS Corpus data to the *Universal Dependencies* (De Marneffe et al., 2021) dataset collection, making it the first sign language corpus to be added.[1] There, we instead worked in a bottom-up fashion, annotating grammatical relations between signs individually and later linking them together into a dependency tree automatically, thus skipping the explicit segmentation step in the annotation process. The STS dataset in Universal Dependencies is still very small, consisting of 1610 sign glosses across 203 sentences.

Although the Universal Dependencies STS dataset provides syntactic segmentation of clause-like units through its dependency trees, there has not been any evaluation of how well these syntactic units correspond to other units. For instance, to what extent do the syntactic units align with the translation units that have been used as placeholder sentence segmentations in previous work? Would either type of utterance unit, whether syntactic or translation-based, have any meaningful prosodic properties – e.g., notable pauses or other articulatory features around the start-/endpoints. We know from other research that sign language utterances display a multitude of prosodic features that can be used to segment and identify them, such as body, head and eyebrow movements and eyeblinks (Crasborn, 2007; Fenlon et al., 2007; Hansen and Heßmann, 2007; Herrmann, 2010; Sandler et al., 2011; Ormel and Crasborn, 2012; Puupponen et al., 2015; Puupponen, 2019; Kimmelman et al., 2020; Dachkovsky, 2022). Such features have in recent years been used in computer vision-based analyses of sign language data, as part of automatically extracting articulation and potentially segmenting continuous signing (Susman, 2022; Moryossef et al., 2023).

In this paper, I aim to:

1. compare and evaluate the alignment of prosodic, syntactic and translation utterance units in the STS Corpus
2. use computer vision-based tools to investigate articulatory correlates of these units

3. Methodology

For this study, I use the six original ELAN (Wittenburg et al., 2006) annotation files (`.eaf`) used in the annotation of the STS Universal Dependencies dataset (Östling et al., 2017). The six corpus files consist of 12 signers engaged in different types of conversation, between 1.5 and 3 minutes long (14 minutes and 5 seconds in total), comprising 1621 sign tokens: two free conversations (more dialogue) and four stories (more monologue).

The data processing, analysis and visualizations were done in R (R Core Team, 2023) with the packages `ggtext` (Wilke and Wiernik, 2022), `glue` (Hester and Bryan, 2022), `pracma` Borchers (2022), `scales` (Wickham and Seidel, 2022), `signglossR` (Börstell, 2022a), `tidyverse` (Wickham et al., 2019) and `udpipe` (Wijffels, 2023). The data and code for this study can be found at: `https://osf.io/fw825/`.

[1] STS is the only sign language represented in Universal Dependencies to date, but see Caligiore et al. (2020) for work on Italian Sign Language (LIS).

3.1. Defining units

The STS data as represented in the Universal Dependencies dataset contains the original sign annotations from the corpus as well as dependency relations between them. These dependency trees form a type of utterance unit segmentation of the STS Corpus data. The utterance units as defined by the Universal Dependencies dependency trees are in the following called *syntactic* utterance units. I compare these *syntactic* utterance units to the so-called *translation* utterance units. The *translation* utterance units are defined as the sign annotations that fall within or overlap with the temporal span of translation tier segmentations. I compare these two unit types also to a third type of utterance unit, labeled *prosodic* utterance units. The *prosodic* utterance units are defined as the sign sequences without any substantial pauses between signs. Here, the pause duration threshold has been set to the median duration of sign pauses between the syntactic units in the Universal Dependencies dataset: 322 milliseconds. Any pause between signs larger than that value forms a segmentation point marking a new prosodic utterance unit. The three types of utterance units – prosodic, syntactic and translation – result in slightly different numbers of utterance units, spanning different numbers of sign annotations (see Table 1).

Unit	# of units	# of signs
Prosodic	264	1621
Syntactic	203	1610
Translation	217	1611

Table 1: The number of utterance units per type and the number of sign annotations covered.

As is visible from Table 1, the largest number of signs is 1621, which is the same as the total number of tokens in the six corpus files of the dataset. This is only found for the prosodic unit segmentation, which is due to the fact that the prosodic segmentation is by definition done on the full dataset of (manual) sign annotations. The translation units have a slightly lower number, because some sign sequences have not been translated (generally short backchannel utterances). The syntactic units have the lowest sign counts because a few sign sequences in the dataset were never annotated for the Universal Dependencies dataset – e.g., due to the annotators being uncertain of the dependency analysis.

While the prosodic and syntactic utterance units always align exactly with the start and end of some sign annotations, since they are defined on the basis of those (sign) annotation segmentations, the translation utterance units do not necessarily align with sign annotation endpoints. Instead, the translation utterance units are treated as temporal segmentations, which can be aligned to the sign annotations based on overlap: if a sign annotation is completely within the boundaries of a translation unit, it is assigned to it; if a sign annotation overlaps with more than one translation unit, it is assigned to the first overlapping translation unit (see Figure 1).

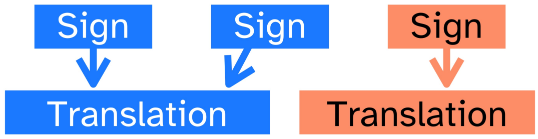

Figure 1: Assignment of signs to translation units.

3.2. Measuring Alignment of Units

Alignment across types of utterance units is analyzed in two ways.

First, the content equivalence of segments across tiers is defined as the intersection between unit types with regard to how many identical segments of sign annotations they share. That is, if the sequence of signs ABCDE is segmented as ABC, DE on one tier and A, BC, DE on the other, the two tiers share exactly one segment (i.e. DE).

Second, the temporal alignment and number of segmentations across the types of utterance units are analyzed with the Staccato algorithm (Lücking et al., 2011) as implemented in ELAN (Version 6.2) [Computer software] (2021). The Staccato algorithm is an implementation of the Thomann graph-theoretical method of segment alignment. This method looks at the so-called *degree of organization* of linear segments across tiers, defined as the correspondence of segments into temporally overlapping "shared nuclei" (core overlapping segments).[2] The metric of agreement (*degree of organization*) is based on the amount of overlap as well as the number of identified segments, compared to a chance baseline from iterated Monte Carlo Simulations, thus arriving at a metric between -1 (low) and 1 (high), where 0 is equal to chance levels in the degree of organization across tiers. Here, the algorithm is run for each pairwise utterance unit tier combination (per file and signer) with 1000 iterations (granularity = 10; α = .05). Thus, a value is obtained for every combination of utterance unit segmentation tiers (n=30).

3.3. Prosody with Computer Vision

Additionally, I extracted articulations through body-pose estimations of the signing in each of the six

[2]See also Rasenberg et al. (2022) for an example of this method used for inter-annotator reliability testing.

corpus files through the computer vision tool *MediaPipe* (Lugaresi et al., 2019). MediaPipe was used to estimate the location of various body landmarks in each of the front-facing videos linked to the ELAN files – thus 12 videos, as there are two signers with one main front-facing video file each for each corpus file. MediaPipe has previously been shown to be successful in analyzing articulatory properties in sign language videos, such as extracting sign articulation onsets and locations (Börstell, 2023) and comparing phonetic features of different text types (Kimmelman and Teresė, 2023).

Here, I focus on the distance moved across frames by 1) the two hands (based on wrist positions in two dimensions) and 2) the head (based on nose position in the vertical dimension), respectively. That is, how far in signing space have the hands and head moved between every sequence of two frames in the video? This is done to identify prosodically prominent points in manual and non-manual articulation – points in time in the files where the hands and/or head move more than usual. The metric used for distance moved is the raw Euclidean distance moved in the MediaPipe coordinate system, but z-scored within each file and signer for cross-signer and cross-file comparison. The measurements for distance moved by the hands and head were then analyzed for peaks to find sequences of increased activity in relative movement. This was done with the `pracma::findpeaks` function, extracting peaks – defined as frames with a previous increase and following decrease in movement activity (±3 frames) – in the hand and head movement data. With this method, 369 peaks were found in the hand movements across files, and 329 peaks were found in the head movements.

4. Results

As seen in Table 1, the syntactic and translation units are more closely overlapping in the total number of units segmented, even though the prosodic unit segmentation was performed on the basis of the median pause duration between syntactic units. When looking at the sign sequences that correspond to each utterance unit (i.e. overlapping sign annotations in the case of translation units), there is a similarity in unit contents that corresponds to the number of units. Table 2 shows the intersection of sign annotation sequence segmentations across utterance unit types, illustrating that the syntactic and translation units have just over 30% overlap in sign sequences resulting from the segmentations, whereas the prosodic utterance units only overlap at around 13–20% with the other utterance unit types. Thus, in terms of content equivalence of sign sequences, it seems the syntactic and translation segmentations have the highest agreement.

Turning to the general temporal alignment between utterance units, Figure 2 shows all segmentations temporally aligned across the six corpus files. There is, unsurprisingly, agreement on when there is articulation happening in general, but the segmentation endpoints are not always aligned. Although the prosodic utterance units are the most numerous, there are examples where they span much longer stretches of signing than either syntactic or translation units, illustrating sequences with only very short "pauses" between sign annotations. However, we can also see that the translation units are the ones most often entirely mismatched in terms of content, such as including an annotation where the others do not. This happens, for example, by translating non-manual content (e.g., translating visible laughter at the end of file `SSLC02_332`) or failing to add a translation annotation in cases of short turns (e.g., several missing annotations in file `SSLC01_104` that constitute short response tokens). The missing segments on the syntactic tier are stretches of glosses that are missing from the dependency annotations, thus lacking a corresponding syntactic unit.

As a second type of alignment measure, I used the Staccato algorithm (Lücking et al., 2011) implemented in ELAN to evaluate the agreement between annotation segmentations across utterance unit types. Figure 3 shows the distribution of scores achieved by each comparison, where circles represent each annotation tier comparison and their relative size corresponds to the number of segments per tier (smaller size means fewer segments to match for those tiers). As is visible from Figure 3, the scores obtained in terms of degree of organization are all quite poor, mostly falling at or below chance levels. Opposite to the patterns found for content equivalence in Table 2, the highest scores come from the alignment between prosodic and syntactic units, followed by syntactic and translation units, and lastly prosodic and translation units. Generally, tiers with only a single annotation (usually a single response token or comment by the addressee at the end of a narrative) receive perfect alignment scores, but tiers with many more annotations display much lower agreement.

Turning to the MediaPipe data, Figure 4 shows the movement (distance traveled) of hands and head (solid and dotted lines) within each of the six corpus files. It also shows the major points of segmentation agreement (vertical lines; n=89), defined as points in time at which all three utterance unit types have marked the start or end of an annotation segment. The movement data is z-scored within signers to show relative movement and smoothed with a LOESS function: the solid lines show the articulation of the hands (distance moved by the

		Unit 2 (comparison)		
		Prosodic	Syntactic	Translation
Unit 1	Prosodic	———	41/264; 15.5%	34/264; 12.9%
	Syntactic	41/203; 20.2%	———	65/203; 32.0%
	Translation	34/213; 16.0%	65/213; 30.5%	———

Table 2: The overlap of sign annotation sequences between utterance unit segmentations.

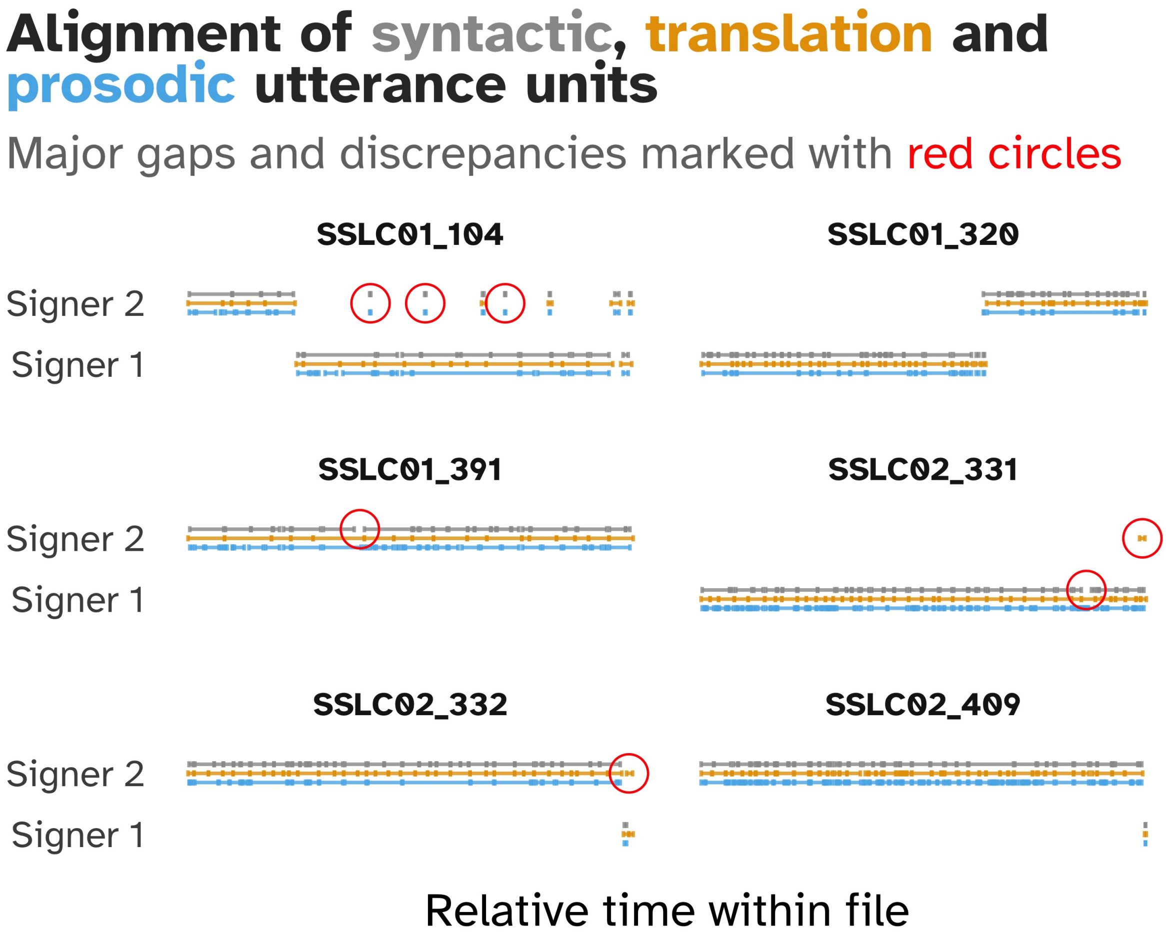

Figure 2: Alignment across utterance units. Red circles mark areas of major discrepancies.

wrist landmarks) and dotted lines show the articulation of the head (vertical distance moved by the nose landmark). The articulation activity can clearly show the main contributor in a text, thus show the major turn-taking events in a conversation (see file `SSLC01_320`; NB: Signers with minimal signing in a file have been filtered out here).

Based on Figure 4, there are no obvious visual correlations between the major segmentation points across utterance units and the articulatory activity of the hands and head. Despite some of the segmentation points matching up with either peaks, valleys or changes in overall contour, the picture is too varied to show any obvious patterns of alignment. Out of the identified peaks in the MediaPipe movement data, only 7 (1.9%) of the hand peaks and 9 (2.7%) of the head peaks occurred within 3 frames of a major segmentation points (i.e. start- or endpoints aligned across all three utterance unit types). Similarly, only 7 (8.1%) and 8 (9.3%) of segmentation points occurred within three frames of a hand or head peak, respectively.

5. Discussion and Conclusion

The goal of this study was to evaluate the equivalence and potential usefulness of various types of utterance units in the STS Corpus based on prosodic, syntactic and translation-based segmentations. Seeing as a subset of the STS Corpus is annotated syntactically, these segmentations could

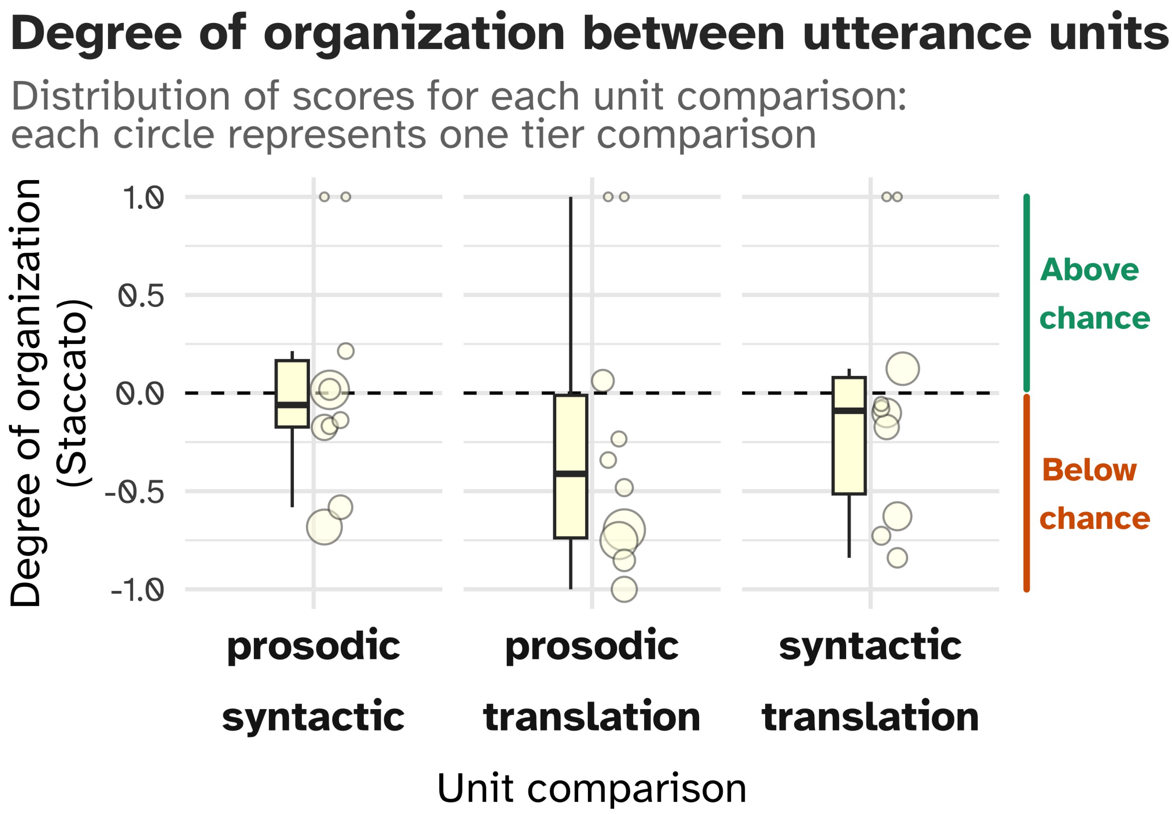

Figure 3: Degree of organization between utterance types using the Staccato algorithm. Circles represent each annotation tier comparison, sizes corresponding to number of annotations per tier. Box plots show the distribution of scores. Dashed lines show chance level.

form a starting point for analyzing the distribution of clause-like units in the corpus, potentially informing automated methods of extracting them. Before such syntactic segmentations were available, the translation tier segmentations had been used as a proxy for a more clause- or sentence-like unit. Segmenting sign annotations into utterance units based on pauses between annotations is another approach, using a type of prosodic (pause duration) information to identify segmentation points.

In this study, it was found that the three methods for identifying utterance units arrive at quite different exact sequences of signs, with at most around 30% overlap in the sequences of signs identified through the different segmentation methods. This shows a low degree of content equivalence between the methods, suggesting that the translation segmentations used in some previous work as a proxy for a sentence-like unit (cf. Sjons, 2013; Östling et al., 2015) do not correspond very closely to the clause-like units identified through manual syntactic annotation (Östling et al., 2017). Nonetheless, the overlap across sign sequence segmentations was higher between syntactic and translation units than any other pairwise comparison. However, the agreement of segment alignment using the Staccato algorithm (Lücking et al., 2011) pointed to a higher similarity between prosodic and syntactic utterance units than any other pairwise comparison. I suspect this to be the result of the start- and endpoints of these units always aligning exactly with sign annotation start- and endpoints, whereas the translation segments are made independently of the sign gloss annotations and rarely align exactly with them at the ends. Additionally, the translation tier segmentations had more instances of complete mismatches compared to the other two tiers, by either adding translations where there were no manual sign annotations or lacking annotations for short manual response tokens (see Figure 2). It is possible that the algorithm is less suitable for this type of data, for which there is often a continuous stream of annotations (i.e. many throughout the file) rather than fewer annotations more sparsely spread out in time. If so, it may not be ideal for evaluating segmentations if the goal of a segmentation is to find the *contents* of what falls within its span, rather than finding its exact endpoints. Another issue is that the number of segments matters for the Staccato algorithm, and the granularity of

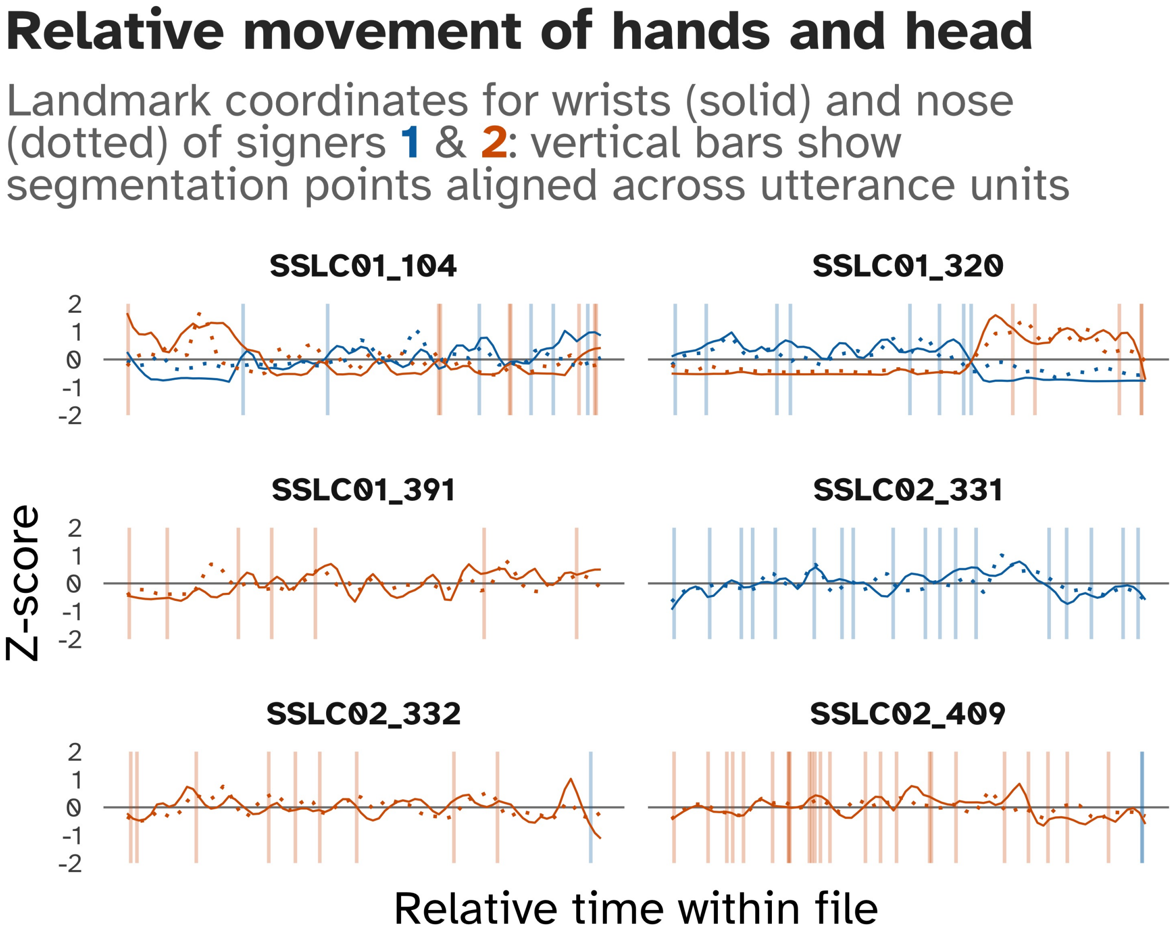

Figure 4: Relative distance moved by hands and head. Solid lines show hand articulation and dotted lines show head articulation (both smoothed with a LOESS function). Vertical lines correspond to utterance unit segmentation points (start or end) matched across all three utterance unit types.

the different methods is quite different as they are based on different motivations: what matters syntactically, what is a convenient content chunk, or what is defined as "pauses".

The second part of this study looked at prosodic correlates between the identified utterance units and articulatory data extracted from the corpus videos using MediaPipe (Lugaresi et al., 2019). Whereas the extracted data can clearly show patterns such as major turn-taking events between signers in conversation, it was not possible to identify any obvious correlations between shared segmentation points (start or end) across utterance unit types and articulatory patterns in the movement of hands and head. However, seeing as this dataset is only a small subset of the STS Corpus, the lack of found patterns/correlates may simply be due to the lack of sufficient data. A type of hybrid approach was proposed by Chizhikova and Kimmelman (2022), who in their analysis of headshakes and negation used computer vision-based methods together with manual inspection. As the STS Corpus continues to grow in terms of features annotated for, there will be better opportunities to measure correlations between manually annotated prosodic features and those extracted automatically, as well as using aggregated data from multiple layers of linguistic information – e.g., prosodic, semantic and interactional (cf. Bono et al., 2020) – to arrive at meaningful utterance units.

In summary, this study has shown that the currently available utterance units (whether annotated or inferred) in the STS Corpus do not align to any greater extent. This means that researchers using these units – possibly as a proxy of "sentences" – need to take great care in choosing motivated unit types and be aware of their limitations. The future goal for the STS Corpus should be to segment the sign annotations into some meaningful larger unit, whether conversational turns or utterances or syntactic sentences or clauses. This would increase the potential of the corpus as a language resource substantially, as it would allow for analyses of language structure beyond the individual signs.

6. Acknowledgements

I am grateful to Mark Dingemanse for directing me to the Staccato algorithm. I thank three anonymous reviewers for comments and suggestions.

7. Data Availability

Data and code are available in an online OSF repository: `https://osf.io/fw825/`

8. Bibliographical References

Mayumi Bono, Rui Sakaida, Tomohiro Okada, and Yusuke Miyao. 2020. Utterance-Unit Annotation for the JSL Dialogue Corpus: Toward a Multimodal Approach to Corpus Linguistics. 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*, pages 13–20, Marseille, France. European Language Resources Association (ELRA).

Hans W. Borchers. 2022. *pracma: Practical Numerical Math Functions*.

Carl Börstell. 2022a. Introducing the signglossR Package. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 16–23, Marseille, France. European Language Resources Association (ELRA).

Carl Börstell. 2022b. Searching and Utilizing Corpora. In Jordan Fenlon and Julie A. Hochgesang, editors, *Signed Language Corpora*, number 25 in Sociolinguistics in deaf communities, pages 90–127. Gallaudet University Press, Washington, DC.

Carl Börstell. 2023. Extracting Sign Language Articulation from Videos with MediaPipe. In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, NEALT Proceedings Series, No. 52, pages 169–178, Tórshavn, Faroe Islands. University of Tartu Library.

Carl Börstell, Tommi Jantunen, Vadim Kimmelman, Vanja de Lint, Johanna Mesch, and Marloes Oomen. 2019. Transitivity prominence within and across modalities. *Open Linguistics*, 5:666–689.

Carl Börstell, Johanna Mesch, and Lars Wallin. 2014. Segmenting the Swedish Sign Language corpus: On the possibilities of using visual cues as a basis for syntactic segmentation. In *Proceedings of the LREC2014 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel*, pages 7–10, Reykjavik, Iceland. European Language Resources Association (ELRA).

Carl Börstell, Mats Wirén, Johanna Mesch, and Moa Gärdenfors. 2016. Towards an Annotation of Syntactic Structure in the Swedish Sign Language Corpus. In *Proceedings of the LREC2016 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*, pages 19–24, Portorož, Slovenia. European Language Resources Association (ELRA).

Gaia Caligiore, Cristina Bosco, and Alessandro Mazzei. 2020. Building a Treebank in Universal Dependencies for Italian Sign Language. In *Proceedings of the Seventh Italian Conference on Computational Linguistics*, pages 1–6, Bologna, Italy. CEUR.

Anastasia Chizhikova and Vadim Kimmelman. 2022. Phonetics of Negative Headshake in Russian Sign Language: A Small-Scale Corpus Study. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 29–36, Marseille, France. European Language Resources Association (ELRA).

Onno Crasborn. 2007. How to recognise a sentence when you see one. *Sign Language & Linguistics*, 10(2):103–111.

Svetlana Dachkovsky. 2022. Emergence of a subordinate construction in a sign language: Intonation ploughs the field for morphosyntax. *Glossa: a journal of general linguistics*, 7(1).

Marie-Catherine De Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, pages 1–54.

ELAN (Version 6.2) [Computer software]. 2021. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen. [link].

Jordan Fenlon, Tanya Denmark, Ruth Campbell, and Bencie Woll. 2007. Seeing sentence boundaries. *Sign Language & Linguistics*, 10(2):177–200.

Jordan Fenlon and Julie A. Hochgesang, editors. 2022. *Signed Language Corpora*. Number 25 in Sociolinguistics in deaf communities. Gallaudet University Press, Washington, DC.

Martje Hansen and Jens Heßmann. 2007. Matching propositional content and formal markers: Sentence boundaries in a DGS text. *Sign Language & Linguistics*, 10(2):145–175.

Annika Herrmann. 2010. The interaction of eye blinks and other prosodic cues in German Sign Language. *Sign Language & Linguistics*, 13(1):3–39.

Jim Hester and Jennifer Bryan. 2022. *glue: Interpreted String Literals*.

Trevor Johnston. 2014. The reluctant oracle: Adding value to, and extracting of value from, a signed language corpus through strategic annotations. *Corpora*, 9(2):155–189.

Trevor Johnston. 2019. Clause constituents, arguments and the question of grammatical relations in Auslan (Australian Sign Language): A corpus-based study. *Studies in Language*, 43(4):941–996.

Vadim Kimmelman, Alfarabi Imashev, Medet Mukushev, and Anara Sandygulova. 2020. Eyebrow position in grammatical and emotional expressions in Kazakh-Russian Sign Language: A quantitative study. *PLOS ONE*, 15(6):e0233731.

Vadim Kimmelman and Anželika Teresė. 2023. Analyzing literary texts in Lithuanian Sign Language with Computer Vision: a proof of concept. In *NAIS 2023: The 2023 symposium of the Norwegian AI Society*, volume 3431 of *CEUR Workshop Proceedings*, page 5, Bergen, Norway. Technical University of Aachen.

Maria Kopf, Marc Schulder, and Thomas Hanke. 2022. The Sign Language Dataset Compendium: Creating an Overview of Digital Linguistic Resources. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 102–109, Marseille, France. European Language Resources Association.

Maria Kopf, Marc Schulder, and Thomas Hanke. 2023. The Sign Language Dataset Compendium. Technical report. Version Number: 1.3.

Andy Lücking, Sebastian Ptock, and Kirsten Bergmann. 2011. Staccato: Segmentation Agreement Calculator according to Thomann. In *Proceedings of the 9th International Gesture Workshop: Gestures in Embodied Communication and Human-Computer Interaction*, pages 50–53, Athens, Greece.

Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. MediaPipe: A Framework for Building Perception Pipelines. Publisher: arXiv Version Number: 1.

Johanna Mesch. 2016. Manual backchannel responses in signers' conversations in Swedish Sign Language. *Language & Communication*, 50:22–41.

Johanna Mesch, Krister Schönström, and Sebastian Embacher. 2021. Mouthings in Swedish Sign Language: An exploratory study. *Grazer Linguistische Studien*, 93:107–135. Publisher: Universität Graz.

Johanna Mesch and Lars Wallin. 2015. Gloss annotations in the Swedish Sign Language Corpus. *International Journal of Corpus Linguistics*, 20(1):103–121.

Johanna Mesch, Lars Wallin, and Thomas Björkstrand. 2012. Sign Language Resources in Sweden: Dictionary and Corpus. In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon [Language Resources and Evaluation Conference (LREC)]*, pages 127–130, Paris. European Language Resources Association (ELRA).

Hope Morgan, Onno Crasborn, Maria Kopf, Marc Schulder, and Thomas Hanke. 2022. Facilitating the Spread of New Sign Language Technologies across Europe. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 144–147, Marseille, France. European Language Resources Association.

Amit Moryossef, Zifan Jiang, Mathias Müller, Sarah Ebling, and Yoav Goldberg. 2023. Linguistically Motivated Sign Language Segmentation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 12703–12724, Singapore. Association for Computational Linguistics.

Zrajm Öqvist, Nikolaus Riemer Kankkonen, and Johanna Mesch. 2020. STS-korpus: A Sign Language Web Corpus Tool for Teaching and Public Use. 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*, pages 177–180, Marseille, France. European Language Resources Association (ELRA).

Ellen Ormel and Onno Crasborn. 2012. Prosodic Correlates of Sentences in Signed Languages:

A Literature Review and Suggestions for New Types of Studies. *Sign Language Studies*, 12(2):279–315. Publisher: Gallaudet University Press.

Robert Östling, Carl Börstell, Moa Gärdenfors, and Mats Wirén. 2017. Universal Dependencies for Swedish Sign Language. In *Proceedings of the 21st Nordic Conference on Computational Linguistics*, pages 303–308, Gothenburg, Sweden. Association for Computational Linguistics.

Robert Östling, Carl Börstell, and Lars Wallin. 2015. Enriching the Swedish Sign Language Corpus with Part of Speech Tags Using Joint Bayesian Word Alignment and Annotation Transfer. In *Proceedings of the 20th Nordic Conference of Computational Linguistics (NoDaLiDa 2015)*, pages 263–268, Vilnius, Lithuania. Linköping University Electronic Press, Sweden.

Anna Puupponen. 2019. Towards understanding nonmanuality: A semiotic treatment of signers' head movements. *Glossa: a journal of general linguistics*, 4(1).

Anna Puupponen, Tuija Wainio, Birgitta Burger, and Tommi Jantunen. 2015. Head movements in Finnish Sign Language on the basis of Motion Capture data: A study of the form and function of nods, nodding, head thrusts, and head pulls. *Sign Language & Linguistics*, 18(1):41–89.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Marlou Rasenberg, Asli Özyürek, Sara Bögels, and Mark Dingemanse. 2022. The Primacy of Multimodal Alignment in Converging on Shared Symbols for Novel Referents. *Discourse Processes*, 59(3):209–236.

Wendy Sandler, Irit Meir, Svetlana Dachkovsky, Carol Padden, and Mark Aronoff. 2011. The emergence of complexity in prosody and syntax. *Lingua*, 121(13):2014–2033.

Johan Sjons. 2013. Automatic induction of word classes in Swedish Sign Language. Master's thesis, Stockholm University.

Margaux Susman. 2022. Eye Blinks in French Sign Language Definition of eye blink types and automatic detection of eye blinks using computer vision, rule-based and machine learning-based methods. Master's thesis, University of Bergen, Bergen, Norway.

Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Pedersen, Evan Miller, Stephan Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. 2019. Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43):1686.

Hadley Wickham and Dana Seidel. 2022. *scales: Scale Functions for Visualization*.

Jan Wijffels. 2023. *udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit*.

Claus O. Wilke and Brenton M. Wiernik. 2022. *ggtext: Improved Text Rendering Support for 'ggplot2'*.

Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, pages 1556–1559.

9. Language Resource References

Mesch, Johanna and Wallin, Lars and Nilsson, Anna-Lena and Bergman, Brita. 2012. *Dataset. Swedish Sign Language Corpus project 2009–2011 (version 1)*. Sign Language Section, Department of Linguistics, Stockholm University. PID https://hdl.handle.net/1839/b9b9c88a-f8df-4fa5-8eb0-53622108764d.