Manual vs. automated annotation of non-manual signs: ELAN *vs.* FaceReader

Leticia Hanada*

Abstract: For more than a decade, signed language linguists have employed the free software ELAN to annotate manual and non-manual signs (NMSs) (Wittenburg et al. 2006). In the past ten years, several studies have sought to advance the automatic recognition of NMSs (Antonakos, Roussos, & Zafeiriou 2015; Bhuvan et al. 2016), including Hanada (2023), who proposes using the paid software FaceReader for the automated recognition of facial expressions and head movements. The present study aims to investigate whether manual transcription of NMSs in ELAN is consistent with automated transcription in FaceReader – that is, whether both transcriptions are equivalent and reliable. To this end, the same signed data from Hanada (2024) were used to produce both transcriptions, with ELAN labels matched to the Action Units (AUs) and head positions coded by FaceReader. In total, 649 NMSs were manually transcribed in ELAN, and 2,240 were automatically coded in FaceReader. The degree of agreement and disagreement regarding whether a given NMS was marked as active or inactive was then analyzed. The results show that the two tools agreed in 48.03% of the data. The main discrepancies were due to: (1) variations in the baseline used to define the participant's neutral face in the human annotations, which depended on what the annotator had recently perceived as neutral during their analysis; (2) the fact that when two or more movements co-occurred, only the most salient or intense one was annotated in ELAN; and (3) FaceReader is not designed for linguistic analysis and annotates all detected facial movements, including those that may not have a linguistic function, while also assigning a degree of deviation to head positions even when the head is at rest. Thus, although it can serve as a second annotator to support intercoder reliability (ICR) in qualitative studies, it is not recommended to rely solely on FaceReader for analyzing NMSs. The findings of this study highlight the advantages and limitations of each transcription method and contribute to a better understanding of which research contexts may benefit more from one type of annotation than the other. Future research could examine how often related muscles co-occur and how this affects the annotation of distinct NMS movements.

Keywords. ELAN; FaceReader; reliability; transcription; non-manual signs

1. Introduction. The articulators responsible for expressing information in a sign language include not only manual signs, that is, two hands, but also non-manual signs (NMSs). In sign language linguistics, the term "non-manual" was introduced to describe aspects of signing that go beyond the actions of the hands, such as body movements and facial expressions (Liddell 2003) - eyebrows, eyes, eyelids, nose, mouth, tongue, cheeks, chin, head, torso, and shoulders. While the literature features numerous studies focusing on manual signs (Ann 2005; Eccarius & Brentari 2007; Henner et al. 2013), relatively few have addressed facial

* I would like to express my sincere gratitude to Professor Corrine Occhino for her invaluable guidance in developing ideas for this paper. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.
Leticia Hanada, The University of Texas at Austin (leticia.hanada@fulbrightmail.org).

articulation (Wilbur & Patschke 1998; Boyes Braem & Sutton-Spence 2001) or head and torso movements (Liddell 1986; Schalber 2006; Lackner 2015).

One primary reason NMSs have not been the central focus of signed linguistics research is that manual signs are often treated as analogous to words in spoken language, serving as the core conveyors of semantic content. A common misconception is that in spoken languages, non-vocal information is entirely paralinguistic, that is, utilized only for emotional expression. Even though Stokoe (2001) titled his book "Language in Hand," he refused the argument that limits sign language analysis strictly to the manual signs.

Instead of considering NMSs to play a mere complementary role or only add information to manual signs, it is possible to argue that manual signs interoperate with motions of the face, head, and torso. The paper by Baker (1976) about ASL was one of the first to acknowledge this. It describes four non-manual channels used alongside manual signs: eyes, face, head, and body. Baker and Padden (1978) introduced the "non-manual" terminology for the identification of a set of NMSs. In their first attempt at description, they noticed variation in the use of non-manuals among signers. They suggested, "it is the configuration or pattern of co-occurring behaviors that serves a given function rather than any specific behavior" (Baker & Padden 1978: 33).

Neither of these ideas considers that NMSs, beyond emotional expressions, can also have a linguistic role (Baker & Padden 1978; Baker 1980; Bergman 1983; Liddell 1980; Pizzuto et al. 1990; Sutton-Spence & Woll 1999). The limited analyses available on NMSs, such as those by Baker and Padden (1978), Baker-Shenk (1983), and Liddell (1978, 1980), have not typically examined these features independently. Coulter (1979) was a pioneering study that addressed NMSs individually to understand their functions. Studies on American Sign Language (ASL) have demonstrated that the head, body, and upper face can alter sentence polarity, establish sentence types, mark topicalized signs, convey agreement, and accompany various embedded clauses and person distinctions in pronouns (Pfau et al. 2010).

Building on this, Aarons (1994) identified that head and upper torso movements manifest over the c-command domain of nodes. Bahan (1996) described the association among head tilt, eye gaze, and phi-features in agreement-projection heads. Liddell (1986) linked head thrust to conditional marking in ASL, while Wilbur (1994) demonstrated that eyeblinks are sensitive to syntactic structure. Wilbur (2003) shows that NMSs can provide lexical or morphemic information or indicate the end of phrases or phrasal extent. Wilcox S. and Wilcox P. (2005) stated that NMSs would be like prosodic changes, such as voice qualities, intonation, and rhythm in spoken languages. This is also assumed by Dachkovsky and Sandler (2009), who argue that shoulders and torso movements represent intonation contour and allow stress marking. According to Wilbur (2022), facial expressions and head movements can also be used to mark intonational contour; for example, topics, negation, and questions are marked by head movement; conditions, yes-no questions, and pronominal references are marked by the eyes and eyebrows; evaluation is denoted by the nose. Similarly, Nguyen and Ranganath (2008) claimed that in ASL, eyebrow, eye, and head movements are responsible for marking wh-questions, yes-no questions, rhetorical questions, topics, conditional clauses, relative clauses, and negation. Ladd (1996) emphasized the intonational role of NMSs and noted that they also participate in the phonology of lexical signs, marking adverbial and adjectival phrases (Anderson & Reilly 1998; Liddell 1980; Meir & Sandler 2007), as well as iconic gestures and emotions. Body movements were observed to play an important role in constructing action and to be a key strategy for the use of space in signed discourse, coinciding with various discourse-level units (Boyes Braem 1999; Crasborn & van der Kooij 2013; Engberg-Pedersen 1993; Hodge & Ferrara 2014; Wilbur & Patschke 1998).

In Brazilian Sign Language (Libras), NMSs have syntactic functions (Quadros & Karnopp 2004), such as agreement, emphasis, topicalization, and sentence modalities, like partial and total interrogatives, as well as phonological functions, like lexical items, pronominal reference, negative particles, affirmation, degree, and adverbial value. Additionally, Paiva et al. (2018) cite the function of NMSs in meaning intensification. These findings, along with a growing body of research, continue to link NMSs to grammatical structures and to different subfields of Linguistics (phonology, phonetics, morphology, syntax, semantics, pragmatics, discourse, prosody, etc.), highlighting their crucial role in signed languages.

2. Background. For more than a decade, linguists have employed the free software ELAN for much of their video annotation (Wittenburg et al. 2006), as it enables researchers to link annotations to audio-visual data in real time and add unlimited tiers (Crasborn & Sloetjes 2010), allowing precise frame-by-frame annotation and facilitating various types of analysis. ELAN also allows the simultaneous visualization of multiple files within a single transcription. This feature is particularly important for phonetic and prosodic analyses, as it enables researchers to view the same realization from different camera perspectives. Another advantage is the ability to align collected data with the stimuli that elicited it, facilitating teamwork where multiple annotators, including those unfamiliar with the data, can contribute to the annotation process. Additionally, ELAN provides tools for efficient data collection across existing transcripts, such as word counts and tier-based searches. These advantages make ELAN widely recognized among sign linguists as a tool for efficient transcript exchange. Notably, ELAN supports the annotation of both manual signs and NMSs, which is crucial for sign language linguistics research.

However, a significant drawback is the time required for manual transcription, especially when working with large corpora or conducting quantitative studies. Researchers must manually transcribe each instance of the movement under study and determine its start and end. This process becomes particularly challenging when analyzing NMSs, which involve intricate movements of the eyebrows, eyes, mouth, nose, cheeks, head, and torso (Ferreira-Brito 1995). Another critical challenge is accommodating corrections and intercoder reliability checks. This is because transcriber effects are inevitable; a single transcriber can unintentionally influence how the data is transcribed due to personal interpretation, attention level, biases, etc.

Over the past decade, studies have sought to advance the automatic recognition of manual signs (Starner & Weaver & Pentland 1998; Cooper & Holt & Bowden 2011). However, given the importance of NMS transcription in sign languages, other studies have focused on automatically recognizing NMSs (Antonakos & Roussos & Zafeiriou 2015; Bhuvan et al. 2016) and on developing methods to integrate them with manual sign coding (Yang & Lee 2013; Quesada & Marín & Guerrero 2016). Despite significant progress in automatic annotation tools with the potential for reliable large-scale analyses (Karppa et al. 2012; Puupponen et al. 2015), it remains necessary to assess the effectiveness of automatic analysis in accurately capturing the structure and expression of signed languages. In this sense, automated tools like FaceReader can offer an alternative. FaceReader is a paid software program developed by Noldus for the automated recognition of facial expressions and head positions. Using the Facial Action Coding System (FACS), FaceReader identifies the activity of up to 43 facial muscles (20 in the most affordable package), referred to as Action Units (AUs), and their intensity levels: trace, slight, pronounced, severe, or max.

Beyond coding individual AUs, FaceReader also annotates combinations of AUs when multiple of them are activated simultaneously. This automation has the potential to significantly benefit sign language research by enabling analysis of larger datasets and saving

valuable research time. By automating the transcription of NMSs, researchers can allocate more resources to other aspects of their studies and analyses. The data are coded by FaceReader as numerical values (on a gradient scale) and are then translated into intensity categories according to the FaceReader manual. That is, AUs intensities[1] are classified in a scale from 0 to 1, in which Not active is from 0.00 to 0.100; Trace (A) is from 0.100 to 0.217; Slight (B) is from 0.217 to 0.334; Pronounced (C) is from 0.334 to 0.622; Severe (D) is from 0.622 to 0.910; and Max (E) is from 0.910 to 1.000 (Face Reader Reference Manual 9, p. 265).

Using a deep learning-based face-detection algorithm, FaceReader first identifies the face's position in an image and then synthesizes an artificial face model with nearly 500 key points to estimate facial landmarks. This facial modeling method enables FaceReader to capture the face's state. Finally, it identifies facial expressions using a trained deep artificial neural network capable of recognizing face patterns. This network was trained on more than 20,000 manually annotated images of people from diverse backgrounds worldwide (FaceReader 9 Help 2019).

In addition to coding facial features, the program analyzes head orientations relative to the x, y, and z axes: pitch, yaw, and roll, respectively. These head orientations represent the degrees of deviation from looking straight. The default range for pitch and yaw values is -30° to 30° in either direction, while any minimum or maximum limits do not constrain roll. If the angle exceeds this threshold, FaceReader rejects the face model. It can also recognize not only the degrees of deviation in pitch, yaw, and roll but also the head's position relative to the camera (horizontal, vertical, and depth).

The program uses pitch and yaw values to assess the Model Quality, which indicates the image quality required for the program to accurately detect and classify a face (FaceReader Reference Manual, 340). The minimum quality threshold for a model to be considered valid is 0.5, on a scale ranging from 0 (low quality) to 1 (high quality). The quality indicator is represented by a color-coded bar: red signifies low quality, orange indicates medium quality, and green denotes good quality. For a model to be considered of good quality, the green portion of the bar must cross the dashed line at 0.5. Another advantage of the program is its ability to normalize or calibrate the participant's face. This can be done manually by selecting a segment of the video where the participant exhibits a neutral expression, or automatically, with the program continuously adjusting its analysis based on changes in the participant's facial expressions. Hanada (2023) also cites two significant disadvantages of FaceReader for linguists: its high cost and the lack of support for analyzing torso movements.

As mentioned, ELAN is a trusted program already integrated into the practices of signed linguists. However, when working with extensive corpora, transcribing both manual and non-manual signs can be time-consuming. Therefore, when considering a quantitative approach, it is worth exploring tools that automatically transcribe such data. Although ELAN and FaceReader serve different functions, both can be used to transcribe NMSs in sign language.

Thus, the present study aims to investigate the reliability of these two tools in transcribing the same video sample. Specifically, the goal is to determine whether the transcriptions from each tool refer to the same data or appear to interpret different data points. Reliability means that the transcriptions are consistent across different tools (Abbuhl & Mackey 2008). Therefore, this is a proof-of-concept study with the purpose of comparing two transcription tools for NMSs and exploring the pros and cons of each in the context of NMS

[1] If such categories are not imposed, however, the data could instead be analyzed phonetically using FaceReader to annotate the intensity of muscle movements activated during a sentence, akin to a spectrogram in phonetic analysis. In contrast, the labels used in ELAN can be viewed as a more categorical, phonological form of analysis, comparable to the use of the IPA.

analysis. However, the intention here is to understand what is captured or missed by a human versus an automatic transcription, and not to provide linguistic results about how NMSs operate in Libras, or determine which transcription method is superior.

3. Method. The present study uses data from Libras to examine potential differences in manual transcription in ELAN and automated transcription in FaceReader. The FaceReader data[2] utilized here was obtained from Hanada (2024). Below are the settings in FaceReader chosen by Hanada (2024): 1) The use the of General Face Model, as the corpus contains only Western adult signers, 2) Continuous calibration of the face, since it was not possible to capture a neutral facial expression from the participants before the experiment, and 3) All the data are in the green portion of the Model Quality bar – above 0.5, which means that the videos have good quality for facial recognition.

Because the purpose of the present study is to compare two programs, the ELAN labels utilized by Hanada (2024) were matched with the AUs coded in FaceReader[3] (Table 1). The transcription in ELAN was carried out without having access to FaceReader's automatic coding. This guarantees that one transcription will not bias the other.

	ELAN transcription	FaceReader transcription
Eyebrows	134 (Raised)	286 (Inner Brow Raiser - AU 01) + 184 (Outer Brow Raiser - AU 02) = 470
	16 (Furrowed)	287 (Brow Lowerer - AU 04)
Eyes	8 (Wide open)	161 (Upper Lid Raiser - AU 05)
	14 (Squinted)	146 (Lid Tightener - AU 07)
Nose	1 (Wrinkle)	15 (Nose Wrinkler - AU 09)
Head	142 (Head tilted right or left)	387 (Roll)
	21 (Head tilted forward or backward) + 73 (Facing up or down) + 163 (Head nod) = 257	387 (Pitch)
	75 (Facing left or right) + 2 (Head shake) = 77	387 (Yaw)
Total:	649	2,240

Table 1. NMSs equivalency between ELAN and FaceReader

It is important to note that mouth movements were excluded because participants mouthed almost every sign in the sentences. The NMS Cheek Raiser (AU 06) movements were also excluded because, if analyzed, they would be transcribed as either Squint Eyes or Smiling Eyes, even though the AU itself refers to the cheek movement. Therefore, only raised and furrowed eyebrows, wide-open and squinted eyes, and wrinkled nose were part of the analysis.

[2] The data was collected from 9 deaf signers from the deaf community of Sao Carlos (Sao Paulo, Brazil), six men and five women, aged between 20-46 years old. They self-identified as fluent in Libras. The signers use Libras daily with their families, both hearing and deaf friends, and in their work or university settings.

[3] For a more systematic comparison between Libras Facial Expressions associated with the Facial Action Coding System, visit da Silva et al. (2020).

Only the maximum activation value[4] of AUs and head orientations in FaceReader that co-occurred with MSs[5] were coded, as they reflect the peak intensity of the AU (the goal to be reached). Given that the signs were generally produced rapidly (averaging 700 ms), the default frame rate of 5 frames per second (FPS) was adopted when coding AUs and head orientations. Finally, to facilitate the comparison of head movements between the programs, we tabulated head movement degrees coded by FaceReader as the following six group ranges: Up to 5°; 5° to 10°; 10° to 15°; 15° to 20°; 20° to 25°; 25° to 30°. This decision relies on the fact that FaceReader also codes AU intensity into 6 groups, enabling comparison of AUs and head movement intensity levels. The data from both transcriptions were entered into an Excel table for comparison.

4. Results. In total, the author of the present study manually transcribed 649, while FaceReader automatically coded 2,240 NMSs. In other words, FaceReader annotated approximately 3.5 times more movements.

However, if trace and slight active AUs movements are disregarded, as well as head movements with less than 10° of deviation, one could say that FaceReader annotated a similar number of movements as those annotated in ELAN, as the total amount recognized by FaceReader would be 680 active movements[6]. From that number, a first impression was that only intense movements are perceived by the human brain, which would explain why the total number of intensely activated movements in FaceReader is comparable to the number of transcriptions made in ELAN by a human annotator. However, before drawing this conclusion, it is essential to analyze each result comparatively, considering all intensity levels of muscle activation and head orientations.

4.1 AGREEMENT. Chart 1 shows that 48.03% (1453) of the manual annotations in ELAN are equivalent to the number of automated annotations generated by FaceReader. This means that the eyebrows, eyes, nose, and head movements that were labeled as “active” in ELAN were also detected by FaceReader with some degree of activation (trace, slight, pronounced, severe, or max), and those labeled in FaceReader as “inactive” were also not detected by the human annotator. Regarding our first assumption that the human annotator would be able to capture only more intense movements, it is now possible to argue that this is not the case, as 24.23% of the annotations in ELAN corresponded to trace or slight levels of AUs and head movements (223 trace + 129 slightly active movements = 352), which are the two least intense activation levels of muscles detected in FaceReader.

[4] I acknowledge that selecting the maximum AU value within the data timeframe may not be the most effective approach for analyzing individual signs in FaceReader, as NMSs, like MSs, can be influenced by coarticulation effects.

[5] This decision was made because, in most cases, more than one NMS was coded for the same MSs. For instance, the sign BAR was produced from 00:30:565 to 00:31:785.

[6] AU 01+02 = 247 (36,3%), AU 04 = 21 (3,1%), AU 05 = 37 (5,4%), AU 07 = 5 (0,73%), AU 09 = 1 (0,14%), pitch = 141 (20,7%), yaw = 71 (10,4%), and roll = 157 (23%). 680 annotations in total.

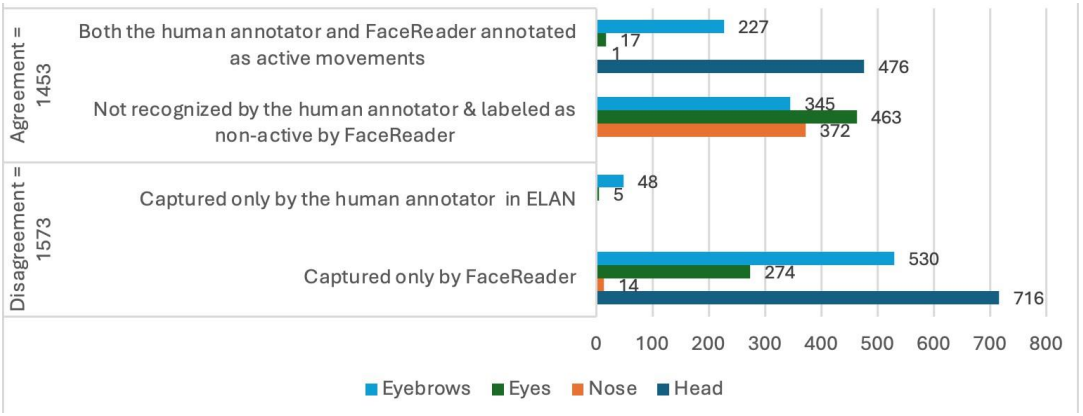

Chart 1. Differences between ELAN and FaceReader transcriptions

4.2. DISAGREEMENT. The two transcriptions disagreed on 51.97% of the annotated data. 3.36% (53 = 48 eyebrows and 5 eye movements) of the annotations captured by the human annotator in ELAN were not identified by FaceReader. Out of 48 annotations of active raised eyebrows only transcribed in ELAN, just 11 were coded by FaceReader as inactive for both AU 01 (inner brow raiser - IBR) and 02 (outer brow raiser - OBR). Upon reviewing these instances, it was observed that eight of them belonged to the only female participant wearing glasses, which could explain the discrepancies between the transcriptions. Hanada (2024) notes that the recognition quality of a participant's face may be affected by the presence of glasses or a beard, potentially impacting FaceReader's accuracy in detecting facial expressions, as glasses may obscure eyebrow analysis. This might lead us to infer that FaceReader can face difficulty detecting eyebrow and eye movements in the presence of a facial accessory. However, it is challenging to determine the exact cause of this difference, since the average maximum value for the Model Quality of that participant is 0.698 (Hanada 2024), which exceeds the default threshold of 0.5 and thus indicates medium-good to good facial modeling quality. Furthermore, another (male) participant was also wearing glasses and the same difference wasn`t observed.

For raised eyebrows considered active in ELAN, 91.2% matched at least one AU coded as active. Given that FaceReader categorizes raised eyebrows as two distinct movements, one might ask whether there is a perceptual difference between them during manual facial expression transcription. When comparing each of these AUs to ELAN annotations, certain patterns emerge. Whenever OBR was categorized as trace, IBR was categorized as either trace or a higher level – the same pattern held for other intensity levels. A complementary pattern was identified when analyzing inactive raised eyebrows: in many cases where IBR was inactive, OBR was also inactive (92.2%), aligning with ELAN's transcription. However, when OBR was inactive, IBR were inactive (53.2%) or with some degree of activation (46.8%). This suggests that the Inner Brow Raiser (IBR) may be more salient to human perception when determining whether a raised eyebrow is active.

If this is indeed the case, why then 20.47% of OBR annotations were labeled as pronounced or severe? A closer examination of these cases revealed two main explanations: either the movement was not labeled as raised eyebrows in ELAN because, when contrasted with previous raised eyebrow movements and neutral eyebrows, it appeared closer to a neutral state; or FaceReader categorized the frame as a raised eyebrow because the preceding (or, in a few cases, following) sign was produced with raised eyebrows, meaning the frame captured the transition back to a neutral position or the anticipation of a raised position.

When considering only IBR, 72.2% (182 instances) of the movements labeled as inactive in ELAN were categorized as either inactive or activated (trace or slight) in FaceReader. From this result, it appears that, even though the human annotator can identify slightly

activated movements, this is done by contrasting them with what was recently considered a neutral face. In other words, it seems that the human annotator memorized the participant's neutral face from the most recently analyzed data and compared it with the current data, updating the information as the transcription process progressed. Thus, the baseline for the participant's neutral state varied with the transcription process, and the face's neutral state was updated in the human annotator's memory each time they created a new annotation of an active facial movement. It is important to note that this could be avoided by having a single picture of the participant's neutral face, which can then serve as a base for annotating all active facial expressions. However, this should not be a problem for the present study, as the calibration type selected in FaceReader was the "Continuous Calibration". This means the program adapted the analysis to the user's neutral state throughout the recording, and it is especially useful when a neutral facial sample is unavailable. The other type of calibration would be the "Participant Calibration", which, is like having one picture of the neutral face of the participant during the transcription task, as FaceReader takes a photo as the baseline a or a short video of the participant looking at a neutral screen. Finally, five instances of squinted eyes (lid tightener - AU 07) were marked as active only in ELAN. They involved head-backward movements, which made it more difficult for FaceReader to capture the eyes' movements.

Again, the two annotators disagreed on 51.97% of the transcribed data. From this, 48.61% (1534) represents transcriptions made only by FaceReader. As observed in chart 1, 818 annotations were of facial expressions (530 eyebrows, 274 eyes, 14 nose), and 716 were of the head. Regarding the 818 facial expressions, patterns were found across all of them.

The main reasons for discrepancies were coarticulation of NMSs, non-linguistic factors, and the presence of facial accessories, such as glasses. The "coarticulation of NMSs" refers to an unconscious preference of the human annotator for transcribing intense NMS over co-occurring ones with less perceptual visibility. For example, if a participant produced both an intense raised eyebrow and a trace of wide-open eyes, the annotator transcribed only the raised eyebrow, interpreting that the trace of wide-open eyes was a consequence of the eyebrow movements rather than as an intentional activation. This means the annotator coded the eye movement if and only if they assumed it would exist independently of eyebrow movement. Here are some articulators that co-occurred: raised eyebrows and wide-open eyes; furrowed eyebrows; squinted eyes[7], with a smile or cheek movements; and the three axes of head movement[8]. This distinction can also be described as a difference in annotation strategy: to make the manual and automated coding more comparable, the human annotator opted to separate eyebrow movements from eye and mouth-and-cheek movements. This highlights the need for further research into how often movements occur simultaneously versus independently and suggests that FaceReader can detect movements even in contexts where the researcher might not have considered them relevant to linguistic analysis. This underscores the need for annotators to assess whether such factors affect annotation accuracy.

The second reason for the discrepancy was non-linguistic factors, such as transitions between movements, the final-sentence position, and eye and head movements with the purpose of looking at the computer. Finally, another possible reason for the discrepancies was the presence of glasses, which

[7] In the manual transcription, the criterion for distinguishing squinted eyes from smiling eyes was that the eyes were labeled as squinted when they were not the result of a smile, or the Cheek Raiser (AU 06) associated with a smile. Although FaceReader uses AU 06 to describe cheek movement that affects the eyes, this movement appears to significantly affect the eyes' appearance. It is unclear whether FaceReader distinguishes between smiling and squinting eyes. This could be another reason for the discrepancy between the two transcriptions: the manual transcription treats squinted and smiling eyes as distinct eye appearances, whereas FaceReader may consider both as part of AU 07, marking AU 06 only for the cheek movement.

[8] It is important to acknowledge that when a change in head orientation occurs, all axes are affected.

made it difficult for the human annotator to capture the eyebrows and eye movements. However, as previously noted, this was also a challenge for FaceReader, suggesting that certain factors pose challenges for both manual and automated transcription in detecting facial movements, such as accessories (glasses, piercings, etc.) or other facial features (beard, bangs, etc.) that obscure the face.

5. Discussion. The results show that a human annotator can identify different intensity levels of NMSs and code them as active, including less intense levels such as trace and slight. However, because humans might categorize the world contrastively (Prince & Alvarez & Konkle 2024), trace and slight active movements were coded only when contrasted with facial expressions considered as neutral by the human annotator. On the other hand, FaceReader's system calibrates the participant's face, establishing a neutral baseline that matches the participant's habitual expression, enabling consistent activation and intensity coding. According to Artuso, Palladino, and Ricciardelli (2012), memory is a cognitive system that constantly updates information it deems useful, including visual information. In doing so, memory serves to maintain goal-relevant information and inhibit irrelevant information (Morris & Jones 1990; Palladino et al. 2001; Oberauer 2005). This same function applies to facial processing, which is crucial for individuals to stay engaged in interpreting socially salient facial stimuli. One hypothesis is that it may be difficult for a human annotator to maintain a single, neutral facial state in mind while observing many facial expressions produced at varying intensities. This is because the human memory tends to be continuously updated by the most recent facial expression the transcriber observed. As a result, manual transcription in ELAN tends to be dichotomous, classifying movements as either active or not, or using just a few intensity categories (such as not active, slightly active, pronounced active). This dichotomous approach makes it difficult to categorize movements into six intensity levels, since the difference between a trace and a slight movement may not be obvious to human perception. Therefore, it is still unclear whether a human annotator can categorize these intensity levels as efficiently and with the same precision as FaceReader. The automated software can also code a larger volume of data in less time. It can detect annotations that were missed or excluded by the human annotator, providing detailed information on the intensity levels of facial expressions and the degrees of head-orientation deviations, as well as co-occurring AUs and head movements. The automated recognition of various degrees of head-orientation deviation is also a valuable resource for understanding the constraints of NMSs, which is essential to the fields of Articulatory Phonetics, Phonology, and Prosody of sign languages. Since this is not an easy task for human perception, FaceReader can be a useful tool in this context. It can be employed in studies aimed at understanding variations in the intensity of facial expressions and head movements, as it provides objective, consistent analysis.

Additionally, the human annotator focuses on more intense/relevant movements and unconsciously ignores other movements that might co-occur but with lower intensity. The task of coding facial expressions becomes even harder when movements are co-articulated alongside other movements, such as eyebrows and eyes (e.g., raised), cheeks and mouth (e.g., smile), eyes and mouth (e.g., smile), eyes and nose (squint and wrinkle), or nose and mouth (e.g., wrinkle and upper lip raiser). Our understanding of the relationship between different articulators remains quite limited. While studies such as Wilbur (2000) highlight the independence and activity of various NMSs, noting that different articulators can produce a single NMS, it is still necessary to investigate whether certain NMSs tend to co-occur or occur independently. From this perspective, signed linguists should ask themselves whether specific NMSs are typically activated in the presence of others due to anatomical connections or shared linguistic functions, or whether these channels operate independently to convey meaning.

Gaining this knowledge would inform decisions about whether to transcribe NMSs separately or collectively and help clarify how they relate to prosodic, phonological, syntactic, and discourse elements. Puupponen (2018) is one of the few studies to explore these questions, investigating whether head and torso movements are independent or coordinated based on how often they occur separately versus in tandem.

The process of transcribing data manually involves making important decisions based on the researcher's interests and the type of data collected. In this study, the researcher decided to annotate only NMSs that accompany linguistic content, excluding other NMSs motivated by non-linguistic factors or by transitional movements associated with the articulation of a preceding or following sign. This was the main drawback of using FaceReader for linguistic analysis, as it coded every single movement of the face and head, even when those movements were not linguistically relevant to the analysis of NMSs.

ELAN enables quick and easy analysis of annotation durations, which is especially important for NMSs in sign language, as they are suprasegmental and span multiple words. This becomes challenging in FaceReader because the program’s analysis is based on frames, which do not always align with the exact moments when signs or utterances begin and end.

In ELAN, it is also possible to annotate the direction of eye gaze, including when eyes are directed at manual signs. In contrast, FaceReader only has Action Units (AUs) referring to wide-open or closed eyes. Similarly, the annotator can differentiate between various types of head movements in ELAN, such as head nods, shakes, and tilts, along with their respective directions, and annotate the movements the researcher deems activated. On the other hand, FaceReader does not readily allow distinguishing when head orientation is inactive, as it assigns a specific degree of deviation to every frame. Additionally, while the researcher can examine other aspects of the video in ELAN, such as torso orientation and position, FaceReader is limited to facial expressions and changes in head orientation.

Finally, Chart 2 indicates that the degree of agreement and disagreement between annotations was similar across levels of muscle activation intensity, except for trace movements. Further research is necessary to understand why “trace” intensity level was the main reason for disagreement between manual and automated transcriptions. One possible hypothesis is that humans might not be able to capture the trace activity of movements unless they contrast with what was recently considered the neutral state of the participant`s face or perceived as carrying communicative or linguistic function/meaning. If this is the case, one can also hypothesize that the trace level might represent the threshold between active and inactive NMSs. This underscores the distinction between human coders - who evaluate both form and meaning in the linguistic analysis of sign language - and automated tools like FaceReader, which rely solely on formal, visual features. If trace movements are excluded from the analysis, the agreement percentage between the two annotators increases from 48.03% to 61.81%.

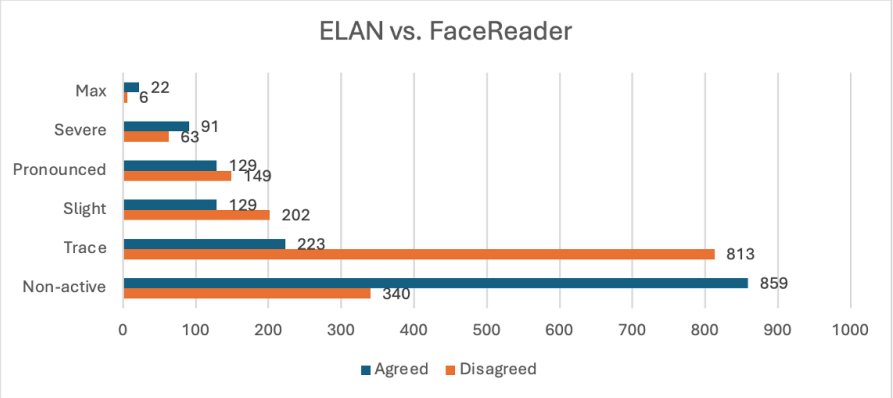

Chart 2. Distribution of NMSs` intensity degree according to agreement

6. Conclusion. In terms of reliability, given that FaceReader is not designed for linguistic analysis, it is preferable not to rely solely on it for linguistic transcription, as the program may capture unintended movements that do not align with the research goals. Similarly, it is not ideal to rely exclusively on a single human annotator, as it depends on subjective judgment shaped by the transcriber's interpretation, attention level, and biases, which can affect the results and conclusions drawn from the observed data. This variability may stem from differences in the transcribers' training, linguistic backgrounds, ideologies, and other factors. Therefore, one possible use of the automated software is to serve as a second annotator to ensure intercoder reliability[9].

The comparison between human and automated annotations highlights the strengths and limitations of each approach in analyzing NMSs in sign languages. The adoption of either transcription method depends heavily on the researcher's objectives, the type of data being analyzed, whether the study is qualitative or quantitative, and whether the data is elicited. A combined approach is the most beneficial, utilizing the strengths of both methods. For instance, if a researcher wants to investigate duration information, they can use ELAN to extract precise durations by identifying the start and end times of signs. On the other hand, if one's goal is to explore varying intensities of NMSs, FaceReader provides a more precise and accurate categorization of the data.

The findings of this study contribute to a better understanding of the differences between manual and automated transcriptions of facial expressions and how the choice of transcription methods and research criteria influences the observed results.

7. Limitations. The main limitation of this study was the absence of a second human annotator for manual transcription, which would have enabled a more reliable comparison and a better assessment of consistency in facial and head movement annotations. While automated tools like FaceReader show promise, the lack of a second annotator in this study limits the ability to fully evaluate the reliability and consistency of manual transcriptions. This is especially important, as manual transcriptions can vary from one annotator to another, and having two or more annotators would allow for more consistent coding. Additionally, it is important to include a native Libras signer among the annotators to transcribe the NMSs in the data. According to Kuhl et al. (1992), linguistic experience influences phonetic perception at an early age. This means there is also the possibility that the annotator may approximate their phonetic annotation to what is familiar in their first language (the magnet effect). Although the coded

[9] According to O`Connor and Joffe (2020), intercoder reliability (ICR) is a numerical measure of agreement between different coders regarding how the same data should be coded, which can be calculated using different statistical tests, such as Cohen`s kappa, Krippendorff`s alpha, Scott`s pi, Fleiss`K, Analysis of Variance binary ICC, and the Kuder- Richardson 20.

data is in a different modality from the author's first language (Brazilian Portuguese), it is necessary to keep in mind that this might result in differences in the manual transcription.

Finally, it is important to note that, since only FaceReader data from Hanada (2024) was utilized here, the AU Eyes Closed (AU 43) was excluded from her analysis. Although participants were instructed to look at the camera while signing their responses during the experiment, they did not consistently maintain direct eye contact with the camera. This was particularly evident at the beginning and end of utterances when participants briefly looked at the computer slide presentation to read the glossed sentences. As a result, FaceReader frequently annotated a significant portion of the signed data as if the participants had their eyes closed. The same pattern was observed with other AUs analyzed in the present study. This represents a limitation of the present study, as the researcher did not identify this during data collection from the first participants, which means that, if the movements of the eyes and mouth hadn't been excluded, additional data could contribute to understanding the effects of different types of transcription. In future research, it is crucial to filter out such occurrences beforehand, as they may significantly influence the results. This highlights the risk of concluding based solely on formal features without examining the actual context of the production. To ensure linguistic relevance, it is essential to code not only the form but also the intended meaning behind the articulation, determining whether the muscle activation contributes to the expression of linguistic information.

References

Aarons, Debra. 1994. *Aspects of the syntax of American Sign Language*. Boston: Boston University.

Abbuhl, Rebekah & Alison Mackey. 2008. Second language acquisition research methods. *Encyclopedia of Language and Education* 10. 99–111.

Anderson, Diane E. & Judy Reilly. 1998. PAH! The acquisition of adverbials in ASL. *Sign Language & Linguistics* 1(2). 117–142.

Ann, Jean. 2005. A functional explanation of Taiwan Sign Language handshape frequency. *Language and Linguistics* 6(2). 217.

Antonakos, Epameinondas, Anastasios Roussos & Stefanos Zafeiriou. 2015. A survey on mouth modeling and analysis for sign language recognition. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, 1–7. IEEE.

Artuso, Caterina, Paola Palladino & Paola Ricciardelli. 2012. How do we update faces? Effects of gaze direction and facial expressions on working memory updating. *Frontiers in Psychology* 3. 362.

Bahan, Benjamin J. 1996. *Non-manual realization of agreement in American Sign Language*. Boston: Boston University.

Baker, Charlotte. 1976. What's not on the other hand in American Sign Language. In *Papers from the Regional Meeting of the Chicago Linguistic Society* 12. 24–32.

Baker, Charlotte. 1980. Sentences in American Sign Language. In *Sign language and the deaf community*, 75–86.

Baker, Charlotte & Carol Padden. 1978. Focusing on the nonmanual components of American Sign Language. In Patricia Siple (ed.), *Understanding language through sign language research*, 27–57.

Baker-Shenk, Charlotte Lee. 1983. *A microanalysis of the nonmanual components of questions in American Sign Language*. Berkeley: University of California.

Bergman, Brita. 1983. Verbs and adjectives: Morphological processes in Swedish Sign Language. *Language in Sign* 8(2). 3–9.

Bhuvan, M. S., D. V. Rao, S. Jain, T. S. Ashwin, R. M. R. Guddetti & S. P. Kulgod. 2016. Detection and analysis model for grammatical facial expressions in sign language. In *2016 IEEE Region 10 Symposium (TENSYMP)*, 155–160. IEEE.
Boyes-Braem, Penny & Rachel Sutton-Spence. 2001. *The hands are the head of the mouth: The mouth as articulator in sign languages*. Hamburg: Signum Press.
Braem, Penny Boyes. 1999. Rhythmic temporal patterns in the signing of deaf early and late learners of Swiss German Sign Language. *Language and Speech* 42(2–3). 177–208.
Cooper, Helen, Brian Holt & Richard Bowden. 2011. Sign language recognition. In *Visual analysis of humans: Looking at people*, 539–562. London: Springer.
Coulter, Geoffrey. 1979. *American Sign Language typology*. San Diego: University of California.
Crasborn, Onno A. & Han Sloetjes. 2010. Using ELAN for annotating sign language corpora in a team setting. Manuscript.
Crasborn, Onno & Els van der Kooij. 2013. The phonology of focus in Sign Language of the Netherlands. *Journal of Linguistics* 49(3). 515–565.
Dachkovsky, Svetlana & Wendy Sandler. 2009. Visual intonation in the prosody of a sign language. *Language and Speech* 52(2–3). 287–314.
da Silva, Emely Pujólli, Paula Dornhofer Paro Costa, Kate Mamhy Oliveira Kumada, José Mario De Martino & Gabriela Araújo Florentino. 2020. Recognition of affective and grammatical facial expressions: A study for Brazilian Sign Language. In *Computer Vision – ECCV 2020 Workshops*, 218–236. Springer.
Eccarius, Petra & Diane Brentari. 2007. Symmetry and dominance: A cross-linguistic study of signs and classifier constructions. *Lingua* 117(7). 1169–1201.
Engberg-Pedersen, Elisabeth. 1993. *Space in Danish Sign Language: The semantics and morphosyntax of the use of space in a visual language*.
Ferreira-Brito, Lucinda. 1995. *Por uma gramática das línguas de sinais*. Rio de Janeiro: Tempo Brasileiro.
Hanada, Leticia K. 2023. Automated coding non-manual markers: Sign language corpora analysis using FaceReader. *Revincluso – Revista Inclusão & Sociedade* 3(1).
Hanada, Leticia K. 2024. *Experimental study of the expression of focus of new information and replacing contrastive focus in Brazilian Sign Language (Libras)*. Campinas: Universidade Estadual de Campinas master's thesis. https://hdl.handle.net/20.500.12733/17892
Henner, Jonathan, Leah C. Geer & Diane Lillo-Martin. 2013. Calculating frequency of occurrence of ASL handshapes. In *LSA Annual Meeting Extended Abstracts* 4. 16–1.
Hodge, Gabrielle & Lindsay N. Ferrara. 2014. Showing the story: Enactment as performance in Auslan narratives. In *Selected papers from the 44th Conference of the Australian Linguistic Society*, 372–397. Melbourne: University of Melbourne.
Karppa, Matti, Tommi Jantunen, Ville Viitaniemi, Jorma Laaksonen, Birgitta Burger & Danny De Weerdt. 2012. Comparing computer vision analysis of signed language video with motion capture recordings. In *LREC*, 2421–2425.
Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens & Björn Lindblom. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255(5044). 606–608.
Lackner, Andrea. 2015. Linguistic functions of head and body movements in Austrian Sign Language (ÖGS): A corpus-based analysis. *Sign Language & Linguistics* 18(1). 151–157.
Ladd, D. Robert. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
Liddell, Scott K. 1978. Nonmanual signals and relative clauses in American Sign Language. In *Understanding language through sign language research*, 59–90.

Liddell, Scott K. 1980. *American Sign Language syntax*. The Hague: Mouton.
Liddell, Scott K. 1986. Head thrust in ASL conditional marking. *Sign Language Studies* 52. 244–262.
Liddell, Scott K. 2003. *Grammar, gesture, and meaning in American Sign Language*. Cambridge: Cambridge University Press.
Meir, Irit & Wendy Sandler. 2007. *A language in space: The story of Israeli Sign Language*. New York: Psychology Press.
Morris, Neil & Dylan M. Jones. 1990. Memory updating in working memory: The role of central executive. *British Journal of Psychology* 81. 111–121.
Nguyen, Tan Dat & Surendra Ranganath. 2008. Tracking facial features under occlusions and recognizing facial expressions in sign language. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, 1–7. IEEE.
O'Connor, Cliodhna & Helene Joffe. 2020. Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods* 19.
Oberauer, Klaus. 2005. Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General* 134. 368–387.
Paiva, Francisco Aulísio dos Santos, Plínio Almeida Barbosa, José Mario De Martino, Ackley Dias Will, Márcia Regina Nepomuceno dos Santos Oliveira, Ivani Rodrigues Silva & André Nogueira Xavier. 2018. Análise do papel das expressões não manuais na intensificação em Libras. *DELTA* 34. 1135–1158.
Palladino, Paola, Cesare Cornoldi, Rossana De Beni & Francesca Pazzaglia. 2001. Working memory and updating processes in reading comprehension. *Memory & Cognition* 29. 344–354.
Pfau, Roland & Josep Quer. 2010. Nonmanuals: Their grammatical and prosodic roles. In *Sign languages*, 381–402.
Pizzuto, Elena, Enza Guiranna & Giuseppe Gambino. 1990. Manual and nonmanual morphology in Italian Sign Language: Grammatical constraints and discourse processes. In Ceil Lucas (ed.), *Sign language research: Theoretical issues*, 83–102. Washington, DC: Gallaudet University Press.
Prince, Jacob S., George A. Alvarez & Talia Konkle. 2024. Contrastive learning explains the emergence and function of visual category-selective regions. *Science Advances* 10(39). eadl1776.
Puupponen, Anna. 2018. The relationship between movements and positions of the head and the torso in Finnish Sign Language. *Sign Language Studies* 18(2). 175–214.
Puupponen, Anna, T. Wainio, Birgit Burger & Tommi Jantunen. 2015. Head movements in Finnish Sign Language on the basis of motion capture data. *Sign Language & Linguistics* 18(1). 41–89.
Quadros, Ronice Müller de & Lodenir Becker Karnopp. 2004. *Língua de sinais brasileira: Estudos linguísticos*. Porto Alegre: Artmed.
Quesada, Luis, Gabriela Marín & Luis A. Guerrero. 2016. Sign language recognition model combining non-manual markers and handshapes. In *Ubiquitous computing and ambient intelligence*, 400–405. Springer.
Schalber, Katharina. 2006. What is the chin doing? An analysis of interrogatives in Austrian Sign Language. *Sign Language & Linguistics* 9(1–2). 133–150.
Starner, Thad, Joshua Weaver & Alex Pentland. 1998. Real-time American Sign Language recognition using desk and wearable computer-based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(12). 1371–1375.
Stokoe, William C. 2001. *Language in hand: Why sign came before speech*. Washington, DC: Gallaudet University Press.

Sutton-Spence, Rachel & Bencie Woll. 1999. *The linguistics of British Sign Language*. Cambridge: Cambridge University Press.
Wilbur, Ronnie. 1994. Eyeblinks & ASL phrase structure. *Sign Language Studies* 84. 221–240.
Wilbur, Ronnie B. 2000. Phonological and prosodic layering of nonmanuals in American Sign Language. In Karen Emmorey & Harlan Lane (eds.), *The signs of language revisited*, 215–244. Mahwah, NJ: Erlbaum.
Wilbur, Ronnie B. 2003. Modality and the structure of language: Sign languages versus signed systems. In *Oxford handbook of deaf studies, language, and education*, 332–346.
Wilbur, Ronnie B. 2022. Prosody in sign languages. *Hrvatska revija za rehabilitacijska istraživanja* 58 (Spec. Iss.). 143–174.
Wilbur, Ronnie B. & Cynthia G. Patschke. 1998. Body leans and the marking of contrast in American Sign Language. *Journal of Pragmatics* 30(3). 275–303.
Wilcox, Sherman & Phyllis Perrin Wilcox. 2005. *Aprender a ver: O ensino da língua de sinais americana como segunda língua*. Arara Azul.
Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann & Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. In *LREC 2006*, 1556–1559.
Yang, Hee-Deok D. & Seong-Whan Lee. 2013. Robust sign language recognition by combining manual and non-manual features. *Pattern Recognition Letters* 34(16). 2051–2056.