The "how-to" of integrating FACS and ELAN for analysis of non-manual features in ASL

Kristin Mulrooney, Julie A. Hochgesang, Carla D. Morris, Katie Lee
Gallaudet University
Department of Linguistics
800 Florida Avenue, NE
Washington, DC 20002
Email: Kristin.mulrooney@gallaudet.edu, julie.hochgesang@gallaudet.edu, carla.morris@gallaudet.edu, katie.lee@gallaudet.edu

Abstract
The process of transcribing and annotating non-manual features presents challenges for sign language researchers. This paper describes the approach used by our research team to integrate the Facial Action Coding System (FACS) with the EUDICO Linguistic Annotator (ELAN) program to allow us to more accurately and efficiently code non-manual features. Preliminary findings are presented which demonstrate that this approach is useful for a fuller description of facial expressions.

Keywords: American Sign Language, FACS, ELAN, non-manual features

1. Introduction

The process of transcribing and annotating non-manual features presents challenges for sign language researchers. This paper describes the approach used by our research team to integrate the Facial Action Coding System (FACS) with the EUDICO Linguistic Annotator (ELAN) program to allow us to more accurately and efficiently code non-manual features.

Since 2010, researchers in the Department of Linguistics at Gallaudet University have collaborated with avatar developers VCom3D, Inc. The most recent collaboration is part of VCom3D's *Mobile Signing Math Dictionary with Mouth Morphemes* project, which was established because "[e]xisting animations of facial expressions and speech fall short of addressing the full range of "visible speech" and mouth morphemes." The Gallaudet research team has two main tasks on this project. One, to provide feedback on avatar animations as to the accuracy and naturalness of the facial behaviors. The second, to analyze naturally produced ASL discourse in a variety of settings (classroom, narratives), and identify the most frequent range of the most commonly used facial expressions.

Initially, the team utilized a notation system which was supplied by VCom3D. This system provided to the Gallaudet team, which was accompanied by a video clip collection of a model demonstrating each label, "bundled" the actions of individual face muscles, and presents them as one unified behavior. This system used global labels such as 'surprise' or 'oo' to describe an entire facial expression. However, facial expressions, more often than not, contain more than one meaningful part conveying multifaceted information. For example, a facial expression could contain the meaning of both surprise and a WH question, or anger and a WH question. The two meaningful parts being generated via distinct muscle movements in different areas of the face.

The Gallaudet team concluded that the "bundled" approach motivated by complex meanings would not accurately describe the facial behaviors used to create a whole facial expression. This moved the team to create an alternative system, the basis of which included dividing the parts of the face (on separate ELAN tiers), describing what each of the parts did (using FACS), and when the different movements occurred (separate annotations on independent ELAN tiers). By doing so, it has enabled the team to annotate actions of the separate facial muscles, such as the eyebrows, mouth, and cheeks, as they moved independently (or dependently), and identify when those movements occur synchronously or not.

2. Facial Action Coding System

The Facial Action Coding System (FACS), by Paul Ekman, Wallace V. Friesen, and Joseph C. Hager (2002), is a system to taxonomize human facial movements by their appearance on the face. In other words, FACS is a coding system used to define groups of facial behaviors and movements on the basis of shared characteristics and giving names to those defined groups. In FACS, movements of individual facial muscles are encoded based on slight changes in outward facial appearances.

FACS allows researchers to code nearly any anatomically possible facial expression, deconstructing them into specific Action Units (AU). The temporal sequence of those segments result in unique meaningful expressions. The FACS manual defines AU as a contraction or relaxation of one or more of the facial muscles, which are constrained by physical limitations based on the muscular structure of the face and skull. These AUs, by definition, are "independent of any interpretation," which means that the code assigned to each contraction/relaxation, or combination of them, is based on form alone and not on the function or meaning of the physical behavior. This system removes subjectivity from the description of facial expressions, making the resulting transcription a more reliable source for research information.

Researchers interested in examining facial expressions used in sign languages have found FACS an effective tool. Charlotte Baker-Shenk was an early adopter and applied it to her research on questions in ASL (1983) and others have used it since (Corina, Belludi, Reily, 1999; Dachkovsky & Sandler, 2009). It has been used with other programs to code sign languages such as SignStream. (Grossman & Shepard-Kegl 2006). In this work we are further extending the coding system by using it with another annotation tool, ELAN.

3. EUDICO Linguistic Annotator

ELAN (EUDICO Linguistic Annotator) is a time-alignable video/audio annotation tool that can be used with different transcription systems with different analytical goals (i.e., from phonetics to discourse). From the online manual, the developers explain that ELAN "is an annotation tool that allows you to create, edit, visualize and search annotations for video and audio data." Developed at the Max Planck Institute for Psycholinguistics, ELAN was designed "to provide a sound technological basis for the annotation and exploitation of multi-media recordings. ELAN is specifically designed for analysis of language, sign language, and gesture, but it can be used by everybody who works with media corpora, i.e., with video and/or audio data, for purposes of annotation, analysis and documentation".

4. Incorporating FACS in ELAN

The challenge the team faced when providing feedback to VCom3D about the avatars expression of 'natural-like' facial behaviors was that the avatars' facial expressions did not appear to be dynamic. Facial expressions are essentially dynamic, in that expressions are created by different parts of the face, and these parts move independently of one another (e.g., eyebrows are raised while the lower face remains static). The Gallaudet team required a way of demonstrating this discrepancy to VCom3D, and agreed that the best way to do so would be to provide them with 1) a more accurate description of the timing of when the parts of the face changed configuration, and 2) a shared 'language' to describe what the parts on the face did. Coding the expressions in ELAN using FACS facilitated these goals.

To begin, the team created a "No Match" dependent tier system to more accurately analyze the annotations previously identified as not matching the model examples/labels provided by VCom3D. For this process three dependent tiers were established under a parent "No Match" tier; annotations for the parent tier were created based on previous transcription of VCom3D "expression" tiers, wherein the token present in the natural data did not match the model in the VCom3D system. On the dependent tiers, the team was able to provide more detailed information about each "No Match" token. Figure 1 is an example of the ELAN template used.

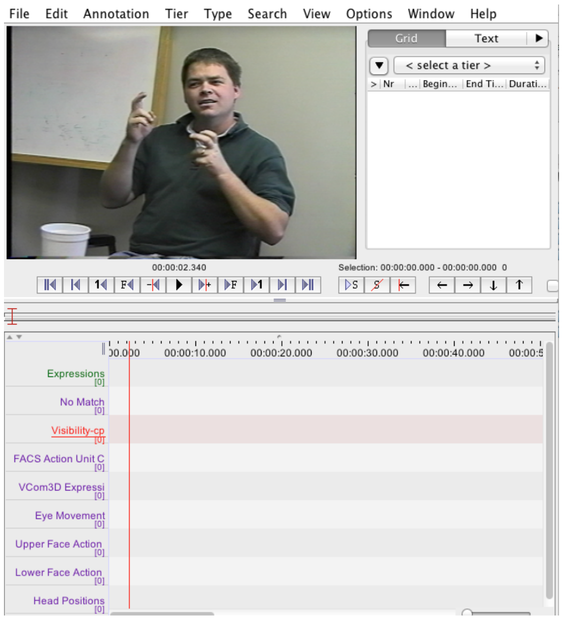

Figure 1: ELAN Template

The first dependent tier used a controlled vocabulary containing all 91 of the VCom3D "expression" labels. Tokens on this tier were coded based on the "closest approximate match" of an expression demonstrated by the VCom3D model and the sign seen in the natural data.

The second dependent tier then provided more detail about what features in the natural data did not match the closest approximate match (which were coded on the first dependent tier). The controlled vocabulary for the second dependent tier was based on material adapted from the FACS manual. Ultimately, there were 15 items in this tier's controlled vocabulary. To create this controlled vocabulary, four broad Action Units (AU) categories were adapted from the FACS manual:

Upper Face Action Units
Lower Face Action Units
Head Positions
Miscellaneous

These 4 categories comprise the initial controlled vocabulary items for this tier; however, combination items were needed for the co-occurrence of features in the data. An additional 10 items were created, from 2 and 3 Action Unit feature combinations, and were created via the use of multi-dimensional tables. Tables 1 and 2 below show the production process for these combination controlled vocabulary codes.

No Match	Upper Face Actions	Lower Face Actions	Head Positions	Miscellaneous Actions	2 features 'No Match'
Upper Face Actions	o	x	x	x	3
Lower Face Actions	x	o	x	x	2
Head Positions	x	x	o	x	1
Miscellaneous Actions	x	x	x	o	Total of 6 combos

Table 1: Two feature combinations

No Match	Upper Face Actions	Lower Face Actions	Head Positions	Miscellaneous Actions	
Upper Face Actions & Lower Face Actions	o	o	x	x	2
Upper Face Actions & Head Positions	o	x	o	x	1
Upper Face Actions & Miscellaneous	o	x	x	o	
Lower Face Actions & Upper Face Actions (redundant)	o	o	x	x	
Lower Face Actions & Head Positions	x	o	o	x	1
Lower Face Actions & Miscellaneous (redundant)	x	o	x	o	
Head Positions & Upper Face Actions (redundant)	o	x	o	x	
Head Positions & Lower Face Actions (redundant)	x	o	o	x	
Head Positions & Miscellaneous (redundant)	x	x	o	o	
Miscellaneous & Upper Face Actions (redundant)	o	x	x	o	
Miscellaneous & Lower Face Actions (redundant)	x	o	x	o	
Miscellaneous & Head Positions (redundant)	x	x	o	o	Total of 4 combos

3 features 'No Match'

Table 2: Three feature combinations

Shown in the above tables, the boxes outlined in red highlight the single instance of 2 and 3 feature combinations (6 and 4, respectively).[1] A third table was generated, wherein 4 feature combinations were shown; however, the combinations in this table are all redundant of each other, since there are only 4 feature categories, and the 4 features co-occurring equates to a full "No Match." Thus, a 15th controlled vocabulary item was added: "Full No Match."

The third dependent tier was also designed to provide detail about which features in the natural data did not match the closest approximate match. This last tier was added to provide information about the visibility of the sign in the data, which was suspected to be the root cause of some "No Match" annotations thus far. The controlled vocabulary for this tier was adapted from the FACS manual. The FACS manual contained 4 codes in its "Miscellaneous" category, which were reallocated for use in this tier. Each of these codes relates information about the visibility of the face in the data. The FACS manual had 4 such codes:

Visibility 70 - Brows and Forehead not visible
Visibility 71 - Eyes not visible
Visibility 72 - Lower Face not visible
Visibility 73 - Unscorable

These four codes were used directly in the controlled vocabulary. Three items were added to accommodate combinations of 'hidden' features. These combinations were generated following a similar multi-dimensional table as that for the preceding tier. In order to aid interpretation of transcript results (i.e. token counts, when images of the data are not provided), an 8th item was added to code whether or not the face in the data had full visibility.

Examples of these controlled vocabularies in application are shown in figure 2 below.

[1] Table items not highlighted red are redundant occurrences of the same feature combinations.

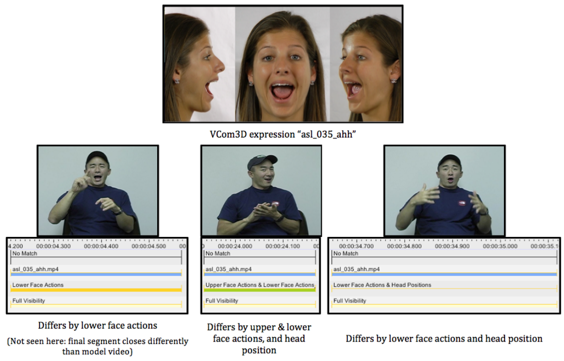

Figure 2: Example of VCom3D model and natural data

The three tokens seen in the middle of figure 1 are examples of instances in which the "closest approximate expression match" is the same for each token (shown in upper half of figure 1, and in ELAN annotations below the tokens coded in blue). They also each match in full visibility (shown on bottom ELAN tier coded in yellow). But, the tokens' FACS tier shows that each token mismatches the VCom3D model in a different way. The left-most token mismatches in the actions of the lower face (coded in gold), particularly in the final segment of production. The center token mismatches in the actions of the lower face, upper face, and head position (coded in green). Finally, on the same dependent tier as the other two tokens, the right-most token mismatches the model in the actions of the lower face and head position. By reviewing portions of ELAN transcripts in this fashion, the team has noted that these dependent tier annotations highlight patterns in the natural data, providing detailed information about the "No Match" tokens.

5. Preliminary Findings

To perform a preliminary test of the No Match dependent tier system, the Gallaudet team chose to code a sample of the natural ASL video data, applying the new tiers to the preexisting ELAN transcripts. The first 60 seconds of a narrative were coded. Within this sample size, the narrative contained a total of 106 facial expression tokens. Of these, the No Match tokens equaled 85 (80% of the total facial expressions). Thus, while at first glance the temporal duration of the preliminary data sample may seem minute at the macro-level, via the coding process (at the micro-level) it becomes clear that this sample size is rich with content and sufficient enough for initial analysis and testing of this new coding system.

As mentioned above, from the sample data 85 No Match tokens were coded for a corresponding 106 facial expression tokens; which equates to 80% of the facial expressions appearing in the natural data not matching those of the VCom3D model. More interestingly, however, are the patterns that emerged in the dependent tiers using the new coding system. First, the coding of the closest approximate match revealed distinct frequencies of 15 different VCom3D "expressions" within the total 85 No Match tokens. Table 3 lists the closest approximate matches and ranks them (left to right) based on their frequency in the data sample.

Total No Match tokens	85				
none	23	asl_008_smile	3	asl_064_pah	3
asl_034_ab	7	asl_044_bop	3	asl_033_regular	2
asl_035_ahh	7	asl_050_fafa	3	asl_040_bah	2
asl_016_relativeclause	5	asl_060_mm	3	asl_047_eee	2
asl_090_disgust	4	asl_063_ooo	3	asl_055_gagaga	2

Table 3: Closest Approximate Match token counts[2]

As can be seen in table 3 above, 23 tokens were identified to have "none" of the VCom3D expressions as their closest approximate match; which equates to 27% of the No Matches. While for those to which an approximate match was identifiable, "ab," "ahh," "relative clause," and "disgust" were the most frequent.

Next, patterns also emerged in the coding of the FACS Action Units tier. Table 4 displays the frequency of the AUs and combinations of AUs that were observed to cause the mismatch between the natural data tokens and their closest approximate matches (VCom3D "expressions").

	Vcom Expressions	FACS Categories	Visibility
No Match	asl_008_smile.mp4	Upper Face Actions & Head Positions	Full Visibility
No Match	asl_008_smile.mp4	Upper Face Actions & Head Positions	Full Visibility
No Match	asl_008_smile.mp4	Upper Face Actions & Head Positions	Full Visibility
No Match	asl_016_relativeclause.mp4	Lower Face Actions & Head Positions	Full Visibility
No Match	asl_016_relativeclause.mp4	Lower Face Actions & Head Positions	Full Visibility
No Match	asl_016_relativeclause.mp4	Upper Face Actions & Head Positions	Full Visibility
No Match	asl_016_relativeclause.mp4	Upper Face Actions & Lower Face Actions	Full Visibility
No Match	asl_016_relativeclause.mp4	Upper Face Actions & Lower Face Actions	Full Visibility
No Match	asl_034_ab.mp4	Upper Face Actions & Lower Face Actions	Full Visibility
No Match	asl_034_ab.mp4	Upper Face Actions & Lower Face Actions	Full Visibility
No Match	asl_034_ab.mp4	Upper Face Actions & Lower Face Actions	Full Visibility
No Match	asl_034_ab.mp4	Upper Face Actions & Lower Face Actions	Full Visibility
No Match	asl_034_ab.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility
No Match	asl_034_ab.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility
No Match	asl_034_ab.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility
No Match	asl_035_ahh.mp4	Lower Face Actions	Full Visibility
No Match	asl_035_ahh.mp4	Lower Face Actions	Full Visibility
No Match	asl_035_ahh.mp4	Lower Face Actions & Head Positions	Full Visibility
No Match	asl_035_ahh.mp4	Lower Face Actions & Head Positions	Full Visibility
No Match	asl_035_ahh.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility
No Match	asl_035_ahh.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility
No Match	asl_035_ahh.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility
No Match	asl_044_bop.mp4	Lower Face Actions	Full Visibility
No Match	asl_044_bop.mp4	Lower Face Actions	Visibility 73 - Unscorable
No Match	asl_044_bop.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility
No Match	asl_050_fafa.mp4	Lower Face Actions & Head Positions	Full Visibility
No Match	asl_050_fafa.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility
No Match	asl_050_fafa.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility
No Match	asl_060_mm.mp4	Head Positions	Full Visibility
No Match	asl_060_mm.mp4	Upper Face Actions & Head Positions	Full Visibility
No Match	asl_060_mm.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility
No Match	asl_063_ooo.mp4	Lower Face Actions	Full Visibility
No Match	asl_063_ooo.mp4	Upper Face Actions & Lower Face Actions	Full Visibility
No Match	asl_063_ooo.mp4	Upper Face Actions & Lower Face Actions	Visibility 73 - Unscorable
No Match	asl_064_pah.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility
No Match	asl_064_pah.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility
No Match	asl_064_pah.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility
No Match	asl_090_disgust.mp4	Upper Face Actions & Head Positions	Full Visibility
No Match	asl_090_disgust.mp4	Upper Face Actions & Head Positions	Full Visibility
No Match	asl_090_disgust.mp4	Upper Face Actions & Head Positions	Full Visibility
No Match	asl_090_disgust.mp4	Upper Face Actions & Lower Face Actions & Head Positions	Full Visibility

Table 4: Frequencies of FACS Action Units and Visibility for "expressions" with 3 or more instances of No Match[3]

In table 4, we can see that for each of the VCom3D expressions there were consistent patterns for how the natural data differed. Namely, the natural data differed by Upper Face Actions, Lower Face Action, Head Positions, or some combination of the three.

Also displayed in table 4 is a portion of the visibility results. This final finding, and perhaps the most definitive thus far, is that 77 of the 85 No Match tokens exhibited full visibility in the natural data (90%). In other words, for identification of the signer's non-manual features and facial expressions, particularly those that did not match the VCom3D model, all were unobstructed visually with the exception if 6, of which only 2 were unscorable. This means that, contrary to our previous supposition, a lack of visibility of the signer's face is not the source of the mismatches we had been noticing up until now. Supporting, more so, that it is the form of the signer's natural facial configurations and movements that are the key to the discrepancies in avatar development.

6. Conclusions

Although this new method of notation has only been applied to a small data set thus far, the team has already been able to find comprehensible patterns within the data that add clarification to the discrepancies previously experienced with the VCom3D notation system. By applying this new set of tiers in the transcription process, the team has been able to identify which model examples/labels are most frequently mismatching with the natural data, which portions of the face are "triggering" the mismatches, and that, despite previous supposition, visibility is not the root of the mismatches.

The representation of non-manual features presents a special set of challenges, and has not received much widespread attention in the field of sign language research. As has been demonstrated in this paper, treating independent non-manual features individually by using FACS in the ELAN transcript allows us to more accurately represent their behavior and better understand their function in language.

7. Acknowledgements

This research was supported by a subcontract agreement from VCom3D under the Department of Education SBIR Phase II Award entitled "Mobile Signing Math Dictionary with Mouth Morphemes" Award number H1335120066, awarded on October 1, 2012 to Vcom3D, Inc.

Thank you also to Sara Malkowski a research assistant on this project as well as students in our LIN 480 Linguistics Research Experience course for their efforts coding and other substantive contributions: Dana Baldiviez, Maryna Epstein , and Eddy Morrison.

8. References

Baker-Shenk, C. (1983). A Microanalysis of the Nonmanual Component of Questions in American Sign Language. PhD diss., University of Cali- fornia–Berkeley.

Castor, A., Pollux, L.E. (1992). The use of user modelling to guide inference and learning. *Applied Intelligence*, 2(1), pp. 37--53.

Corina, D. P., Bellugi, U., & Reilly, J. (1999). Neuropsychological studies of linguistic and affective facial expressions in deaf signers. *Language and Speech*, 42, pp. 307-31.

Dachkovsky , S and Sandler , W. (2009). Visual Intonation in the Prosody of a Sign Language. *Language and Speech* 287 – 314.

Ekman, P., Friesen, W.V., & Hager, J.C. (2002). *Facial Action Coding System: The Manual on CD ROM*. A Human Face: Salt Lake City.

ELAN http://tla.mpi.nl/tools/tla-tools/elan/

Grossman, R. and Shepard-Kegl (2006). To Capture a Face: A Novel Technique for the Analysis and Quantification of Facial Expressions in American Sign Language. *Sign Language Studies*, (6), 3, pp. 273-305.

2 "Expressions" with only one instance are not included here.

3 Tokens coded as "none" not included here.