The Annotation of Gesture Designed for Classroom Interaction

Michael Amory and Olesya Kisselev
The Pennsylvania State University
304 Sparks Building, University Park, PA
mda5004@psu.edu; ovk103@psu.edu

Abstract

In the past decade, the field of Applied Linguistics has witnessed an increased interest in the study of multimodal aspects of language and language acquisition, and the number of multimodal corpora that are designed to investigate classroom interactions, second language acquisition and second language pedagogy is on the rise. The promise is that these digital repositories of video-recordings will be able to take advantage of Corpus Linguistics tools and procedures in order to maximize and diversify analytical capabilities. However, the transcription conventions (i.e., annotation schemas) for multimodal features (such as gestures, gaze, and body movement) that are simple, systematic and searchable are not readily available. The current project focuses on developing an annotation schema for the transcription of gestures, integrating the research traditions of Conversation Analysis, gesture research, American Sign Language, and Corpus Linguistics. The goal of the project is to create a set of conventions that have analytical and descriptive power but are manageable for the transcriber and reader to engage with, as well as systematic in order to allow for searchability. The study utilizes video-recorded data from the Corpus of English for Academic and Professional Purposes developed at the Pennsylvania State University.

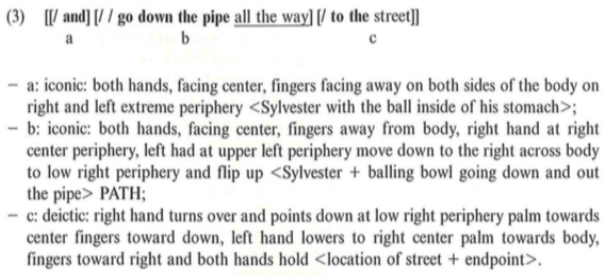
Keywords: Conversation Analysis, Multimodal Corpus Linguistics, Gesture Annotation

In the last decade, the field of Applied Linguistics has been witness to a rise in development of corpora designed to investigate classroom interactions and second language (L2) pedagogy (e.g., Lab School at Portland State University, Reder, 2005). These projects have enabled researchers and language teaching practitioners to conduct cross-case and cross-corpora comparisons on various interactional practices in the classroom environment. One such video-based specialized corpus is the Corpus of English for Academic and Professional Purposes (CEAPP) at the Pennsylvania State University, the focus and testing ground of the present project. In the hopes of creating an important knowledge base upon which practitioners and researchers can draw to identify problems, devise solutions, and enhance efficacy in classroom interactions, CEAPP video-recordings are transcribed using Conversation Analysis (CA) conventions (cf. Jefferson, 2004) and are subjected to CA analyses.

CEAPP functions as a digital repository consisting of approximately 350 hours of classroom interactions that provides corpus resources focusing on the teaching and learning of English as a Second Language (ESL), as well as the New Professoriate Initiatives (NPI) focusing on Science, Technology, Engineering, and Mathematics (STEM) courses. CEAPP does provide basic search capabilities as one can use the search interface to conduct inquiries for transcripts of classroom interactions from a variety of courses and language proficiency levels. For example, one can specify a single or a combination of search criteria, including course type, course level, teaching context, activity type, professor rank, professor education, professor experience, etc. Although CEAPP may be considered a corpus on the basis of being a principled collection of data, it does not possess most of the functionalities of a searchable corpus. We believe, however, that CEAPP will be a significantly more powerful research platform with the addition of Corpus Linguistics (CL) tools, which will yield more automated analyses at both micro-level, i.e., linguistic structures and utterances, and macro-level, i.e., discourse (Walsh, 2013). The ability to annotate data with tags based on CA conventions and other multimodal features and then search the corpus by these tags will significantly enhance a researcher's engagement and profound understanding of data. Of particular interest to the current stage of the project, is the annotation of multimodal components, specifically gestures, which has proven to be a challenging endeavor for both CA and CL.

The study of multimodality has become an area of increasing research interest in the recent decades; recent studies of gesture have created a considerable body of supporting evidence for language's close relationship to bodily movement and argue that gesture and speech are part of a unified system and should not be analyzed separately (McNeill, 1992; Goldin-Meadow, 2005). In these studies, several gesture classification systems have been proposed (i.e., Ekman & Friesen, 1969; Freedman & Hoffman, 1967; McNeill, 1992). However, since these systems were designed with particular research questions in mind and largely used in the study of monologic co-speech gesture language production, they may not be immediately applicable to large databases created and designed for CA or CL research. Consider, for example,

a sample transcript in Illustration 1 below (from Stam, 2014).


(3) [[/ and] [/ / go down the pipe all the way] [/ to the street]]
a b c

- a: iconic: both hands, facing center, fingers facing away on both sides of the body on right and left extreme periphery <Sylvester with the ball inside of his stomach>;
- b: iconic: both hands, facing center, fingers away from body, right hand at right center periphery, left had at upper left periphery move down to the right across body to low right periphery and flip up <Sylvester + balling bowl going down and out the pipe> PATH;
- c: deictic: right hand turns over and points down at low right periphery palm towards center fingers toward down, left hand lowers to right center palm towards body, fingers toward right and both hands hold <location of street + endpoint>.

Illustration 1: Gesture annotation from Stam (2014)

It is not our intent to critique existing methodologies as they have their own analytic purpose and have contributed significant insights to their respective fields. However, they are challenging for the purposes of CA and CL in a number of important ways. Most importantly, annotations as the one presented above usually include a (lesser or higher) degree of interpretation made by the researcher; notice the use of such semantic categories as "iconic" and "deictic" in Stam's transcript. These interpretations vary across studies and research traditions, and may change across time, leading to the lack of systematicity that then leads to issues with searchability. That is, with the lack of a unified, descriptive system to describe the gestures themselves, their location and movement in relation to the body, etc., it becomes more challenging to search across multiple transcripts. Another issue impacting searchability is the lack of simplicity within the transcript: gesture annotation can be difficult for the non-expert reader to understand and for a transcriber to record systematically, in addition to requiring an increasing number of transcription hours.

There do exist systems that are designed for the multimodal annotation of dialogic interactions. One such example is the annotation schema developed by researchers from the Nordic Network on Multimodal Interfaces (MUMIN, Allwood et al., 2004). Having made important strides in creating a general schema for the study of facial expressions and gestures, the researchers paid "particular regard to the role played by multimodal expressions for feedback, turn management and sequencing" (Allwood et al., 2005, p. 1). However, MUMIN is not as of now readily applicable to the transcription of gesture in CEAPP. The system is based solely on Handedness and Trajectory, which is limiting to the descriptive and analytical power of gesture annotation; the researchers themselves recognize that "[t]here are thus a number of ways in which the coding of gesture shapes could be further developed for different purposes and applications" (ibid., p. 12). More importantly, being based on McNeill' (1992) system, MUMIN annotation schema requires the annotator to make an analytical decision regarding the function of the gesture being transcribed: it incorporates McNeill's semiotic categories of deictic, iconic, and symbolic types of gesture and adds interactional semiotic categories of feedback-giving, feedback-eliciting, etc., which we consider to be in opposition with the methodological assumptions of both CA and CL.

Since clear distinction between form and function is key from our perspective, the system proposed in this paper is more closely aligned with Conversational Gesture Transcription System (CoGesT, Trippel et al., 2014), that proposes a feature-based transcription system (ibid., p. 1), i.e. a system where the form of a gestural movement is first described as perceived visually, and only after analysis is assigned a functional gloss and other interpretations. However, in terms of the compromise between simplicity and readability for CA transcripts, CoGesT is less intuitive for the annotator and reader than the system we wish to create for CEAPP.

CEAPP attempts to create a system that is based on the previous gesture research but is grounded in the tradition of Jefferson (1974), a CA analyst who wished to create a methodology to annotate transcripts of spoken interaction that was a compromise between two objectives: to preserve the details of talk (in this case, gesture) as it is actually produced (*description before interpretation*), while at the same time remaining simple enough to yield transcripts that are accessible to a general audience (*simplicity and readability*). In addition, the current project will attempt to take this one step further and offer a systematic transcription of gesture designed for both CA and CL research while taking into account the large database such CEAPP (*searchability*). Thus, a balanced must be reached between the simplicity and readability of multimodal tags through the lens of CA and their searchability and analytic power in a large corpus.

To satisfy the requirements of descriptive power, systematicity and simplicity, we draw upon previous research in gesture, particularly McNeill (1992), as well as adopt parameters and classifications from American Sign Language research and sign language phonology (Valli & Lucas, 2000). The parameters that are considered to be pertinent are the following: *handshape, movement, palm orientation, movement, and trajectory*. In addition, we have added handedness for descriptive purposes. Some parameters are more elaborated upon (e.g., handshape, movement) since previous research has linked handshape and type of movement to cognitive-linguistic categories (e.g. type of movement corresponds with linguistic categories of motion such as PATH and MANNER, see for example Stam, 2008; Cadierno, 2010).

What follows is by no means an exhaustive list of the gesture annotation system. Rather, we present a broad overview of each category with select, representative examples and classifications.

1. Handedness indicates which hand is used in gesturing.

Right Hand (RH)
Left Hand (LH)
Both Hands (BH)

Illustration 2: right-handed gesture

2. Handshape description is based on the complex visual-spatial system (Nakamura, 2002) used in American Sign Language (ASL). Handshape is literally the shape (or shapes) in which we form our hand during the production of a gesture (i.e., hand configuration). The utilization of this system presents us with a purely descriptive account to represent handshapes while trying to avoid implying meaning. In illustrations 2-5 below, a few examples of handshapes are presented. For reference, handshape is indicated by HS. The number or letter follow HS represents the form that the hand has taken. For reference, these numbers or letters are based upon ASL.

Illustration 3: HS-1

Illustration 4: HS-V

Illustration 5: HS-O

3. Palm orientation (palm) while making the handshape.

Left
Right
Up
Down
Front
Back

Illustration 6: palm-up

Illustration 7: palm-front

4. Location of handshape in relation to the body. This parameter is a modified and significantly simplified account of McNeill's (1992) original proposition, which suggested 21 different locales around in space around the body (such as Extreme Periphery, Lower Left, Center-Center, etc.)

Center (C)
Left (L)
Right (R)
Upper (UP)

Upper left (UL)
Upper right (UR)
Lower (LW)
Lower left (LL)
Lower right (LR)

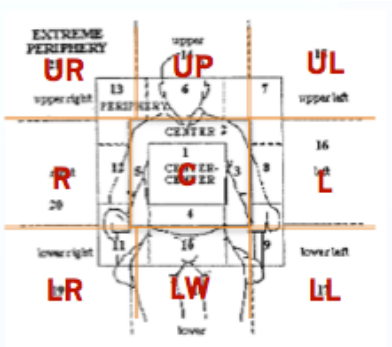


Illustration 8: Simplification of McNeill (1992)

5. Movement of the handshape and/or arm. This parameter will be used to represent the trajectory and type of movement.

Single movement (SM)
Repetitive movement (RM), 3x, 2x, etc.
Clockwise (CW)
Counter-clockwise (CCW)
Sinuous (SI)
Straight (ST)
Wrist rotation/movement (WR)

6. Trajectory of the gesture in relation to the body, indicating initial and final position.

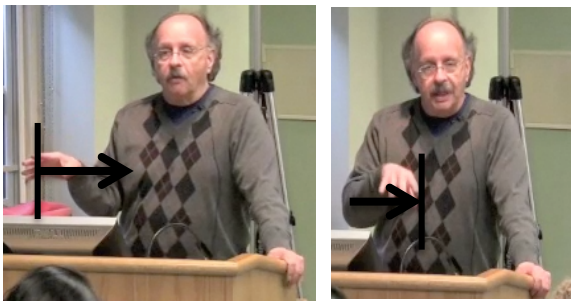


Illustration 9: Initial to final position; from R to C

Illustration 9 above is a depiction of the gesture and its trajectory that will be incorporated in the below example of a gesture annotated alongside the co-occurring utterance in Excerpt 1. The gesture made in Illustration 9 is HS-bent-5. Excerpt 1 on the following page represents how the annotation of gesture may be incorporated into a CA transcript using the annotation schema.

Excerpt 1: Example of CA transcript

8	*TEA:	{>th- the< gestures show up,
9		{RH HS-bent-5 palm-down moves from C to R
10	*TEA:	{in interesting wa:yz.
11		{RH HS-bent-5 palm-down moves from R to C

Conclusion

With the enhancement of technology, digital repositories such as CEAPP will be able to take advantage of CL tools and procedures in order to maximize and diversify analytical capabilities. Utilizing coding schemes for the annotation of multimodal corpora, in our case tagging of gestures and incorporation of searchable functions, may facilitate cross-case studies, cross-corpora, and longitudinal analyses. At the same time, CA methodology has a lot to offer to corpora studies of speech and communication, especially in terms of accounting for multimodality of naturally-occurring speech. With a commitment to "naturalistic inquiry" (Schegloff, 1997, p. 501) and rigorous transcription procedures, CA can provide a theoretical and practical framework to transcribing and analyzing video recordings along with transcripts of these recordings in a systematic, simple and yet descriptive way. This would allow researchers to maintain a clear distinction between form and function in gesture transcription and annotation. By combining CA and CL approaches to data transcription, coding and analysis, can reveal new insights into the relationship between interaction patterns, language use, and learning (O'Keeffe & Walsh, 2012; Walsh, 2012).

Bibliographical References

Allwood, J., Cerrato, L., Dybkær, L., Jokinen, K., Navarretta, C. and Paggio, P. (2004). The MUMIN multimodal coding scheme. Technical report available at www.cst.dk/mumin/stockholmws.html.

Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C., & Paggio, P. (2005). The mumin annotation scheme for feedback, turn management and sequencing. In *Proceedings from the Second Nordic conference on Multimodal Communication. Gothenburg Papers in Theoretical.*

Bucholtz, M. (2000). The politics of transcription. *Journal of pragmatics*, *32*(10), pp. 1439-1465.

Cadierno, T. (2010). Motion in Danish as a second language: Does the learner's L1 make a difference? In Z. Han & T. Cadierno (Eds.), *Linguistic Relativity in SLA* (pp. 1-33). Bristol: Multilingual Matters.

Ekman, P. & Friesen, W.V. (1969). The repertoire of

nonverbal behavior: Categories, origins, usage, and coding. *Semiotic, 1*, pp. 49-98.

Freedman, N., & Hoffman, S. P. 1967). Kinetic behavior in altered clinical states: Approach to objective analysis of motor behavior during clinical interviews. *Perceptual and Motor Skills*, 24, pp. 527-539.

Goldin-Meadow, S. (2005). *Hearing gesture: How our hands help us think*. Harvard University Press.

Jefferson, G. (1974). Error correction as an interactional resource. *Language in Society*, *3*(2), pp. 181-199.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.) *Conversation Analysis: Studies from the first generation* (pp. 13-23). Philadelphia: John Benjamins.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

Nakamura, K. (2002). Sign Language Linguistics. About American Sign Language. Deaf Resource Library. Retrieved 12, Feb. 2016 <http://www.deaflibrary.org/asl.html>.

O'Keeffe, A., & Walsh, S. (2012). Applying corpus linguistics and conversation analysis in the investigation of small group teaching in higher education. *Corpus Linguistics and Linguistic Theory*, *8*(1), pp. 159-181.

Reder, S. (2005). The "lab school.". *Focus on Basics*, *8*, pp. 1-7.

Schegloff, E. A. (1997). Whose text? Whose context?. *Discourse & Society*, *8*(2), pp. 165-187.

Stam, G. (2008). What gestures reveal about second language acquisition. *Gesture: Second Language Acquisition and Classroom Research*, 231.

Stam, G. (2014). Further changes in L2 thinking for speaking? In C. Mller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and S. Tessendorf (Eds.) *Body Language Communication: An international handbook on multimodality in human interaction*. *Volume 2*, 1875-1886. Handbooks of Linguistics and CommunicationScience/Handbcher zur Sprach- und Kommunikationswissenschaft. Berlin, New York: Mouton De Gruyter.

Trippel, T., Gibbon, D., Thies, A., Milde, J. T., Looks, K., Hell, B., & Gut, U. (2004, May). CoGesT: a Formal Transcription System for Conversational Gesture. In *LREC*.

Valli, C., & Lucas, C. (2000). *Linguistics of American sign language: An introduction*. Gallaudet University Press.

Walsh, S. (2013). Corpus linguistics and conversation analysis at the interface: Theoretical perspectives, practical outcomes. In *Yearbook of Corpus Linguistics and Pragmatics 2013* (pp. 37-51). Springer Netherlands.