Annotation of Mouth Activities with iLex

Thomas Hanke
Institute of German Sign Language and Communication of the Deaf,
University of Hamburg
thomas.hanke@sign-lang.uni-hamburg.de

Abstract

This paper describes the support for mouth activity annotation provided by the iLex annotation workbench on a holistic level connected to the lexical database, on a feature level, as well as in the context of semi-automatic annotation.

Keywords: Annotation, mouthing, mouth gesture, lexical database

1. Introduction

In a purely bottom-up approach an annotation practice used for mouth activities would try to describe the phenomena and leave it to a second step to classify (e.g. between mouthing and mouth gestures) and relate (e.g. to spoken language words) (cf. Keller, 2001). For practical reasons, however, the first step is often skipped, and separate coding systems are applied to what is categorised either as mouthing derived from spoken language or mouth gesture where there is no obvious connection between the meaning expressed and any spoken language words expressing that same meaning. This happens not only for time (=budget) reasons, but also because it is difficult for coders to describe mouth visemes precisely if the sign/mouth combo already suggests what is to be seen on the mouth. While there are established coding procedures to avoid influence as far as possible (like only showing the signer's face, provided video quality is good enough), they make the approach very time-consuming, even if not counting quality assurance measures like inter-transcriber agreement. Some projects undertaken at the IDGS in Hamburg therefore leave it with a spoken-language-driven approach: The mouth activity is classified as either mouth gesture or mouthing, and in the latter case the German word is noted down that a competent DGS signer "reads" from the lips, i.e. that word from the set of words to be expected with the co-temporal sign in its context that matches the observation. Standard orthography is used unless there is a substantial deviation. For mouth gestures, holistic labels are used. These two extremes span a whole spectrum of coding approaches that can be used for mouth activities. We present different aspects of how iLex, the Hamburg sign language annotation workbench, supports the whole range of solutions from more time-series-like systems to those evaluating co-occurrence and semantic relatedness, from novice-friendly decision trees to expert-only modes to support semi-automatic annotation.

2. iLex Background

Unlike other transcription environments, iLex does not follow a document-centric approach, but keeps all annotation in a relational database. Consequently, tags are not simply text, but are structured database entities themselves, such as tokens describing an instance of a type. This allows the user to immediately access other tokens of the same type as (phonetic and context) data, as a video snippet, or an avatar performance. The complete integration of a lexical database into the annotation process in our view is crucial when transcribing a language not having an established written form.[1]

Mouth activities, are not part of the token records, but are annotated as text tags on a separate tier.[2] Being text tags, mouthings are not considered as instantiations of spoken language lexemes, the tag is a mere form description. However, this does not mean that mouth activity annotation does not profit from the integrated approach:

3. Mouthing in the Lexical Database

In the iLex lexical database, types have a field to store a default mouth activity typically co-occurring with the sign. In some cases, certain mouth gestures are an integral part of the sign, these would be stored here. In the case of lexicalised form-meaning combinations[3], one or more mouthings can be stored here that typically occur in this context.

As these mouthings are good candidates for the mouth tier tags overlapping with a certain token, iLex provides easy access to them via a context menu to create the mouth tag.

The iLex database can be set up to provide extra suggestions here, e.g. all mouthings that the informant currently being transcribed has already used in combination with the token's type.

Only if the observed mouth activity does not match with any of the suggestions, the user needs to open a specialised editor in order to describe the observation (cf. section 4).

[1] A more detailed description of the iLex workbench can be found in Hanke, 2002, Hanke/Storz, 2008 and Hanke et al., 2010.

[2] The reason for this is evident: Mouthings can stretch over more than one sign (token).

[3] For details on the type hierarchy implemented in iLex and how it is explored for modelling the sign lexicon, cf. Konrad et al., 2012.

At the same time, mouthings are an essential bit of information in the lemma revision process: When a few tokens for a lexicalised form-meaning combination co-occur with mouthings that derive from spoken language words not semantically related, this might be an indication that they actually belong to another type, even if they share the same (manual) form.

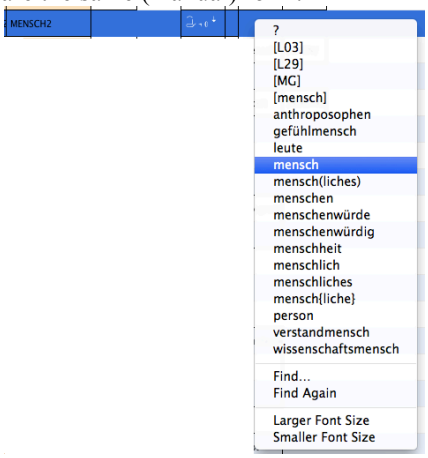

Figure 1: Context menu for mouthing to be cotemporal with the sign MENSCH2

4. Mouth Editor

iLex currently supports three conventions how to store mouthings as text: Orthography, IPA, and SAMPA.

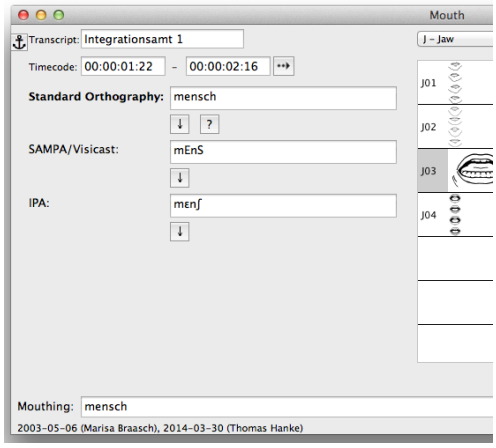

Figure 2: Left side of mouth activity editor: Mouthing

As visemes are equivalence classes of visually indistinguishable phonemes, any member of the class can represent the viseme, allowing visemes to be encoded by a subset of IPA. Whether one always uses the same IPA letter for one class or keeps with the original phoneme, is irrelevant for describing the observation, but certainly makes a difference when testing two annotations for equalness. SAMPA (cf. Gibbon et al., 1997) was suggested in the context of the ViSiCAST and eSIGN projects (cf. Hanke, 2004) to describe visemes as SAMPA text is (was) easier to handle (being ASCII text) than IPA. However, for the purpose of viseme labelling, SAMPA can simply be considered a coding variation of IPA.

As said in the introduction, using spoken language orthography seems weird to describe visemes, but has its advantages, not limited to the transcribers' convenience. The pronunciation data in iLex allow the program to derive the viseme sequence from the orthography entered. For German, iLex also manages to derive the viseme gestalt for abbreviated mouthings from the abbreviated orthography as well as to compute the viseme gestalt for compounds.

iLex allows the user to annotate mouth gestures on separate tiers or in line with mouthings. In the latter case which seems preferable to us, some distinguishable code set is needed to tell mouth gesture codes apart from mouthing. For this reason, we use the convention to include mouth gesture codes in square brackets.

A specialised editor for using a mouth gesture code set introduced in the ViSiCAST project (Hanke et al., 2001) is implemented in iLex. As these codes are rather arbitrary, is is most important that the system supports the user by showing a textual and video description for the code selected.

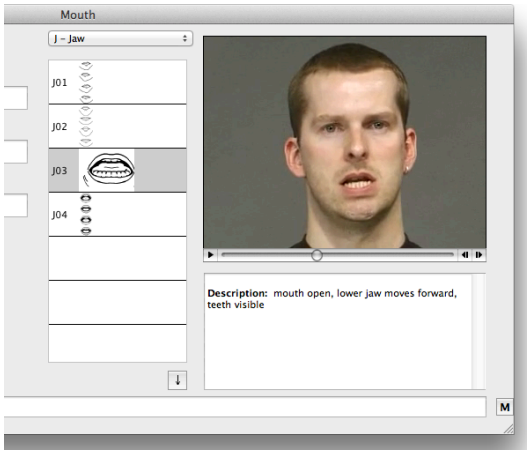

Figure 3: Right side of mouth activity editor: Mouth gestures

Nevertheless, as some mouth gestures occur very rarely, iLex also offers an experimental "expert system" to determine the right code: Following the ideas of Sutton-Spence/Day (2001), the user has to answer a number of relatively easy questions on his/her observation, and system then provides the code.

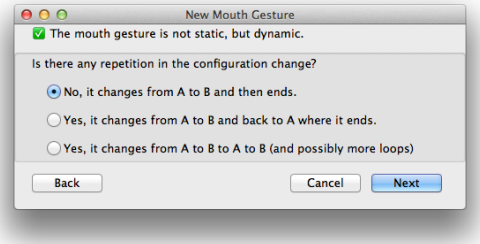

Figure 4: "Expert system" for entering mouth gestures by answering a series of questions

5. Tag Alignment

Unless one is interested in the exact timing of mouth activities, iLex allows the user to set up the mouth tier to depend on the token tier in order to save time: Tag boundaries are then shared between these tiers, but mouth tags can still span several token tags.

For DGS, we observe some signers who (sequentially) combine mouthing and mouth gesture within one manual

sign. In the case of separate tiers for mouthings and mouth gestures, this means that a sub-sign granularity is necessary, i.e. the tiers have to be set up not to depend on each other. For one common tier for mouthings and mouth gestures, codes can simply be concatenated into one tag spanning the whole sign duration.

6. Compatibility with the eSIGN Approach

The main components of the eSIGN software are an avatar system that is able to sign from phonetic data (cf. Elliott et al., 2004) and an editor that allows scripting of such avatar performances (cf. Hanke, 2004). In order to avoid re-writing the necessary phonetic information, the editor works with a local database or links into the iLex database. However, from within iLex it is also possible to save a transcript as an eSIGN document. Obviously, for this to work the transript needs to contain all necessary phonetic descriptions. With respect to mouthings and mouth gestures, this means that the data is coded in one of the aforementioned systems. If orthography is used, the conversion relies on available pronunciation data. If another coding system is used for mouth gestures, the user can still provide a mapping onto the eSIGN formats for the conversion to work.
For the iLex user, this approach has the advantage that an anonymised version of a sign performance can be created with minimal effort.

7. Feature-Level Annotation

For detailed phonetic analysis iLex provides another mechanism than simple textual tags: Binary features. By assigning a closed vocabulary to a binary features tier, iLex prompts the user with a list of all the features (the elements of the vocabulary) in order to check those that apply for the tagged time stretch. This approach still works with rather large sets of features when it is no longer feasible to reserve one tier per feature.

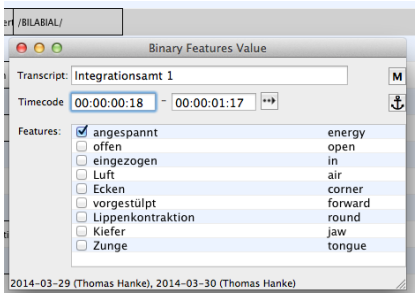

Figure 5: The feature angespannt=tense alone from the full set displays in the transcript as /BILABIAL/ from Bergman/Wallin's reduced set

Here we show an example using the feature set from Bergman/Wallin, 2001.
The display in the transcript can just be the selected features or any function thereof. In the example, the display is automatically computed from the input features by means of a user-provided mapping table, in this case implementing the Bergman/Wallin reduced feature set.

8. Towards Semi-Automatic Annotation

While lipreading is known to be a hard problem both for humans and automatic systems, it is a lot easier to identify the mouthing given the identity of the sign coarticulated as that sign narrows down the search space to only a couple of probable mouthings. We currently experiment with feature vectors obtained from short-range 3D sensors imported into iLex transcripts in order to first determine whether there is mouth activity during a sign, and if so, which of the candidate mouthings best fits with the feature vectors observed. Even when applying some thresholding, this approach increases the risk that unusual sign/mouthing combinations remain undetected. It therefore remains to be seen if this automation is a time saver when a certain annotation quality is to be guaranteed.

9. Acknowledgements

This publication has been produced in the context of the joint research funding (DGS Corpus) of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.
The research leading to these results has also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 231135 (Dicta-Sign).

10. References

Bergman, B., Wallin, L. (2001) A preliminary analysis of visual mouth segments in Swedish Sign Language. In Boyes Braem, P. & Sutton-Spence, R. (eds.): *The hands are the head of the mouth*. Hamburg: Signum, pp. 51-68.

Elliott, R., Glauert, J., Jennings, V., Kennaway, R. (2004). An overview of the SiGML notation and SiGMLsigning software system. In: O. Streiter, O. & Vettori, C. (eds.): *From SignWriting to Image Processing. Information techniques and their implications for teaching, documentation and communication. Proceedings of the Workshop on Representing and Processing of Sign Languages.* 4th International Conference on Language Resources and Evaluation, LREC 2004, Lisbon, Portugal. Paris: ELRA, pp. 98–104.

Gibbon, D., Moore, R., Winski, R. (eds., 1997): *Handbook of Standards and Resources for Spoken Language Systems*. Berlin: de Gruyter.

Hanke, T. (2002). iLex. A tool for Sign Language Lexicography and Corpus Analysis. In González Rodriguez, M. & Paz Suarez Araujo, C. (eds.): *Proceedings of the Third International Conference on Language Resources and Evaluation. Las Palmas de Gran Canaria, Spain*; Vol. III. Paris: ELRA, pp. 923–926.

Hanke, T. (2004). Lexical Sign Language Resources: Synergies between Empirical Work and Automatic Language Generation. Paper presented at the Fourth

International Conference on Language Resources and Evaluation, LREC 2004, Lisbon, Portugal.

Hanke, T., Langer, G., Metzger, C. (2001). Encoding non-manual aspects of sign language. In Hanke, T. (ed.), ViSiCAST Report D5.1: Interface Definitions.

Hanke, T., Storz, J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In Crasborn, O., Efthimiou, E., Hanke, T., Thoutenhoofd, E.D, Zwitserlood, I. (eds.): *Construction and Exploitation of Sign Language Corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages.* 6th International Conference on Language Resources and Evaluation, LREC 2008, Marrakech, Maroc. Paris: ELRA, pp. 64–67.

Hanke, T., Storz, J., Wagner, S. (2010). iLex – Handling multi-camera recordings. In Dreuw, P., Efthimiou, E., Hanke, T., Johnston, T., Martínez Ruiz, G., Schembri, A. (eds.): *Corpora and Sign Language Technologies. Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages.* 7th International Conference on Language Resources and Evaluation, LREC 2010, Valletta, Malta. Paris: ELRA, pp. 110–111.

Keller, J. (2001). Multimodal representations and the linguistic status of mouthings in German Sign Language (DGS). In Boyes Braem, P. & Sutton-Spence, R. (eds.): *The hands are the head of the mouth*. Hamburg: Signum, pp. 191–230.

Konrad, R., Hanke, T., König, S., Langer, G., Matthes, S., Nishio, R., Regen, A. (2012). From form to function. A database approach to handle lexicon building and spotting token forms in sign languages. In Crasborn, O. Efthimiou, E., Fotinea, S.-E., Hanke, T., Kristoffersen, J., Mesch, J. (eds.): *Interaction between Corpus and Lexicon. Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages.* 8th International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey. Paris: ELRA, pp. 87–94.

Sutton-Spence, R., Day, L. (2001). Mouthings and mouth gestures in British Sign Language (BSL). In Boyes Braem, P. & Sutton-Spence, R. (eds.): *The hands are the head of the mouth*. Hamburg: Signum, pp. 69–85.