An annotation scheme for the linguistic study of mouth actions in sign languages

Onno Crasborn & Richard Bank
Radboud University Nijmegen, Centre for Language Studies
P.O. Box 9103, NL-6500 HD Nijmegen, The Netherlands
E-mail: o.crasborn@let.ru.nl, r.bank@let.ru.nl

Abstract

This paper describes the annotation scheme that has been used for research on mouth actions in the Corpus NGT. An orthographic representation of the visible part of the mouthing is supplemented by the citation form of the word, a categorisation of the type of the mouth action, the number of syllables in the mouth action, (non)alignment of a corresponding sign, and a layer representing some special functions. The scheme has been used for a series of studies on Sign Language of the Netherlands. The structure and vocabularies for the annotation scheme are described, as well as the experiences in its use so far. Annotations will be published in the second release of the Corpus NGT annotations in late 2014.

Keywords: sign language, annotation scheme, transcription, non-manual features, mouth actions, mouthings, mouth gestures

1. Goal

This paper aims to describe the annotation scheme that has been developed for a series of studies of mouth actions in Sign Language of the Netherlands (NGT), based on the Corpus NGT (Crasborn, Zwitserlood & Ros, 2008; Crasborn & Zwitserlood, 2008). These studies are targeted at achieving a better understanding of the role of the mouth as an articulator in NGT, with a focus on mouth actions that consist of or are derived from spoken language words ('mouthings'). While it is clear that such mouthings form a case of simultaneous code mixing, dubbed 'code blending' by Emmorey et al. (2005), it has only recently been argued that mouthings form an integral part of deaf communication in the Netherlands (Bank et al., 2013). They are used in virtually every utterance by every user of the language (Bank et al., submitted).

Psycholinguistic studies have demonstrated that deaf people are proficient lip-readers (e.g., Auer & Bernstein, 2007), and it is likely that this information contributes to successful interaction between deaf people also when they use sign language as their primary and preferred mode of communication. While the nature and function of mouth actions have received considerable attention in the sign language literature for a variety of (primarily European) languages (cf. the contributions to Boyes Braem & Sutton-Spence, 2001), no large-scale corpus studies had been performed until recently.

To be able to study the various properties of mouth actions in a corpus, we devised an annotation scheme that systematically separates form from meaning, and that aims to increase efficiency by using Dutch orthographic representations rather than a visual phonetic representation in terms of 'visemes' for the basic transcription layer.

2. The annotation scheme

In this paragraph, we describe the six tiers that we use for every signer in an ELAN annotation file. The transcription (par. 2.1) is independently aligned, while the other tiers containing annotations to the transcription are dependent on this parent tier. This leads to the tier structure displayed in Figure 1.

Mouth	par. 2.1
MouthLemma	par. 2.2
MouthType	par. 2.3
MouthSpr	par. 2.4
MouthSyll	par. 2.4
MouthAdd	par. 2.5

Figure 1: Tier structure for mouth actions

In section 3, we will further discuss how this structure is further implemented in the Corpus NGT.

2.1 Transcription

2.1.1. Preliminary considerations

The start of any investigation into mouth actions will be based on a description of their forms. This immediately leads to problems, as there is no standard transcription system that can be used. One option is to focus purely on the visible properties of articulations, using a classification of the amount of lip rounding, lip opening, and visibility of the tongue, for instance. This appears attractive as it is these properties that are accessible in deaf communication, any possible acoustic accompaniments not being perceivable to deaf people. Although proposals for such 'viseme' categories have been proposed in the literature (see Massaro, 1998; Cappalletta & Harte, 2012; Nonhebel et al., 2004), they lead to a description that in a sense is true to the function of the forms, but that is hard to read. The same holds for a detailed articulatory transcription of mouth actions by use of the action units available in the Facial Action Coding System (FACS; Ekman & Friesen, 1978).

As has become clear from earlier research, the majority of mouth action tokens are mouthings, articulations that consist of (parts of) spoken words. It is thus attractive to somehow use knowledge of speech in the transcription of mouth actions, if only for mouthings.

We know however that any attempt at speech reading involves a lot of interpretation, all aimed at reconstructing words from a spoken language from a small number of visible contrasts. Only a subset of the phonological distinctions in a spoken language has a visible correlate. For vowels, for instance, lip rounding and to a limited extent also tongue/jaw height can be visually perceived, but front-back distinctions in vowels are almost impossible to perceive visually. Thus, if we would use a phonetic or orthographic transcription of a spoken language, we need to make a lot of inferences about what the signer might be saying, on the basis of relatively little phonetic evidence. Comparing the meaning of the perceived mouthings with the co-occurring sign may help in deciding on the transcription, but it may also be misleading.

A different problem with using a transcription system that is based on a representation of the spoken language is that not all mouth actions can be related to spoken language words. In most, if not all sign languages, not only mouthings but also mouth gestures are used (papers in Boyes Braem & Sutton-Spence, 2001; Crasborn et al., 2008). These mouth gestures are by definition not composed of (parts of) spoken words, and may include a variety of articulations (see Crasborn et al., 2008, and Woll, 2001 for discussion). Transcribing them by using a system that is made for speech creates the false suggestion that mouth gestures have always somehow evolved from spoken language words.

Despite these drawbacks, we decided to use an orthographic representation of the spoken language (primarily Dutch, in our case) to transcribe mouth actions. The most powerful argument in favour of this choice is efficiency: not using (a visual version of) a phonetic notation like IPA but using spoken language orthography saves enormous amounts of time during the annotation phase, and the same holds for the exploitation phase. Because of the good readability of orthographic transcriptions as compared to regular phonetic (let alone visual phonetic) transcriptions, the chances that the information about mouth actions will be taken into account in a variety of future studies based on our corpora, orthographic transcriptions are also to be preferred from the point of view of the general user of corpus data. Based on our research findings for NGT that will be briefly discussed in section 4 below, we argue that in addition to glosses and a sentence-level translation, a transcription of mouth actions should be a basic layer of annotation that is needed for any sign language corpus.

The arguments relating to efficient annotation and efficient exploitation are rather similar in nature to the arguments for using a gloss representation for manual signs. Although spoken language glosses have all kinds of disadvantages (including the representation in another language), they are unrivalled in their usability (Johnston, 2010).

Aside from these practical considerations for the corpus annotator and corpus user, filling in details of spoken language articulations that cannot be perceived visually is not all that unnatural: it is what deaf speechreaders do all the time, and are highly proficient at (Woll, 2012). Where (deaf and hearing) communicators are constantly using limited visual information to arrive at an interpretation of what is being said (a process not unlike auditory speech perception in noisy circumstances or in the case of fast speech, for instance), it is important to keep the task of transcription in mind when we annotate mouth actions for corpus annotation. The goal here is not to correctly lemmatise the spoken word, but merely to transcribe the parts of spoken language words that the annotator observes, or in the case of mouth gestures, to arrive at a consistent written representation of the visible form irrespective of any possible spoken language origin. More concretely, what we propose to use for the transcription of mouthings is to only include the segments or syllables that are actually produced, and not any deleted segments or syllables. Reference to the spoken language lemma that the articulation is hypothetically an instance of can be made on the Lemma tier (see section 2.2 below).

2.1.2. Conventions

Mouth action transcriptions are made on a tier called 'Mouth'. Articulations that are perceived as being (fragments of) spoken language words (mouthings) are written in lowercase without any special markers. All other mouth actions (any type of mouth gesture) are put between single quotation marks ('…'). If a mouth gesture cannot be easily described in terms one or more spoken language segments, we use a phonetic description of the mouth articulation between pipes (|…|). This set of descriptors was based on what was developed for the ECHO project (Nonhebel et al., 2004), and adapted on an ad hoc basis.

Acoustic correlates of the mouth action such as phonation are not annotated. We acknowledge that for studies on code mixing, for instance, this could be important information. We suggest that this type of information could best be annotated on a separate tier, with conventions to be established in accordance with the purpose of a specific research goal.

As on other tiers used in the Corpus NGT, uncertainty about the correct representation can be labelled with a single question mark following the transcription. As with manual signs, false starts are prefixed with a tilde symbol (~).

Especially in the case of mouth gestures, the nature of the transcriptions will be influenced by the research findings on this topic for the language at hand (whether in linguistic publications or implicit in dictionary representations or teaching materials). While consistency will be difficult to achieve in the absence of a vocabulary of mouth gestures, the creation of such a vocabulary can be the result of multiple revisions of the set of transcriptions created by a variety of annotators in a first annotation pass. The ECHO conventions for mouth gestures referred to above may serve as a basis for this, but are in need of an evaluation and possibly adaptation,

as they have never been used for a large-scale corpus, as far as we know.

2.2 Lemma

As was already referred to above, the MouthLemma tier is a child tier of the transcription of the Mouth tier, and is the place where the presumed uninflected lemma can be notated of which the observed mouthing is an instance. By using a lemma rather than a full (inflected) form of the spoken word, we stay clear from any overinterpretation of (the morphological specificity of) the mouthing.

The lemma information allows for the searching for mouth actions based on a spoken word type, and will thus facilitate the extraction of various instantiations of the word, whether inflected or not inflected and no matter how reduced or repeated (see section 2.4 below) a Mouth token may be. For this reason, it would be advisable to include a lemma annotation for all mouth annotations, also when they do not differ.

2.3 Classification

On the tier MouthType, we classify the mouth action transcribed on the Mouth tier. We adopt the five-part classification proposed in Crasborn et al. (2008), distinguishing the following categories:

M	Mouthing
E	'Empty' mouth gesture: a lexicalised phonological component of a sign that is not derived from a spoken word
A	Adverbial mouth actions, lexicalised independently of a manual sign
4	'Mouth for mouth' actions: instances where the mouth represents the mouth (as in pantomiming drinking or chewing)
W	Whole-face actions that include a specific mouth articulation, as in affective facial expressions

Figure 2: Types of mouth actions

In addition to these five main types, the Mouthing category is further specified into five subtypes, presented in Figure 3.

M	Regular mouthing
M-back	Mouthing used as backchannel signal
M-add	Mouthing that is not related to a manual sign but temporally overlaps with manual signs.
M-solo	Mouthing that does not overlap with manual signs
M-spec	Mouthing that is co-articulated with a manual sign that serves to specify the semantics of the manual sign

Figure 3: Types of mouth actions for different uses of mouthings

This latter subdivision has arisen in the context of our investigations into NGT mouthings, briefly discussed in section 4. A similar investigation into mouth gestures is likely to lead to a further specification of the four types of mouth gestures listed in Figure 2 (see e.g. Sandler's (2009) category of 'iconic mouth gestures').

2.4 Phonetic properties

Two types of phonetic properties are encoded each on their own tier. First of all, the alignment of the mouthing with the manual glosses is characterised on the MouthSpr tier ('Mouth spreading', following the description of spreading as a prosodic process in Sandler, 2006). As in feature spreading in spoken language segmental phonology, spreading refers to the phenomenon that certain articulatory features may be lengthened to co-occur not only with their source, but also with neighbouring elements. In the case of spreading mouthings, mouthings that have a clear 'source' sign with which the mouthing semantically overlaps are articulated in such a way that they also overlap with the preceding or following sign(s).

The annotation on the MouthSpr tier contains information on the glosses that overlap with the mouth annotation. Angled brackets are used to encode the direction of spreading (< for regressive, > for progressive). For example, the MouthSpr annotation 'BIER > DRINKEN', together with the Mouth annotation *bier* 'beer', means that the mouthing that accompanies the manual sign BEER is either lengthened or maintains it final state so long as to also cover the manual sign DRINKEN 'to drink'. Signers are usually not maximally synchronised in their articulation of sign/mouth pairs, so MouthSpr annotations should not be applied every time that there is a single-frame difference in start or end, irrespective of the duration of the actions and/or the signing speed, for instance. In our own investigations, a mouthing is categorised as spreading over an adjacent sign when it overlaps that sign with at least 50% or 10 or more video frames, whichever applies first.

A second type of phonetic information can be encoded on the MouthSyll tier. It is used to specify the number of syllables of the observed mouth articulation. For mouthings, the number of syllables of the visible word would be transcribed, while for mouth gestures, if countable, the number of cycles of the articulation would be encoded. We have not yet used this tier for our ongoing investigations, but it is devised to study the alignment of manual and oral actions. There are cases in our data where the first syllable of mouthings is reduplicated, seemingly to correspond to the number of movement cycles (syllables) in the manual sign. To investigate the hypothesis that 'the hand drives (the prosody of) the mouth', systematic annotation of the MouthSyll together with the number of movements on the 'NOM' tier (a child of the gloss tiers in the Corpus NGT) will be needed.

2.5 Semantic role

While in our data most mouthings appear to be clearly linked to manual signs both in terms of their semantics (typically overlapping with, if not equal to, that of the sign) and in terms of their timing (typically being co-articulated), there are also mouthings that cannot be analysed as linked to a manual sign. We call these 'added mouthings', as they add an element to the semantics of the whole utterance (rather than specifying the semantics of an individual sign). Solo mouthings (specified as such on the MouthType tier, see Figure 3), have the same function as added mouthings but do not overlap with manual signs. They occur often at the start or end of a signed phrase, before the signing starts or after the signing has ended.

In order to efficiently analyse these utterances, the annotations on the MouthAdd tier consist of a string of manual glosses (ignoring differences between one-handed and two-handed signs and various types of two-handed constructions) followed by a string of mouthings.

Although these annotations are made on sentence level or phrase level, they can still be rather short. For example, utterances like *BEGINNEN begin maar* 'START start go-ahead' are not uncommon.

3. Application of the scheme to the Corpus NGT

We are using the tier structure described above for annotating the Corpus NGT with the ELAN annotation tool. In order to systematically separate annotations for the two signers in the dialogues, we create a double set of tiers, one set per participant in the dialogue. The tiers are suffixed by "S1" and "S2" for the two signers, a system that is used throughout the Corpus NGT and that could easily be adapted for multilogues. A participant tag (S001, S002, ..., S092) for each tier makes it possible to uniquely link each annotation to an individual signer.

The two tiers are 'linked' by having the same 'linguistic type' property in the ELAN documents. This linguistic type is an obligatory specification for each tier, and is in turn specified among other things for its independent or child status, and in the latter case, for the name of the parent tier and the nature of the relation of (one or more) annotations on the child tier to an annotation on the parent tier. In the tier hierarchy outlined in Figure 1 above, the Mouth tiers are independent tiers, not having a parent tier to which they are associated, while all other tiers are child tiers of a Mouth tier. The linguistic types of these child tiers are all specified with the restriction 'symbolic association', meaning that there is a one-on-one relation between child annotation and parent annotation, and that the child annotations cannot be independently aligned with the time axis. Figure 4 presents the names of the linguistic types for the six mouth tiers. Following the conventions for the Corpus NGT, tier names have initial capitals for each word, while linguistic types only use lowercase in combination with underscores to separate words. These conventions help to highlight the distinction between tiers and types both in ELAN and when working with the XML code in the ELAN document.

Tier name	**Linguistic Type**
Mouth	mouth
MouthLemma	mouth_lem
MouthType	mouth_type
MouthSpr	mouth_spr
MouthSyll	mouth_syll
MouthAdd	mouth_add

Figure 4: Tiers and their linguistic types in ELAN

4. Use of the annotation scheme in recent and on-going research

The above annotation scheme has been developed for a series of studies on mouth actions in NGT, with a focus on mouthings. A small subset of the Corpus NGT of over 94 minutes (40 sessions containing data from 40 signers) was fully annotated for the Mouth tiers at the time of writing. In the whole corpus, over 250 sessions contained some Mouth annotations, counting almost 12,000 tokens for a total of 70 different participants. These Mouth tier annotations were all classified according to type on the MouthType tier, and formed the basis of all our studies. Depending on the specific research goal, data from the whole corpus were used or from the smaller subset identified above.

In a first study (Bank et al., 2011), we investigated the variation in Dutch lexical items used as mouthings for twenty highly frequent signs. We used the MouthLemma tier to find all tokens of a certain type, and the MouthType classification to make a distinction between mouthings and mouth gestures. The main source of variation turned out to be between using a mouthing versus a mouth gesture, rather than between different spoken words occurring with the same manual sign. This dichotomy between mouthings and mouth gestures was established by using the MouthType tier.

We continued to investigate mouthings by looking at their spreading behaviour, encoding this information on the MouthSpr tiers (Bank et al., 2013). This allowed us to easily classify regressive and progressive spreading, as well as determining the scope of spreading by counting the number of angled brackets in an annotation. The finding here confirmed the findings of Crasborn et al. (2008) for the ECHO fable stories, namely that spreading is a frequent phenomenon: more than one in ten mouthings are spread out over two or more signs. The MouthSyll tiers could be used in future investigations on spreading that aim to analyse the phonological length of words, comparing those with the length of signs. Although we report some findings on this subject, we did not systematically annotate the number of syllables in each mouthing.

While in this study on spreading, no sociolinguistic differences were found based on distinctions in gender,

age, or region, we continued to look at sociolinguistic differences in the use of mouth actions more generally. In Bank et al. (submitted) we report the finding that while no group differences were found based the variables region, gender, or age, what does stand out is the high frequency of mouthings in comparison to the various types of mouth gestures. Depending on the signer, between 65 and 100% of all mouth actions are mouthings. We concluded that spoken language is an important element of deaf interaction in the Netherlands, even for native signers signing to other native signers whom they know well. Although the semi-spontaneous interaction was recorded in a lab setting, the further conclusion appears warranted that there simply is no 'pure' NGT in the sense of not being accompanied by elements of the spoken language, even though we consider NGT to be a language with its own lexicon and its own grammar.

In a final study, we are building on this conclusion by making use of the MouthAdd tiers (Bank et al., forthcoming). The MouthAdd tier is the only place where oral and manual information is combined, information that cannot otherwise be retrieved in an automated search in ELAN. In this study, we will analyse the structure of utterances where mouthings do more than contribute redundant information to manual signs or specify the semantics of manual signs.

The data for all of these studies will be published in the second release of the Corpus NGT annotations foreseen for the autumn of 2014.

5. Conclusion

We hope to have described an annotation scheme for mouth actions that could benefit a large number of sign language corpora. Many of the phenomena at its basis have been observed for many sign languages, albeit often on the basis of rather small data sets. We recommend the transcription of mouth actions on the Mouth tier as a basic element of corpus annotation for all sign languages, especially ones in which mouthings are not uncommon.

Admittedly, the validity of the distinctions that we propose to some extent remains to be confirmed by more research. As with other types of sign language corpus annotation, the annotation and analysis of many elements of signed interaction remains a constant process of improvement and revision based on new research methods and new insights into the functioning of sign languages and deaf interaction more generally. This should not withhold us from striving towards annotation standards (cf. Schembri & Crasborn, 2010).

Unlike the validity, the inter-annotator and intra-annotator reliability of the various elements of the annotation scheme is something that could be established relatively easily by dedicated studies. This is one of the steps we plan to take next.

References

Auer, E. T., Jr., & Bernstein, L. E. (2007). Enhanced Visual Speech Perception in Individuals With Early-Onset Hearing Impairment. *Journal of Speech, Language, and Hearing Research, 50*(5), 1157-1165.

Bank, Richard, Crasborn, Onno, & Hout, Roeland van. (2011). Variation in mouth actions with manual signs in Sign Language of the Netherlands (NGT). *Sign Language and Linguistics*, 14(2), 248–270.

Bank, Richard, Crasborn, Onno, & van Hout, Roeland. (2013). Alignment of two languages: The spreading of mouthings in Sign Language of the Netherlands. *International Journal of Bilingualism*. doi: 10.1177/1367006913484991

Bank, Richard, Crasborn, Onno, & van Hout, Roeland. (Submitted). The prominence of spoken language elements in a sign language.

Bank, Richard, Crasborn, Onno, & van Hout, Roeland. (forthcoming). Bimodal code-mixing: speech supported signing is the norm in NGT signers. Ms., Radboud University Nijmegen.

Boyes Braem, P., & Sutton-Spence, R. (Eds.). (2001). *The hands are the head of the mouth. The mouth as articulator in sign languages*. Hamburg: Signum Verlag.

Cappalletta, Luca, & Harte, Naomi (2012). *Phoneme-to-viseme mapping for visual speech recognition. Proceedings of the International Conference on Patter Recognition, Applications and Methods*, pp. 322-329.

Crasborn, Onno, Kooij, Els van der, Mesch, Johanna, Waters, Dafydd, & Woll , Bencie (2008). Frequency distribution and spreading behavior of different types of mouth actions in three sign languages. *Sign Language and Linguistics, 11*(1), 45-67.

Crasborn, Onno, & Zwitserlood, Inge. (2008). The Corpus NGT: an online corpus for professionals and laymen. In Onno Crasborn, Eleni Efthimiou, Thomas Hanke, Ernst Thoutenhoofd & Inge Zwitserlood (Eds.), *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages*. Marrakech, Morocco: ELRA, pp. 44-49.

Crasborn, Onno, Zwitserlood, Inge, & Ros, Johan. (2008). The Corpus NGT. An open access digital corpus of movies with annotations of Sign Language of the Netherlands (Video corpus). from Centre for Language Studies, Radboud University Nijmegen http://hdl.handle.net/hdl:1839/00-0000-0000-0004-DF8E-6

Ekman, Paul, & Friesen, Wallace V. (1978). *The facial action coding system. Investigator's guide.* Palo Alto, CA: Consulting Psychologists Press.

Emmorey, Karen, Borinstein, H.B., & Thompson, Robin. (2005). Bimodal bilingualism: Code-blending between spoken English and American Sign Language. In J. Cohen, K.T. McAlister, K. Rolstad & J. MacSwan (Eds.), *ISB4: Proceedings of the 4th International Symposium on Bilingualism*. Somerville, MA:

Johnston, Trevor. (2010). From archive to corpus: Transcription and annotation in the creation of signed

language corpora. *International Journal of Corpus Linguistics*, 15(1), 104-129.

Massaro, Dominic W. (1998). *Perceiving talking faces. From speech perception to a behavioral principle*. Cambridge, MA & London: The MIT Press.

Nonhebel, Annika, Crasborn, Onno, & Kooij, Els van der. (2004). Sign language transcription conventions for the ECHO Project: BSL and NGT mouth annotations. Ms, Radboud University Nijmegen.

Sandler, Wendy. (2006). From phonetics to discourse: the nondominant hand and the grammar of sign language. In Louis Goldstein, D.H. Whalen & Catherine Best (Eds.), *Laboratory Phonology 8* (pp. 185-212). Berlin: Mouton de Gruyter.

Sandler, Wendy. (2009). Symbiotic symbolization by hand and mouth in sign language. *Semiotica*, 174(1), 241-275.

Schembri, Adam, & Crasborn, Onno. (2010). *Issues in creating annotation standards for sign language description*. Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, Valletta, Malta, pp. 212-216.

Woll, Bencie. (2001). The sign that dares to speak its name: echo phonology in British Sign Language (BSL). In Penny Boyes Braem & Rachel Sutton-Spence (Eds.), *The hands are the head of the mouth* (pp. 87-98). Hamburg: Signum Verlag.

Woll, Bencie. (2012). Speechreading revisited. *Deafness & Education International*, 14(1), 16-21.