First Steps Towards a Procedure for Annotating Non-Manual Markers in Sign Languages

Oomen, M.; Esselink, L.; de Ronde, T.; Roelofsen, F.

Publication date
2023
Document Version
Author accepted manuscript
Published in
NELS 53: Proceedings of the Fifty-Third Annual Meeting of the North East Linguistic Society

Citation for published version (APA):
Oomen, M., Esselink, L., de Ronde, T., & Roelofsen, F. (2023). First Steps Towards a Procedure for Annotating Non-Manual Markers in Sign Languages. In S.-Y. Lam, & S. Ozaki (Eds.), *NELS 53: Proceedings of the Fifty-Third Annual Meeting of the North East Linguistic Society: January 12-14, 2023, University of Göttingen* (Vol. 2, pp. 257-266). GLSA.

First Steps Towards a Procedure for Annotating Non-Manual Markers in Sign Languages*

Marloes Oomen, Lyke Esselink, Tobias de Ronde & Floris Roelofsen

University of Amsterdam

1. Introduction

We report on the development, application, and evaluation of a procedure for annotating non-manual markers (NMM) in experimentally obtained sign language data. We also share resources to enable other researchers investigating NMM in sign languages or multimodal communication to utilize the annotation protocol we developed in their own research.

In Section 2, we outline our motivations for developing a new protocol for annotating NMM. Section 3 provides a brief description of the study for which the annotation protocol was specifically developed, which is about biased polar questions in Sign Language of the Netherlands (NGT). We then describe how the annotation guidelines were developed and used to annotate the study data in Section 4. In Section 5, we evaluate the reliability of the annotation procedure by assessing inter-annotator agreement. Section 6 concludes.

2. Motivation

In sign languages, facial expressions, body movements, and other NMM serve a wide range of linguistic functions, in addition to the gestural and affective functions they may fulfil more generally. There are plenty of examples in the literature tying particular NMM (or clusters of NMM) to particular grammatical functions (for a recent overview, see Wilbur 2021). For instance, Bahan (1996) has argued that eye gaze (or head tilt in the case of first person) can be used to mark verb agreement in American Sign Language (ASL); Göksel and Kelepir (2013) have claimed that (forward or backward) head tilt in Turkish Sign Language marks interrogative mood while specific combinations of head tilt and head movement distinguish polar (forward + head nod) and content (backward + headshake) questions; Wilbur and Patschke (1998) have proposed, again for ASL, that body leans are

*We thank Marc Schulder and James Trujillo for helpful discussion. This work is part of the project *Questions in Sign Language* (grant number VI.C.201.014, PI Roelofsen) financed by the Dutch Science Foundation (NWO).

used to convey contrast at the prosodic, lexical, semantic, and pragmatic level. Works such as these provide highly valuable descriptive, analytical and theoretical insights, but they tend to be based on relatively small sets of examples, for which it is often unclear exactly how they were obtained or analyzed. The analyses also generally do not involve detailed qualitative annotation of NMM, or the annotation procedure is not discussed.[1] Moreover, (individual) variation in NMM use is often not considered. This means that many claims about NMM and their properties and functions in sign languages still await robust empirical verification, which cannot be done without in-depth analysis of NMM patterns by means of careful linguistic annotation of linguistic data.

Facial expressions and other NMM also play an important role in multimodal communication, where they have been shown to be connected to a wide variety of semantic, pragmatic, and social functions (e.g., Bavelas and Chovil 2018, González-Fuente et al. 2015, Nota et al. 2021, Tomasello et al. 2019). Thus, research in this domain likewise requires (and sometimes already includes; e.g., González-Fuente et al. 2015, Nota et al. 2021) fine-grained annotation of facial expressions and other visual cues in video data.

Annotation of NMM is highly time-consuming and also poses challenges for data analysis, given the considerable number of possible NMM and the fact that temporal information is ideally also taken into account. Even so, as we have discussed, such work is vital both for empirical verification of theoretical claims as well as to gain more insight into the (extra-)linguistic factors that lead to variation in NMM use in sign language and multimodal communication. We conducted a study with both these goals in mind and found that there was no existing annotation protocol that fully satisfied our research needs. We therefore set out to develop one. The annotation protocol and associated resources can be freely used – and adapted – by other researchers. Indeed, we hope that further refinement of the protocol will become a community effort. Before we discuss the development and evaluation of the annotation procedure, we first provide some details about the aims and design of our specific case study.

3. The case study: Biased polar questions in NGT

We conducted a production experiment to elicit different polar question forms in Sign Language of the Netherlands (NGT). We expected NMM to play a crucial role in the marking of different question forms, making a high level of detail in NMM annotation a necessity.

Polar questions in sign languages have frequently been claimed to be marked by raised eyebrows, with other oft-cited NMM including wide-opened eyes, addressee eye contact, and a body and/or head forward position (Zeshan 2004). However, some studies have attested rather more variation in NMM use than is typically acknowledged (e.g., for NGT, Coerts 1992, de Vos et al. 2009). We hypothesize that at least part of this variation can be accounted for by taking *bias* into account. That is, a signer's *original speaker bias* (cf. Ladd 1981) about a topic before the current situational and conversational context, and the *contextual evidence* (cf. Büring and Gunlogson 2000) directly provided within that context,

[1]Notable exceptions are Pendzich (2020) on lexical NMM in German Sign Language and Lackner (2017) on the various functions of head and body movements in Austrian Sign Language.

may affect the way in which polar questions are asked. In spoken languages, speaker bias may be reflected in sentence structure, e.g., by adding specific words ("Does he *really* have a car?") or tags ("He has a car, doesn't he?"), or through high or low negation ("Does*n't* he have a car"; "Does he *not* have a car?"). Prosody also plays an important role (Reese and Asher 2009). For sign languages, we expect bias to be reflected, to a great extent, in NMM use, although we are also investigating (and annotating) sentence structure and manual bias marking strategies. In this paper, we discuss only the annotation procedure for NMM.

The study design was inspired by a sentence selection task designed by Domaneschi et al. (2017) for a study on biased polar questions in English and German, which in turn builds on Roelofsen et al. (2013). However, we had to substantially modify the task to turn it into a production experiment. In a role-play setting, participants were prompted to ask questions to two confederates, both native signers of NGT, whose responses introduced different original speaker expectations and contextual evidence (positive, negative, neutral). These exchanges triggered a target question – with potential bias – directed toward confederate B at the end of each role play.

Data were elicited from seven deaf NGT signers in a controlled production experiment. Data from one signer had to be excluded because the signer's productions differed from those of the other participants to such an extent that they could not be meaningfully analyzed along with the data from the other participants. From each participant, we elicited 35 target items across 7 conditions (different combinations of positive/neutral/negative speaker bias and contextual evidence), for a total of 205 items. Along with baseline declarative and interrogative sentences, the full data set includes 260 sentences. A full description of the experimental design and set-up, including experimental stimuli, can be found at `https://doi.org/10.21942/uva.21701954`. All elicited video data are available at `https://doi.org/10.21942/uva.21666203`.

All participant utterances were video-recorded and prepared for annotation using Adobe Premiere Pro. Video files were created such that each contained five target questions elicited from the same participant in the same experimental condition (e.g., PosNeut: positive original speaker bias and neutral contextual evidence). Video files were subsequently loaded into ELAN (2022), along with the tier template created for this study (see Section 4). A generalized version of this template, with corresponding Controlled Vocabulary, can be downloaded at `https://doi.org/10.21942/uva.22732616`.

We then prepared the files for further annotation. On the main tier ('Item_ID' in published template; 'Question' in study template), codes were added reflecting the participant number, experimental condition, and question number, e.g., 05-PosNeut-02. The scope of the annotations made on this tier corresponds to the scope of every target question. On the next two tiers, 'Glosses_EN' and 'Glosses_NL' (latter not included in published template), glosses corresponding to the ID glosses listed in Global Signbank for NGT (`https://signbank.cls.ru.nl/`) were added in English and in Dutch. Signs for which no ID gloss could be found in Signbank, such as fictional name signs or the sign VEGETARIAN, were prefixed with '%'.

After these steps, the transcription files were ready for NMM annotation.

4. Creating annotation guidelines for NMM

The annotation guidelines are based on guidelines initially developed by Gaasbeek (2022) for an exploratory study on polar questions in naturalistic data from the Corpus NGT (Crasborn et al. 2008, Crasborn and Zwitserlood 2008), in the context of the larger project that the present study is also a part of. Since the video data collected for our experimental study are of higher definition than the corpus data, they allow for much finer-grained annotation of non-manuals. We therefore added and changed various NMM features and values included in Gaasbeek's (2022) guidelines, and we also made an effort to align the annotation values with Action Units from the Facial Action Coding Scheme (FACS; Ekman et al. 2002[1978]) whenever possible; see Section 4.2. Below, we describe the steps we took in the initial development, subsequent refinement, and application of the annotation guidelines.

4.1 Development of initial annotation guidelines

We aimed to be as comprehensive as possible in the analysis of NMM in our data. We therefore aspired to enable annotation of all (potentially) relevant facial and body features, creating individual tiers in ELAN for each of them. Tiers were named according to the format 'NMM-[non-manual feature]', e.g., 'NMM.eye-shape'. Based on Gaasbeek's (2022) annotation scheme as well as careful initial observation of a subset of the video data, we created preliminary inventories of annotation values in the form of descriptive labels (e.g., 'raised'; 'tilt') for each tier. These values were listed in Controlled Vocabularies (an ELAN function). For all tiers, we also included the options 'neutral' and 'other'. In the latter case, annotators had to add an explanation on the comment tier. Our original coding scheme contained an inventory of twelve NMM tiers and three to nine possible values per tier.

4.2 Refinement of initial guidelines

In a second step, we evaluated how the drafted annotation guidelines worked in practice. Two coders, one L2 signer with ten years of signing experience in primarily professional contexts (first author) and one native signer with a background as an NGT teacher (third author) annotated a small portion of the data (18 items) in parallel. After every item, the coders compared their annotations and discussed cases of disagreement to evaluate whether certain parts of the annotation guidelines needed clarification, improvement, or revision. This led to several changes, the most notable one concerning head position. In the initial coding scheme, head position was annotated on a single tier, with possible values including 'tilt', 'chin-up', 'chin-down', 'head-forward', and 'head-backward'. This proved insufficient, as we frequently observed combinations of these values. The new annotation scheme therefore contains three separate tiers for head position ('NMM.head-x/y/z'). Another significant change was the removal of the tier 'NMM.chin' with possible values 'raised' and 'neutral', since chin raise is strongly conflated with mouth configuration (e.g., pressed lips and chin raise systematically co-occur) and therefore need not be annotated separately. The revised coding scheme contains twelve NMM tiers and three to nine values per tier.

These updated guidelines served as the basis for the publicly available annotation manual that can be downloaded here: https://doi.org/10.21942/uva.24080868. This document is similar to the annotation manual used in the described study, with two exceptions: (i) the published manual has been translated from Dutch into English; (ii) it includes images (screenshots from the study data) to illustrate NMM features.

The annotation manual lists all possible annotation values per tier, and includes descriptions and images for each NMM. It also shows, where applicable, how the annotation values in our coding scheme map onto Action Units from the Facial Action Coding Scheme, which has been extensively used in research on affective facial expressions, but also by Pendzich (2020) in her work on lexical NMM in German Sign Language. Motivated by the specific (linguistic) purposes of our study, our coding scheme does differ from FACS in several respects. For instance, annotation for repeated head movements, such as headshaking, is relevant for our study, but this is not part of FACS. We therefore introduced a separate tier 'NMM.head-move' to represent such dynamic features. For the tiers 'NMM-lip-corners' and 'NMM-lips', we added the option 'mouth-action' to be used in cases of mouthing or other linguistic mouth patterns. For 'NMM.eye-gaze', we did not use the FACS Action Units for right, left, upward, and downward directed eye gaze, but rather included the conversation-setting oriented labels addressee, researcher, and space. We also added tiers for body and shoulder position, which are not part of FACS. The annotation manual also includes a list of basic annotation rules.

4.3 Annotation of the data set

Next, 20% of the items in the data set were randomly selected for independent coding by the same two coders as in the previous iteration, using the revised guidelines resulting from that phase (see annotation manual). These two sets of annotations were used to assess the reliability of the annotation procedure (see Section 5). The remaining 80% of the data were then annotated by one master coder (the first author of this paper). For an example of a question fully annotated for NMM, see figure 1.

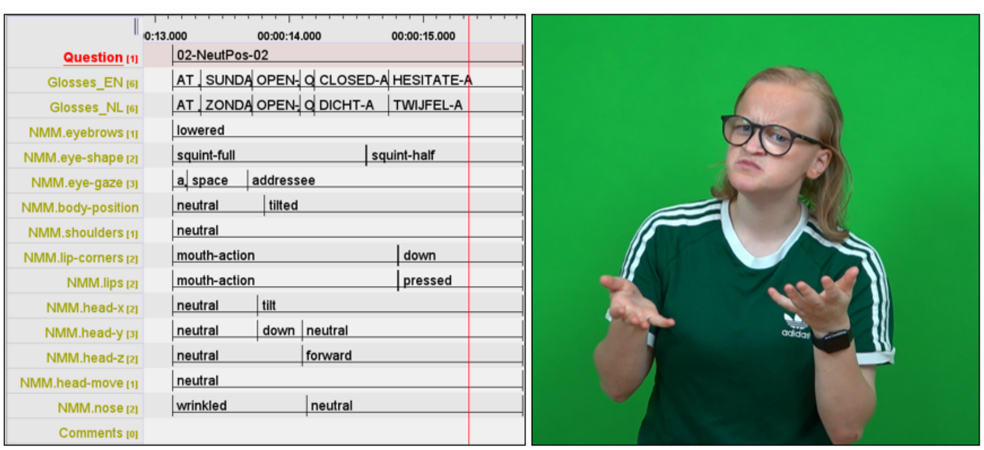

Figure 1: Screenshot of an annotated video in ELAN; video still aligns with red line.

The data annotation files (.eaf extension) are publicly available at https://doi.org/10.21942/uva.22737074. Note that the annotations include a number of additional tiers not included in the description above. The tiers NMM.mouthing, MM.PU, and MM.sign target potential manual markers of question sentences, as well as possibly relevant mouthings of Dutch spoken words accompanying manual material. The tiers 'Structure', 'Structure-components', and 'Structure-elements' represent information about sentence structure.

5. Assessing the reliability of the annotation procedure

Sign language studies are often based on annotated video data. However, the reliability of the annotation procedure is rarely, if ever, systematically assessed. For instance, six out of the twelve research articles published in *Sign Language & Linguistics* in 2021-2022 investigate linguistic properties of sign languages based on annotated video data, but none of them report inter-annotator agreement indices. It is important to carry out such an assessment, for at least the following two reasons. First, only when we know how reliable the annotation procedure was, can we weigh how solid the conclusions are that are drawn from the data. And second, when we do not only assess the *overall* reliability of the annotation procedure but also the extent to which multiple annotators agreed on the *different labels* that they applied in the annotation process, we learn which labels were easier to agree on and which ones were harder, making it possible to improve the annotation guidelines in future iterations. This is particularly relevant for our current purpose, namely, to develop a reliable annotation procedure for NMM.

5.1 Two approaches to assess inter-annotator agreement for timed-event sequential data

The type of data that sign language linguists work with is a particular kind of so-called *timed-event sequential data* (Bakeman et al. 2009, Bakeman and Quera 2011). In general, such data involve recordings of sequences of events, each with a particular time duration. Besides sign linguists, researchers in other domains (e.g., speech, multimodal communication, or animal behavior) also work with this kind of data, make similar use of annotations, and have devised several methods to assess inter-annotator agreement for this type of data. Broadly, two approaches can be distinguished: *frame*-based approaches and *event*-based approaches (Bakeman et al. 2009).[2] For reasons of space, we only present a frame-based approach here, and restrict the scope of the discussion to the *eyebrow* tier. We intend to discuss the event-based approach and comprehensively present the results for all tiers in a future extension of the present paper. Scripts for determining inter-annotator agreement under both approaches are available at https://doi.org/10.21942/uva.24080724.

5.2 The frame-based approach

On the frame-based approach, we simply consider each individual frame in all videos that were annotated by both coders, C1 and C2, and compare the labels applied to that frame.

[2] Frame-based approaches are also referred to as *time*-based approaches.

We do this separately for each tier. For the *eyebrows* tier, the findings are summarized in the so-called *confusion matrix* in the table in (1). We see, for instance, that C1 applied the label 'raised' to 1570 frames in total. Of these, 1111 were also labeled as 'raised' by C2, but 7 were labeled as 'inner-raise', 7 as 'lowered', 90 as 'neutral', and 355 as 'raised-low'. We can also present the confusion matrix in terms of percentages. The table on the left in (2) does this from the perspective of C1 (all rows add up to 100%), while the table on the right does this from the perspective of C2 (all columns add up to 100%).

(1) *Confusion matrix for the eyebrow tier.*

C1/C2	inner-raise	lowered	neutral	raised	raised-low	Total
inner-raise	416	69	35	179	415	1114
lowered	411	2068	74	5	659	3217
neutral	34	0	817	19	133	1003
raised	7	7	90	1111	355	1570
raised-low	14	262	233	4	478	991
Total	882	2406	1249	1318	2040	7895

(2) *Percentage-wise confusion matrices from the perspective of (a) C1 and (b) C2.*

(a) From the perspective of C1:

C1/C2	in-ra	lo	neu	ra	ra-lo	Total
in-ra	**38**	6	3	16	37	100
lo	13	**64**	2	0	21	100
neu	3	0	**82**	2	13	100
ra	0	0	6	**71**	23	100
ra-lo	1	27	24	0	**48**	100

(b) From the perspective of C2:

C1/C2	in-ra	lo	neu	ra	ra-lo
in-ra	**47**	3	3	14	20
lo	47	**86**	6	0	32
neu	4	0	**65**	2	7
ra	1	0	7	**84**	17
ra-lo	1	11	19	0	**24**
Total	100	100	100	100	100

These tables lend insight into which labels were particularly difficult to agree on, namely, 'inner-raise' and 'raised-low'. For instance, when C2 applied the label 'raised-low', C1 agreed only 24% of the time, and instead applied 'inner-raise' 20% of the time, 'lowered' 32% of the time, 'neutral' 7% of the time, and 'raised' 17% of the time. This indicates that, when it comes to describing when the label 'raised-low' should be applied, the current annotation guidelines need considerable further improvement.

On a more positive note, C1 and C2 agreed relatively often on the labels 'raised' and 'lowered'. Moreover, whenever one of the coders applied one of these labels and the other coder did not agree, the label that they applied instead was nearly always one of the ambiguous labels 'inner-raise' and 'raised-low'. There were no frames which one coder labeled as 'raised' and the other as 'lowered'.

Besides confusion matrices, another way to quantify inter-annotator agreement is to compute *agreement indices* for each label. Here it is important to note that so-called *raw* agreement indices are insufficient. To illustrate this, suppose that two annotators x and y

label 100 items. To 50 items they both apply label A, to 20 items only x applies label A, to 20 items only y applies label A, and to the final 10 items they both apply another label. Then, x and y agree in 50 + 10 = 60 of the cases as to whether label A applies or not. The raw agreement index for label A, then, is 0.6. However, this does not take into account the possibility that, at least in some cases, x and y may have agreed on the application of label A *by mere chance*. Both x and y applied label A to 70% of the items, and other labels to 30% of the items. If they would randomly assign label A to 70% and other labels to 30% of the items, they would agree 58% of the time as to whether A applies or not (because $(0.7*0.7)+(0.3*0.3) = 0.58$). So the raw agreement index, $i_{raw} = 0.6$, is just slightly higher in this case then the chance agreement index, $i_{chance} = 0.58$.

Chance-corrected agreement indices take this factor into account. One widely used chance-corrected index is Cohen's kappa index (Cohen 1960). It is computed by dividing the difference between i_{raw} and i_{chance} by the difference between i_{chance} and the index for perfect agreement, which is 1.

$$\kappa := \frac{i_{raw} - i_{chance}}{1 - i_{chance}}$$

In the example above, κ would amount to $0.02/0.42 = 0.05$. To give some other examples, if $i_{raw} = 0.7$ and $i_{chance} = 0.5$ then $\kappa = 0.4$, and if $i_{raw} = 0.9$ and $i_{chance} = 0.6$ then $\kappa = 0.75$.

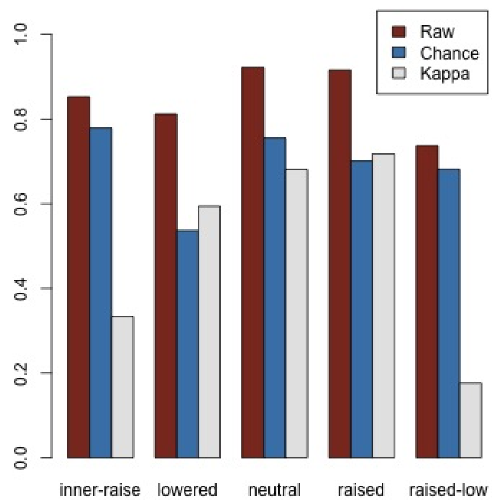

Figure 2: Agreement indices.

Now let us return to our own annotation procedure. Figure 2 displays the raw, chance, and κ agreement indices for all possible labels on the *eyebrows* tier. We see that the κ indices for 'raised' (0.72) and 'lowered' (0.59) are reasonably high, but those for 'inner-raise' (0.33) and 'raised-low' (0.18) are much lower. Again, these are clear indications of where the annotation guidelines need further improvement. Moreover, if, in the future, we would assess the reliability of a new version of the guidelines computing the same κ indices, we would obtain a concrete measure of the extent to which those new guidelines improve on the current guidelines.

6. Upshot and outlook

In this paper, we have presented initial work towards establishing an annotation procedure primarily intended for the investigation of NMM in sign languages, but potentially also useful for the investigation of facial expressions and body poses in multimodal communication. We have also presented methods to assess the reliability of annotation procedures, something that is not yet commonly done in sign language research. Applying these methods to our own annotation procedure for NMM revealed that several aspects of it need to be further improved. We intend to undertake this in future work.

Resources

1. Annotation template: https://doi.org/10.21942/uva.22732616
2. Annotation manual: https://doi.org/10.21942/uva.24080868
3. Inter-annotator agreement scripts: https://doi.org/10.21942/uva.24080724
4. Details experimental design case study: https://doi.org/10.21942/uva.21701954
5. Video data from the case study: https://doi.org/10.21942/uva.21666203
6. Annotation files for the case study: https://doi.org/10.21942/uva.22737074

References

Bahan, Benjamin. 1996. Non-manual realization of agreement in American Sign Language. Doctoral dissertation, Boston University.

Bakeman, Roger, and Vicenç Quera. 2011. *Sequential analysis and observational methods for the behavioral sciences*. Cambridge University Press.

Bakeman, Roger, Vicenç Quera, and Augusto Gnisci. 2009. Observer agreement for timed-event sequential data: A comparison of time-based and event-based algorithms. *Behavior Research Methods* 41:137–147. https://doi.org/10.3758/BRM.41.1.137.

Bavelas, Janet, and Nicole Chovil. 2018. Some pragmatic functions of conversational facial gestures. *Gesture* 17:98–127. https://doi.org/10.1075/gest.00012.bav.

Büring, Daniel, and Christine Gunlogson. 2000. Aren't positive and negative polar questions the same? Ms., UCSC/UCLA. Available at https://semanticsarchive.net/Archive/mYwOGNhO/polar_questions.pdf.

Coerts, Jane. 1992. Nonmanual grammatical markers. An analysis of interrogatives, negations and topicalisations in Sign Language of the Netherlands. Doctoral dissertation, University of Amsterdam.

Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20:37–46.

Crasborn, Onno, and Inge Zwitserlood. 2008. The Corpus NGT: an online corpus for professionals and laymen. In *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages*, ed. by Onno Crasborn, Thomas Hanke, Eleni Efthimiou, Inge Zwitserlood, and Ernst Thoutenhoofd, 44–49. Paris: ELDA.

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

Domaneschi, Filippo, Maribel Romero, and Bettina Braun. 2017. Bias in polar questions: Evidence from English and German production experiments. *Glossa: a journal of general linguistics* 2:26. 1–28. https://doi.org/10.5334/gjgl.27.

Ekman, Paul, Wallace V. Friesen, and Joseph C. Hager. 2002[1978]. *Facial Action Coding System: A technique for the measurement of facial movement*. San Francisco, CA: Consulting Psychologists Press.

ELAN. 2022. Nijmegen, Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan.

Gaasbeek, Marleen. 2022. (Non)manual marker annotation guideline. Ms., University of Amsterdam. Unpublished.

Göksel, Aslı, and Meltem Kelepir. 2013. The phonological and semantic bifurcation of the functions of an articulator: HEAD in questions in Turkish Sign Language. *Sign Language & Linguistics* 16:1–30.

González-Fuente, Santiago, Victoria Escandell-Vidal, and Pilar Prieto. 2015. Gestural codas pave the way to the understanding of verbal irony. *Journal of Pragmatics* 90:26–47. https://doi.org/10.1016/j.pragma.2015.10.002.

Lackner, Andrea. 2017. *Functions of head and body movements in Austrian Sign Language*. Berlin: De Gruyter Mouton.

Ladd, D. Robert. 1981. A first look at the semantics and pragmatics of negative questions and tag questions. In *Papers from the Seventeenth Regional Meeting of the Chicago Linguistic Society*, 164–171. Chicago Linguistic Society.

Nota, Naomi, James P. Trujillo, and Judith Holler. 2021. Facial signals and social actions in multimodal face-to-face interaction. *Brain Sciences* 11:1017. https://doi.org/10.3390/brainsci11081017.

Pendzich, Nina-Kristin. 2020. *Lexical nonmanuals in German Sign Language. Empirical studies and theoretical implications*. De Gruyter Mouton.

Reese, Brian, and Nicholas Asher. 2009. Biased questions, intonation, and discourse. In *Information Structure: Theoretical, Typological, and Experimental Perspectives*, ed. by Malte Zimmermann and Caroline Féry, 139–174. https://doi.org/10.1093/acprof:oso/9780199570959.003.0007.

Roelofsen, Floris, Noortje Venhuizen, and Galit Weidmann Sassoon. 2013. Positive and negative polar questions in discourse. In *Proceedings of Sinn und Bedeutung 17*, ed. by Emmanuel Chemla, Vincent Homer, and Gregoire Winterstein, 455–472.

Tomasello, Rosario, Cora Kim, Felix R. Dreyer, Luigi Grisoni, and Friedemann Pulvermüller. 2019. Neurophysiological evidence for rapid processing of verbal and gestural information in understanding communicative actions. *Scientific Reports* 9:16285. https://doi.org/10.1038/s41598-019-52158-w.

de Vos, Connie, Els van der Kooij, and Onno Crasborn. 2009. Mixed signals: Combining linguistic and affective functions of eyebrows in questions in Sign Language of the Netherlands. *Language and Speech* 52:315–339. https://doi.org/10.1177/0023830909103177.

Wilbur, Ronnie B. 2021. Non-manual markers – theoretical and experimental perspectives. In *The Routledge Handbook of Theoretical and Experimental Sign Language Research*, ed. by Josep Quer, Roland Pfau, and Annika Herrmann, 530–565.

Wilbur, Ronnie B., and Cynthia G. Patschke. 1998. Body leans and the marking of contrast in ASL. *Journal of Pragmatics* 30:275–303.

Zeshan, Ulrike. 2004. Interrogative constructions in signed languages: Crosslinguistic perspectives. *Language* 80:7–39.

Oomen, Esselink, De Ronde & Roelofsen
m.oomen2@uva.nl