Using Computer Vision to Analyze Non-manual Marking of Questions in KRSL

Anna Kuznetsova   kuzannagood@gmail.com
School of Linguistics, NRU HSE, Moscow, Russia

Alfarabi Imashev   alfarabi.imashev@nu.edu.kz
Department of Robotics and Mechatronics, School of Engineering and Digital Sciences,
Nazarbayev University, Nur-Sultan, Kazakhstan

Medet Mukushev   mmukushev@nu.edu.kz
Department of Robotics and Mechatronics, School of Engineering and Digital Sciences,
Nazarbayev University, Nur-Sultan, Kazakhstan

Anara Sandygulova   anara.sandygulova@nu.edu.kz
Department of Robotics and Mechatronics, School of Engineering and Digital Sciences,
Nazarbayev University, Nur-Sultan, Kazakhstan

Vadim Kimmelman   vadim.kimmelman@uib.no
Department of Linguistic, Literary and Aesthetic Studies, University of Bergen

Abstract

This paper presents a study that compares non-manual markers of polar and wh-questions to statements in Kazakh-Russian Sign Language (KRSL) in a dataset collected for NLP tasks. The primary focus of the study is to demonstrate the utility of computer vision solutions for the linguistic analysis of non-manuals in sign languages, although additional corrections are required to account for biases in the output. To this end, we analyzed recordings of 10 triplets of sentences produced by 9 native signers using both manual annotation and computer vision solutions (such as OpenFace). We utilize and improve the computer vision solution, and briefly describe the results of the linguistic analysis.

1   Introduction

Non-manual marking, that is, linguistically significant use of the body, head, facial features, and eye gaze, is a prominent feature of sign languages (Pfau and Quer, 2010). For instance, in most sign languages, polar questions are accompanied with raised eyebrows, and some type of head movement (Cecchetto, 2012). While non-manual markers in many sign languages have been previously studied, many other sign languages have not been analyzed before, including Kazakh-Russian Sign Language (KRSL), which we discuss in this study.

Current developments in computer vision provide an opportunity for a large-scale quantitative research on non-manual markers. In this study, we evaluate whether computer vision solutions can be utilized for the analysis of non-manual marking present in sign language video recordings. The objective of this work is to compare non-manual markers in statements and questions in KRSL, and to test a computer vision solution against manual annotations.

2 Background

2.1 Non-manual markers in sign languages

Sign languages employ not only hands, but also the body, the head, and the face in order to express linguistic information. Non-manual markers have been analysed for many sign languages (see Pfau and Quer (2010) for overviews). These markers function on different linguistic levels: phonological, morphological, syntactic, and prosodic.

Question marking in particular has been studied for many sign languages (Cecchetto, 2012). Polar (yes/no) questions seem to be almost universally marked by eyebrow raise on the whole sentence, while content questions (wh-questions) are more varied: some sign languages use eyebrow raise, some use eyebrow lowering, and some a combination of both (Zeshan, 2004). In addition, some type of head movement is reported as a marker for many languages, including backward head tilt (chin moving upwards) and forward head tilt (chin moving downward/forward).[1] The non-manual markers vary in scope: they can align with different constituents in the sentence. Furthermore, recent corpus-based research shows a high degree of variability of non-manual marking of questions, contrary to previous claims of the obligatory nature of such markers (Hodge et al., 2019).

Because of both typological and language-internal variation in non-manual marking of questions in sign languages, it is clearly necessary to conduct more empirical studies of such marking in languages that have not been described yet. In addition, applying novel computer vision techniques can facilitate reliable quantitative analysis and enable quantitative cross-linguistic comparison in future.

2.2 Quantitative approaches to non-manual marking

While there exist quantitative studies of non-manuals in various sign languages, some also based on naturalistic corpus data (Coerts, 1992; Puupponen et al., 2015; Hodge et al., 2019), until very recently quantitative approaches were limited by the data sets and the available techniques of analysing non-manuals. Many projects in the past employed manual annotation of non-manual markers, which is both extremely time-consuming and potentially unreliable (Puupponen et al., 2015). Moreover, manual annotation rarely provides the amplitude of non-manuals – in most cases annotation only states the existence of the marker. A more reliable alternative has been to use motion tracking to record precise quantitative data (Puupponen et al., 2015) e.g. on head movement. However, using a motion tracking set up is costly, and the signers have to wear trackers, which make the data recorded this way very far from naturalistic.

Currently two developments have made a large-scale quantitative studies of non-manual marking possible. First, large naturalistic corpora have been created for several sign languages (Crasborn et al., 2008; Konrad et al., 2020). Second, computer vision techniques now allow tracking of the body and facial features in video recordings without any trackers (see references in Section 4). The field of sign language translation has already benefited from these and other factors (see (Camgoz et al., 2018) for a brief overview), however most models only consider hand signs (Zimmermann and Brox, 2017) and other models do not interpret the video features at all (Li et al., 2020). Therefore we are just starting a discussion on the applicability of machine learning methods to non-manual feature extraction in sign languages. In this study, we test the applicability of computer vision to studying question marking using a controlled data set of statements and questions in Kazakh-Russian Sign Language.

---

[1]Other markers relevant for question marking include eye aperture, eye gaze direction, and body movements, but we do not consider them in this study (Cecchetto, 2012).

Proceedings of the 18th Biennial Machine Translation Summit, Virtual USA, August 16 - 20, 2021
1st International Workshop on Automatic Translation for Signed and Spoken Languages

Page 50

2.3 Kazakh-Russian Sign Language (KRSL)

KRSL is the language used primarily by the deaf and hard-of-hearing people in Kazakhstan. We use the term KRSL to acknowledge the fact that this language is closely related to Russian Sign Language due to the common history in the times of the Soviet Union. While no formal comparison between the two languages has been conducted, anecdotally the two languages are fully mutually intelligible, and not considered as separate languages by the deaf signers. Note, however, that KRSL signs can be accompanied by Kazakh mouthings.

Question marking in KRSL has not been described before. Based on the typological research cited above, we had a strong expectation that eyebrow position and head movement would be used to mark questions in KRSL, and that polar questions would be marked with raised eyebrows; we did not have a clear expectation about the eyebrow position in wh-questions.

3 Methodology

3.1 Participants and data collection

We collected video recordings from 9 native signers of KRSL: five deaf signers, and four interpreters, who are hearing children of deaf adults (CODAs). The data set analysed here is a part of a larger data set collected for a different project on automatic sign language recognition.[2]

We created a list of 10 simple sentences consisting of a subject and an intransitive verb. Each of the sentences was collected in three forms: statement (1a), polar question (1b), and wh-question (1c). The former two types of sentences thus contained two signs, and the latter type contained three signs due to the presence of a wh-sign.

1. (a) GIRL FALL   (b) GIRL FALL?   (c) WHERE GIRL FALL?
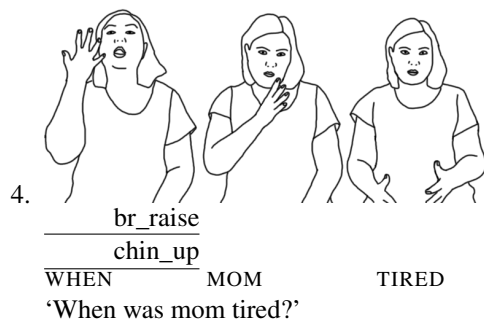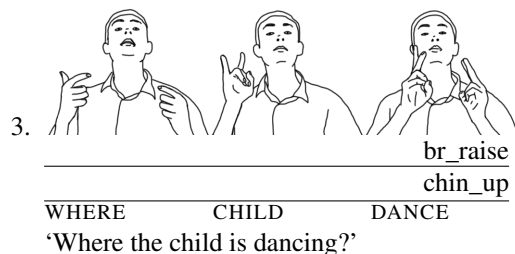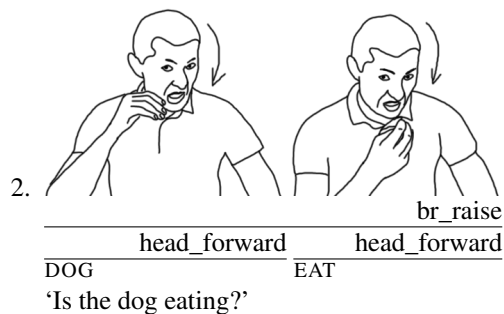   'A girl fell.'   'Did the girl fall?'   'Where did the girl fall?'

The stimuli were presented in written Russian to the hearing signers, and as video recordings in KRSL to the deaf signers. We did not use filler sentences, nor concealed from the signers that we were interested in question marking in KRSL. Our aim was thus not to create a maximally naturalistic data set, but a controlled data set with uniform structures produced by several signers for experimenting with computational approaches. No explicit instructions were given about the non-manuals, so we expected the signers to produce the markers natural to them.

Given that the data set was created for the purposes of automatic sign language recognition, and not for research on the grammar of KRSL, and due to the relatively small pool of signers which also includes hearing CODAs, the current study can only be considered describing non-manual marking of questions in this specific data set, and not in KRSL in natural settings. However, we believe that this is a first step towards a more broader research. Furthermore, the size of the data set is quite small, so further testing of the approach with larger data sets will be required in future.

3.2 Manual annotation

As the first step in research, we watched all the videos in order to get a qualitative picture of the non-manual patterns. For polar questions, it turned out that the main non-manual markers were eyebrow raise on the whole sentence and two consecutive forward head tilts on the subject and verb (2). For wh-questions, it was eyebrow raise on the whole sentence or only on the wh-sign and a backward head tilt on the whole sentence or the wh-sign (3-4). It was also noticeable that, in wh-questions, the signers had less consistency in their marking.

---

[2]All the data used in this study as described below is available at https://github.com/kuzanna2016/non-manuals-2020.

Proceedings of the 18th Biennial Machine Translation Summit, Virtual USA, August 16 - 20, 2021
1st International Workshop on Automatic Translation for Signed and Spoken Languages

Page 51

2.

```
                                    br_raise
            head_forward      head_forward
DOG                           EAT
```

'Is the dog eating?'

3.

```
                                  br_raise
                                  chin_up
WHERE           CHILD         DANCE
```

'Where the child is dancing?'

4.

```
        br_raise
        chin_up
WHEN          MOM          TIRED
```

'When was mom tired?'

With these observations, it was decided that we need to manually annotate eyebrow movement and head tilts. Besides that, the manual signs were annotated to determine the boundaries of the constituents, and the syntactic roles of the constituents (subject, verb, wh-word) were annotated. The annotations were made by the first author, according to the Corpus NGT Annotation Conventions (Crasborn et al., 2015) using the ELAN software (ELAN, 2020).

In order to explore reliability of manual annotation, 20% of randomly selected videos (54 videos) were independently annotated by the last author specifically for eyebrow movement (as later in the paper we assess computer vision measurements of this specific non-manual against manual annotations). Inter-rater agreement was calculated in two different ways: using agreement in category assignment and using the percentage of overlap between the annotations to take duration into account. We found moderate raw agreement for eyebrow movement detection (67%), and even lower agreement in overlap between annotations (57%).

This testing of the reliability of manual annotations is a showcase of the difficulty and subjectivity of this procedure. It is clear that automation of annotation is a necessity. At the same time, the computer vision tools discussed below make it potentially possible to study these subtle phonetic properties of non-manuals.

In some of the videos, the signers produced signs other than the subject, object, and the verb (such as a past tense marker), and in a few videos the subject sign was missing. We removed such videos from further quantitative analysis. Having done that, we had 259 videos in total (88 statements, 82 polar questions, 89 wh-questions).
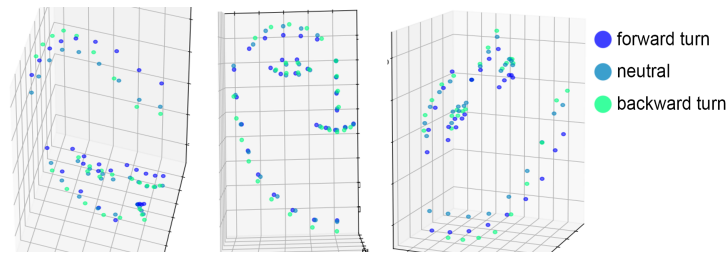
Proceedings of the 18th Biennial Machine Translation Summit, Virtual USA, August 16 - 20, 2021
1st International Workshop on Automatic Translation for Signed and Spoken Languages

Page 52

Figure 1: The behaviour of keypoints with different head turns on the test video.

4 Applying computer vision

The field of computer vision is actively developing thanks to the advancements in deep learning. Achievements in this area allow computers to process a large amount of visual information, such as pictures, photos or videos. One of the tasks of computer vision is landmark detection. It can be described as an object recognition task with localization: the algorithm needs not only to detect an object on the image, but also to estimate the position of this object. We planned to use face and head landmarks recognition tools to estimate the eyebrow and head movement.

OpenFace is a toolkit for facial landmark detection, head pose estimation, facial action unit recognition and other facial behaviour analysis (Baltrušaitis et al., 2018, 2013; Zadeh et al., 2017). OpenFace is able to estimate 3D landmarks position from 2D points using Point Distribution Model, which parametrizes the shape of a face using a limited set of parameters, such as scaling, rotation, translation and individual deformations of the face. We used OpenFace 2.2.0 FeatureExtraction model for getting face landmarks position in 3D and head pose estimation for every frame. Face landmarks and head coordinates were in x,y,z-coordinates in mm, and the head rotation angle was in radians with the camera being the origin. The model also provided us with a confidence score for each frame. By outputting head rotation, OpenFace directly provides a measure of forward/backward head tilt that we are interested in for the analysis. However, the estimation of eyebrow position from OpenFace output is less straightforward.

When we visually investigated the OpenFace output, we noticed that the results conflicted with our initial observations. We expected polar questions to have the largest eyebrow raise but OpenFace output was even smaller than in statements. Furthermore, we saw correlation between rotation angle and eyebrow distance, meaning that, probably, OpenFace predictions are prone to bias by the head rotation.

To demonstrate this bias, we plotted 3 frames from our test video of the forward to backward head tilt without eyebrow movement. We rotated the face keypoints in the reverse angle of the computed head rotation and centred them on the bridge of the nose keypoint number 27. From Figure 1, we can see that the face points, especially eyebrow points, are bending. Hence, the 3D model that OpenFace deduces is likely to be distorted in the presence of head movement. We therefore had to attempt to eliminate such distortions.

General 3D reconstruction from a single camera is a really challenging problem and it is outside of our area of competence, therefore we were not able to modify the OpenFace model itself. Moreover, we do not have the specifically annotated data with facial landmarks to retrain the model. That is why we will deal with the computed output instead of the model.

To deal with the bias we created a machine learning model that would predict the eyebrows distance depending on the head rotation when the eyebrows were not raised or lowered, i.e. the default eyebrow distance with the influence of tilting. We trained linear regression with L2 regularization and alpha 0.001 from *sklearn* library for Python (Pedregosa et al., 2018) on statements that contained no eyebrow raises: 63 sentences (4414 frames). The target of the model was the distance from the eyebrow points (18, 23 - inner, 20, 25 - outer) to the eye line -

Proceedings of the 18th Biennial Machine Translation Summit, Virtual USA, August 16 - 20, 2021
1st International Workshop on Automatic Translation for Signed and Spoken Languages

Page 53

a line between points 36, 45. The mean inner eyebrow distance in the subset was 29.7 and the outer was 27.1. Our choice of the model was based on the observation that the distortion seems to be linear and consistent across signers – Pearson correlation coefficient between vertical head angle and the eyebrow distance to the eye line in sentences with no eyebrow raise is -0.42 for the inner distance and -0.48 for the outer distance. Moreover, the training data set is small and thus a simple model should be sufficient.

Model №	Features	inner MSE	outer MSE
1	cos, sin, tan of vertical and horizontal head angles; nose length (distance between points 27 and 30)	4	
2	+ signer IDs and sentence IDs	1.61	1.54
3	+ the vertical eye distances	1.45	1.36

Table 1: Models' features description and the results

The features of the models that we used are shown in Table 1. We did not use the angle of rotation in the z-axis (from the ears to the shoulders) because it worsened the result for such turns. We did not use pure radians for rotation, because they shifted the model to some incomprehensible extremes.[3] In all our experiments we used 5-split random permutation cross-validation with test proportion of 25% (731 frames) to make sure the model does not overfit.

The first model performed with an MSE of about 4 for each eyebrow distance. However we wanted a better result. We added meta information as one-hot encoded vectors and it significantly increased the quality on the cross-validation. We believe that this allowed the model to learn the individual mean eyebrow height for each signer. We also noticed that blinking affected the distance estimation, so we added eye aperture features, which slightly reduced the error.[4]

After training the final model on statements, the eyebrow distance was predicted for all sentences, and then subtracted from the distance computed on OpenFace output directly. Thus, we subtracted the changes in distance caused by the tilts from the distance based on the output to get the unbiased distance measurement.

To check that this approach produced reasonable results, we looked manually at 35 sentences and compared the predictions with the annotations. In general, the output of the model agreed with our annotations. Furthermore, the correlation between the vertical head tilt and the new eyebrow distance is lower than on the original distance (-0.27 for inner, -0.25 for outer). Having demonstrated that this method of adjusting the eyebrow distance for head movement works well, we conducted the subsequent analysis using it.

5   Statistical analysis

One of the main advantages of applying computer vision to analysis of non-manuals in sign languages is that it opens the possibility of consequent advanced statistical analysis, instead of relying on qualitative observations. Thus, we use the current data set to showcase a possible statistical analysis of the output of the proposed computer vision approach.

We analyzed the data in R (version 3.6.3) using R Studio (version 1.0.143) (R Core Team, 2020; RStudio Team, 2020). For this study, we averaged the eyebrows distance in each video in the parts of sentence areas for inner eyebrows points (20,23) and for outer eyebrows points

---

[3]We also deleted the frames that had low OpenFace confidence (<0.8) - 103 frames from 12 videos in total.

[4]It was our decision to select specific data set features to increase the accuracy of the model. Thus the current model would not generalize to other data sets and we encourage other researchers to retrain it on their data or use the first model without the meta features.

Proceedings of the 18th Biennial Machine Translation Summit, Virtual USA, August 16 - 20, 2021
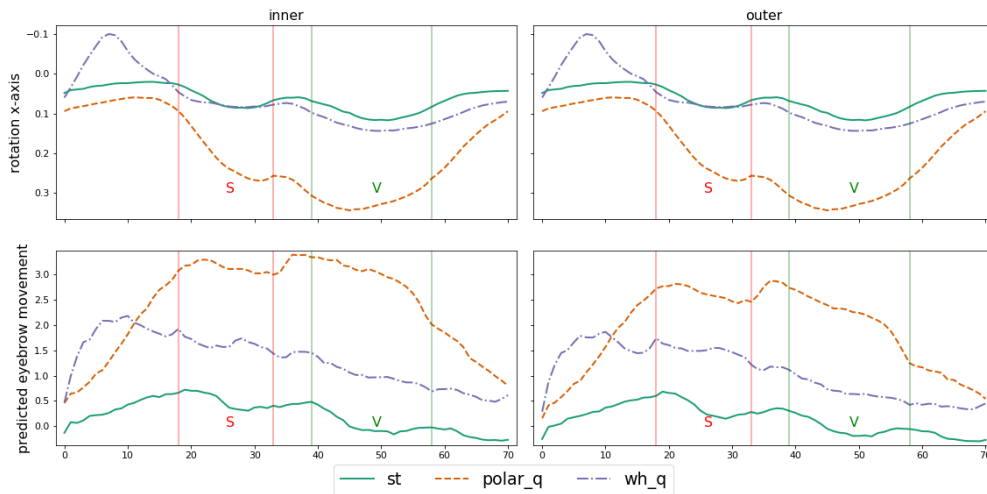*1st International Workshop on Automatic Translation for Signed and Spoken Languages*

Page 54

Figure 2: Mean head rotation angle and predicted difference in eyebrow movement for three sentence types. Vertical lines define the beginning and ending of the part of sentence tag (S - subject, V - verb). Rotation graphs are reversed in y-axis for interpretation (lower angle - backward head tilt, higher angle - forward head tilt).

(18,25) (averaging the left and right eyebrow distance). Thus, we had four eyebrow measurements for all the sentences (inner and outer mean distance for the subject and inner and outer mean distance for the verb) and separately we had six eyebrow measurements for the wh-questions (inner and outer mean distance for the wh-sign, subject, and verb). Besides, we calculated the mean head rotation angle on the x-axis (vertical tilts) again for two signs for all the sentences and for three signs for the wh-questions.

A mixed-effects multivariate linear regression model was picked for the analysis (Baayen et al., 2008; Bates et al., 2014). We made 9 models with the outcome variables of internal and external eyebrow distances and head tilt angle on the subjects, verbs, and wh-questions.

The fixed predictor variables for the first 6 models were sentence type (categorical, three levels: statement, polar question, wh-question), group (categorical, deaf vs. hearing), and all the interactions between the two predictors. For the last three models, the fixed predictor variables were part of sentence (categorical, three levels: wh-word, subject, verb), group (categorical, deaf vs. hearing), and all the interactions between the two predictors. We used orthogonal coding of contrasts for the predictors with three levels. Finally, for all models, the random variables were participant (with a random slope for sentence type or part of sentence), and sentence (with a random slope for group).

For the models, we used the *lme4* package (Chung et al., 2015) with the help of the *blme* package to achieve convergence with a small number of levels for the random effects (Chung et al., 2013). The significance of the contribution of the factors was computed with the *ANOVA* function from the *car* package (Fox and Weisberg, 2019).

6 Results

Firstly, we examined the results visually. From Figure 2, we can see that: polar questions have tilting forward on the subject and verb and eyebrows raise on the whole question, while wh-questions have tilting backwards on the wh-word and eyebrows raise in the beginning on the wh-word, which is slowly declining to the end of the sentence. In all the statistical analyses below, the effect of group and interactions between the group and the other effects were never

Proceedings of the 18th Biennial Machine Translation Summit, Virtual USA, August 16 - 20, 2021
1st International Workshop on Automatic Translation for Signed and Spoken Languages

Page 55

significant, so we do not discuss them further.

6.1 Eyebrow movement

We find that sentence type significantly influences both internal and external eyebrows distance on the subject and the verb (ANOVA χ^2 show significance at p<0.001 for all the comparisons).

For internal eyebrow points, the eyebrows distance is bigger for wh-questions than statements on the subject by estimated 1 mm (*se* = 0.64, *t* = 1.57), and on the verb by 0.87 mm (*se* = 0.5, *t* = 1.76). The average between statement and wh-question is lower than polar questions by 2.05 mm (*se* = 0.46, *t* = 4.48) on the subject and 2.34 mm (*se* = 0.55, *t* = 4.23) on the verb.

For external eyebrow points, the eyebrows distance is bigger for wh-questions than statements on the subject by estimated 1 mm (*se* = 0.65, *t* = 1.5), and on the verb by estimated 0.63 mm (*se* = 0.5, *t* = 1.27). The average between statement and wh-question is less than polar questions by estimated 1.66 mm (*se* = 0.43, *t* = 3.89) on the subject and 1.77 mm (s*se* = 0.6, *t* = 2.91) on the verb. The difference in distance is lower for the external eyebrows than for the internal eyebrows (but note that we did not quantitatively compare these differences).

In wh-questions, we find that the part of the sentence influences the eyebrow distance (ANOVA χ^2 for internal = 7.362, df = 2, p<0.05, for external = 6.824, df = 2, p <0.05).

Internal eyebrows distance on the subject is bigger than on the verb by 0.5 mm (*se* = 0.5, *t* = 1.09). The average between internal eyebrows distance on subject and verb is less than on the wh-word by 0.73 mm (*se* = 0.38, *t* = 1.9). External eyebrows distance on the subject is bigger than on the verb by 0.68 mm (*se* = 0.47, *t* = 1.43). The average between external eyebrows distance on subject and verb is less than on the wh-word by 0.59 mm (*se* = 0.4, *t* = 1.43).

To sum up, we confirmed our initial hypothesis that polar questions and wh-questions are marked with eyebrow raise both internal and external. Besides, we find indications that the contour of the raise is different in these sentence types, even though we have not tested the significance of these differences as the estimates come from different models. The eyebrow raise in polar questions is higher on the verb than on the subject, whereas in wh-questions it is higher on the subject than on the verb. Also, the raise itself is smaller in wh-questions.

Moreover, we analyzed the eyebrow raise in wh-questions separately and found out that eyebrows raise starts at the wh-word and then gradually decreases.

6.2 Head movement

As we stated before, vertical head tilts are measured in radian angles on the x-axis. A positive angle means forward tilt, whereas a negative angle means head tilt backwards.

We find that sentence type influences the head rotation angle on the subject and the verb (ANOVA χ^2 for subject 8.819, df = 2, p <0.05, for verb 10.462, df = 2, p <0.01). The head rotation angle is bigger for wh-questions than statements on the subject by estimated 0.006 radians (*se* = 0.029, *t* = 0.22), and on the verb by estimated 0.03 radians (*se* = 0.02, *t* = 1.4). The average between statement and wh-question is lower than polar questions by estimated 0.13 radians (*se* = 0.046, *t* = 2.9) on the subject and 0.2 radians (*se* = 0.06, *t* = 3.1) on the verb.

In wh-questions, we find that the part of the sentence significantly influences the head rotation angle (ANOVA χ^2 = 39.887, df = 2, p <0.001). Rotation angle on the subject is less by estimated 0.06 radians (*se* = 0.02, *t* =-2.6) than on the verb. Rotation angle is less on the wh-word than on the average between subject and verb by 0.16 radians (*se* = 0.04, *t* = -4.19).

To conclude, as we saw in Figure 2, polar questions differ from wh-questions and statements regarding head tilting forward on the subject and verb. Meanwhile, the difference between wh-questions and statements is not significant (0.006 radians and 0.032 radians on subject and verb respectively). However, when examining wh-questions separately, we confirmed that the wh-word is marked with a head tilt backwards (-0.16 radians) in contrast to the other part of the sentence, and this difference is significant.

7 Discussion

7.1 Non-manual question marking in the KRSL dataset

Our study provides a first description of non-manual question marking in KRSL by analyzing a dataset created for NLP purposes. As discussed above, it is inadvisable to generalize our findings to naturalistic use of KRSL.

Both manual annotations and the quantitative analysis demonstrate that polar questions in the dataset are marked by forward head tilt on the subject and the verb, and eyebrow raise on the whole sentence; wh-questions are marked by backward head tilt on the wh-sign, and by eyebrow raise on the wh-sign that can also spread to the whole sentence, but the raise is smaller than in polar questions. This dataset thus neatly fits the most common typological pattern for non-manual marking of questions in sign languages (Cecchetto, 2012).

7.2 Applicability of computer vision

This study demonstrates that it is possible to apply modern computer vision tools to analyze non-manual markers in sign languages quantitatively.

OpenFace (Baltrušaitis et al., 2018, 2013; Zadeh et al., 2017) provides a solution to the problem of using 2D video recordings for analysis by reconstructing a 3D model of the face from the 2D representation. However, our experiments show that the reconstructed 3D model is still sensitive to distortions due to some types of movement (specifically, due to forward and backward head tilts). We developed a solution for this problem by applying machine learning in order to teach a new model to account for the bias introduced by the head tilts.

Based on this experience, we can also offer a **practical recommendation** for linguists planning to use OpenFace or similar tools for the analysis of non-manual markers. In case the study involves novel data collection, it would be very useful to record each subject rotating their head in various directions without moving the eyebrows or any other articulators on the face. These recordings can be later use to train a model similar to the one described in this study to correct for distortions due to head tilts. In our data set we fortunately had some recordings that could be used as such a training data set, but it is better to plan for such a data set directly.

8 Conclusions

This paper presents the analysis of non-manual marking of simple polar and wh-questions in KRSL produced by nine native KRSL signers for a dataset for automatic sign language translation. To this end, we firstly annotated the data set manually, and then applied computer vision techniques to automate extraction of non-manual marking from video recordings.

Our findings suggest that polar questions in the KRSL dataset are marked by an eyebrow raise on the whole sentence, and by consecutive forward head tilts on the subject and the verb. In addition, wh-questions are marked by backward head tilts on the wh-sign, and by an eyebrow raise on the wh-sign that can spread over the whole sentence.

Additionally, we demonstrated the utility of computer vision solutions, specifically, Open-Face (Baltrušaitis et al., 2018, 2013; Zadeh et al., 2017) that can be applied to sign language data for the purpose of linguistic analysis of non-manual marking. However, we also discovered that head movement leads to distortion of the facial features even though OpenFace reconstructs a 3D model of the face to account for such movement. We addressed this problem with a machine learning solution.

References

Baayen, H., Davidson, D., and Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59:390–412.

Proceedings of the 18th Biennial Machine Translation Summit, Virtual USA, August 16 - 20, 2021
1st International Workshop on Automatic Translation for Signed and Spoken Languages

Page 57

Baltrušaitis, T., Robinson, P., and Morency, L.-P. (2013). Constrained local neural fields for robust facial landmark detection in the wild. In *2013 IEEE International Conference on Computer Vision Workshops*, pages 354–361.

Baltrušaitis, T., Zadeh, A., Lim, Y. C., and Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pages 59–66.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv e-prints*, arXiv:1406.

Camgoz, N. C., Hadfield, S., Koller, O., Ney, H., and Bowden, R. (2018). Neural sign language translation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7784–7793.

Cecchetto, C. (2012). Sentence types. In Pfau, R., Steinbach, M., and Woll, B., editors, *Sign language: An international handbook*, pages 292–315. De Gruyter Mouton.

Chung, Y., Gelman, A., Rabe-Hesketh, S., Liu, J., and Dorie, V. (2015). Weakly informative prior for point estimation of covariance matrices in hierarchical models. *Journal of Educational and Behavioral Statistics*, 40:136–157.

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., and Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, 78:685–709.

Coerts, J. (1992). *Nonmanual Grammatical Markers. An Analysis of Interrogatives, Negation and Topicalisation in Sign Language of the Netherlands*. Doctoral dissertation, University of Amsterdam.

Crasborn, O., Bank, R., Zwitserlood, I., van der Kooij, E., Meijer, A., Sáfár, A., and Ormel, E. (2015). *Annotation Conventions for the Corpus NGT, version 3*.

Crasborn, O., Zwitserlood, I., and Ros, J. (2008). Corpus NGT. an open access digital corpus of movies with annotations of Sign Language of the Netherlands.

ELAN (2020). ELAN (Version 5.9) [Computer software]. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive.

Fox, J. and Weisberg, S. (2019). *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, third edition.

Hodge, G., Fenlon, J., Schembri, A., Johnston, T., and Cormier, K. (2019). A corpus-based investigation of how deaf signers signal questions during conversation. Poster presented at the 13th Theoretical Issues in Sign Language conference, Universität Hamburg, Germany.

Konrad, R., Hanke, T., Langer, G., Blanck, D., Bleicken, J., Hofmann, I., Jeziorski, O., König, L., König, S., Nishio, R., Regen, A., Salden, U., Wagner, S., Worseck, S., Böse, O., Jahn, E., and Schulder, M. (2020). MY DGS annotated. public corpus of German Sign Language, 3rd release MEINE DGS annotiert. Öffentliches korpus der deutschen gebärdensprache, 3. release.

Li, D., Xu, C., Yu, X., Zhang, K., Swift, B., Suominen, H., and Li, H. (2020). Tspnet: Hierarchical feature learning via temporal semantic pyramid for sign language translation.

Proceedings of the 18th Biennial Machine Translation Summit, Virtual USA, August 16 - 20, 2021
1st International Workshop on Automatic Translation for Signed and Spoken Languages

Page 58

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Édouard Duchesnay (2018). Scikit-learn: Machine learning in python.

Pfau, R. and Quer, J. (2010). Nonmanuals: their prosodic and grammatical roles. In Brentari, D., editor, *Sign Languages*, pages 381–402. Cambridge University Press.

Puupponen, A., Wainio, T., Burger, B., and Jantunen, T. (2015). Head movements in Finnish Sign Language on the basis of Motion Capture data: A study of the form and function of nods, nodding, head thrusts, and head pulls. *Sign Language & Linguistics*, 18(1):41–89.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA.

Zadeh, A., Baltrušaitis, T., and Morency, L.-P. (2017). Convolutional experts constrained local model for facial landmark detection.

Zeshan, U. (2004). Interrogative constructions in signed languages: Crosslinguistic perspectives. *Language*, 80(1):7–39.

Zimmermann, C. and Brox, T. (2017). Learning to estimate 3d hand pose from single RGB images. *CoRR*, abs/1705.01389.

Proceedings of the 18th Biennial Machine Translation Summit, Virtual USA, August 16 - 20, 2021
1st International Workshop on Automatic Translation for Signed and Spoken Languages
Page 59