Detecting names in Sign Language Corpus Data

A deaf annotator was asked to view the video and mark each occurrence of a name. The annotator had not seen the original experiment data before and was allowed to stop and review the video as often as necessary. Nineteen minutes of the sample were signed in a dialect unfamiliar to the annotator. As expected, it was harder for the annotator to detect names when informants signed in an unfamiliar dialect. But even if the concrete meaning of a name was not understood, the entity was still identified as a name. The inspection of the DGS films took five times real time of the films. It is therefore a rather costly method, but it was also the most reliable method tested.

In total, there were only four false negatives, one to be neglected, because a name was mentioned in the translation only. However, one of these cases was a name that had to be anonymized. Assuming we had relied only on the manual inspection, we had missed this entity.

Named entities specifically used in the Deaf community were identified almost exclusively by the manual inspection.

For the publication of the data, named entities have to be anonymized in different places of the transcript.

In the translations, names are replaced by numbered placeholders, e.g. #Name1, allowing the reader to follow co-references.

In the transcript, the same applies to mouting annotations and to the gloss tier in most cases.

For the video, the time span to be anonymized is annotated separately. However, some experiments showed that completely blackening that timestamp invalidates the whole sentence for further linguistic analysis as suprasegmental signals are disturbed.

Therefore, we defined several options how to manipulate a stretch of video sufficient to make the sign or mouting component unrecognizable.

For this experiment different approaches to identify named entities were applied to evaluate their reliability.

The sample consisted of 31 minutes of corpus data in total from three different conversations. As we wanted different DGS dialects to be covered in the sample, we chose informants from the North as well as the South of Germany. The dialects are reflected not only in varying signs, but also in divergent mouthings. This might make the identification of names harder for our staff members compared to a monolingual translation.

We provide data on an experiment with a part of the corpus detailing which percentage of the ground truth names are detected with each method. Lacking any better method, the ground truth is to be accomplished. (http://ffmpeg.org) to render the designated blocks black over the timespans specified.

For named entity recognition, we implemented calling pre-defined WebLicht chains into our annotation environment Live. On our data through different named entity recognizers available in WebLicht and finally kept working with two WebLicht chains that showed the best results in combination. As we were well aware that most such systems are trained on written text, we feed them with translations of face-to-face communication, we had to expect some errors, mostly false negatives. Surprisingly, sentences ending with an exclamation mark were not processed properly by the named entity recognizers used. After automatically adding a full stop after the exclamation marks, we got reliable results.

Half of the false negatives are due to the fact that the named entity recognizer were not run on the original language data but on the translations.

In the case of mouthing, only the mouth including cheeks and the chin is to be hidden. In the case of fingernailing, only the dominant hand and the surroundings covering the sideways and downwards movements potentially occurring need to be covered.

For signs in front of the head or the trunk, the whole body region needs to be hidden, as the positioning of the hand itself (let alone its movement) might suffice to identify the signer. Our experiments showed that blackening these areas is less disturbing for the viewer than a neutral good enough to really hide the sign mouthing.

In order to maintain the manual annotation, our annotation environment features some computer vision algorithms, including facemask and hand tracking reliable enough to be used for this purpose as the areas detected need to be labeled anyway. The trackers generate annotation, in this case rectangle coordinates which upon export of the movie files are used to command FFmpeg (a cross-platform multimedia processing framework, cf. http://ffmpeg.org).

For the personal inspection by a signer yielded the best results, but is rather costly. However, automatic procedures with high rates of false positives cause substantial costs for manually identifying the false alarms as such, too. Therefore, a one-pass manual inspection combined with the other automatic methods seems appropriate to gain reliable results. The combination of methods not only achieves slightly better results for the original language data than manual inspection alone, but also provides a good chance to catch names in the translation not present in the original without spending another manual inspection on the translations.