Sign Language Translation in a Healthcare Setting

L.D. ESSELINK, Institute for Logic, Language and Computation, Netherlands
S.E. MENDE-GILLINGS, Institute for Logic, Language and Computation, Netherlands
F. ROELOFSEN, Institute for Logic, Language and Computation, Netherlands
A.S. SMEIJERS, Amsterdam University Medical Centre, Netherlands

It is challenging for healthcare professionals to communicate with Deaf patients, especially in times of COVID-19. Sign language interpreters cannot always enter hospitals and clinics, and face masks make lipreading impossible.

We present an application which allows healthcare professionals to translate sentences that are frequently used in the diagnosis and treatment of COVID-19 from Dutch into Sign Language of the Netherlands (NGT). Translations are displayed by means of videos and avatar animations. The architecture of the system is such that it could be straightforwardly extended to other sign languages.

CCS Concepts: • **Human-centered computing → Accessibility technologies**; • **Applied computing →** *Health informatics*; • **Computing methodologies → Machine translation**.

Additional Key Words and Phrases: Sign language translation, sign language avatar

ACM Reference Format:
L.D. Esselink, S.E. Mende-Gillings, F. Roelofsen, and A.S. Smeijers. 2021. Sign Language Translation in a Healthcare Setting. In *Web4All 2021: 18th International Cross-Disciplinary Conference on Web Accessibility, April 19–20, 2021, Online.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

It is challenging for healthcare professionals to communicate with Deaf[1] patients [1], especially in times of COVID-19 [2]. Sign language interpreters cannot always enter hospitals and clinics, face masks make lipreading impossible [3], and interpreting via video relay is not always viable.

We present an application which allows healthcare professionals to translate sentences that are frequently used in the diagnosis and treatment of COVID-19 from Dutch into Sign Language of the Netherlands. For a limited number of sentences, translations are displayed by means of pre-recorded videos. To allow for larger coverage, however, the system is also able to generate translations that are displayed by means of an animated avatar. The paper concentrates on the latter part of the translation the system.

We would like to emphasise that a qualified sign language interpreter should, whenever available, always be preferred over a machine translation system. Still, it is worth investigating the extent to which a machine translation

[1] It is common to refer to people who are part of the deaf community and primarily communicate using sign language as Deaf people, with a capital 'D'.

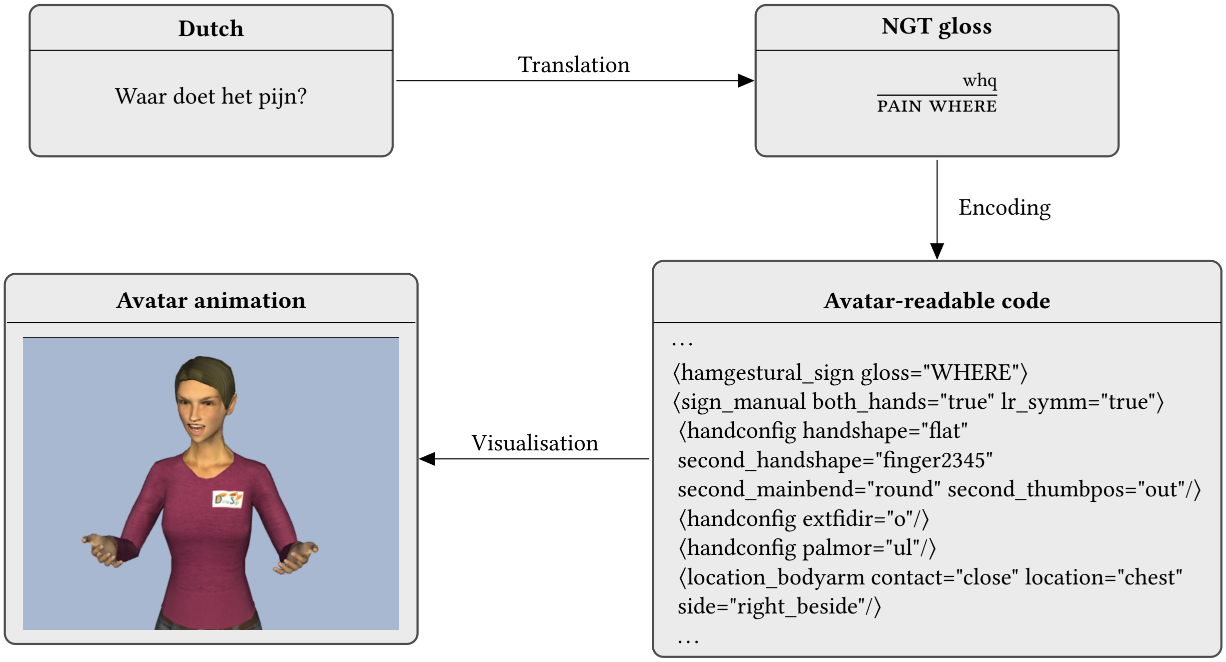

Fig. 1. Overview of the avatar translation workflow.

system making use of pre-recorded videos and avatar technology can be of help in situations in which a human interpreter cannot be employed, especially in the healthcare setting where effective, instantaneous communication between healthcare professionals and patients is of critical importance.

2 AVATAR TRANSLATION WORKFLOW

As depicted in Figure 1, three main steps can be distinguished in the avatar translation workflow: (i) a **translation step**, which consists of translating a given Dutch sentence into the corresponding NGT sentence, represented by means of a so-called gloss, (ii) an **encoding step**, which consists of encoding the NGT gloss in an avatar-readable representation language, and (iii) a **visualisation step**, which consists in converting the encoding into an avatar animation.

Suppose, for instance, we need to translate the sentence in (1).

(1) Waar doet het pijn?
'Where does it hurt?'

The first step is to convert this sentence into the corresponding NGT gloss in (2). In such glosses, the English words in smallcaps stand for manual signs, while the upper tier shows non-manual markers (facial expressions, body/head positions), and the horizontal line indicates the duration of these non-manual markers. The non-manual marker 'whq' in (2) is the one that is characteristic for constituent questions in NGT. While empirical studies have found copious variation in the actual realisation of 'whq' in NGT [4, 5], furrowed eyebrows are usually taken to be characteristic for the most canonical realisation.

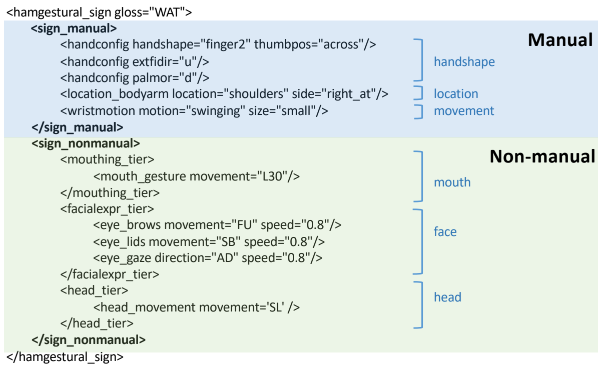

Fig. 2. SiGML encoding of the NGT sign WAT (‘what’).

(2) $\overline{\text{PAIN WHERE}}^{\text{whq}}$

This gloss representation, however, is not readable to software that can generate avatar animations and thus the second step is to encode it in a suitable format. Our system makes use of Java Avatar Signing (JASigning)[2] [8, 9], a software package that uses a representation language called Signing Gesture Markup Language (SiGML) [10, 11]. The completion of the encoding step therefore requires the conversion of the gloss into SiGML. Both manual signs and non-manual markers are encoded.

In the final step, the SiGML encoding is visualised by an avatar. JASigning reads the SiGML encoding which tells it which manual signs to express in order and which of them are accompanied by non-manual markers. JASigning then generates the corresponding movements and orientations of the hands, arms, torso, head and (parts of the) face of the avatar.

3 APPROACH AND MAIN CONTRIBUTIONS

Our objective was to address the immediate need, ensuing from the current COVID-19 crisis, for a translation tool to aid healthcare professionals in communicating with Deaf patients. Two requirements follow from this objective: (i) the tool had to be developed within a short time-frame, and (ii) high accuracy of the delivered translations was more important than broad approximate coverage.
Therefore, our aim has *not* been to automate the entire translation workflow. In particular, automating the process of mapping input sentences to the corresponding NGT glosses would not have been feasible within a short time-frame,[3] and would, even in the somewhat longer term, most likely result in an unacceptably low accuracy rate. This research therefore focused on the (semi-)automation of the encoding step.

[2]Another avatar engine, Paula, provides high quality sign language synthesis, but has been developed specifically for ASL and cannot be straightforwardly extended to other sign languages [6, 7].
[3]Prins and Janssen [12], who investigated the feasibility of automated sign language translation for children television programs, drew the same conclusion.

3.1 Corpus building

We determined a set of statements and questions that are commonly used during diagnosis of COVID-19 and overall care for Deaf patients based on the experience of other automatic healthcare translation systems such as SignTranslate [13]. These sentences were then divided into three categories: video-only, avatar-only, and hybrid.

The first category, video-only, consisted mainly of sentences that could be divided into three further categories: emotional, complex and informed consent. Emotional sentences are those that concern how the patient is feeling and therefore require emotion and empathy to be conveyed. This is difficult to achieve in a satisfactory way with an avatar given the current state of the art, so we deemed that video translations were necessary for these sentences. Sentences were classified as complex when they involved a combination of several statements and/or questions, or required a demonstration of pictures or diagrams along with an explanation. Finally, in the case of questions and statements concerning informed consent it is especially important to leave no room for potential misunderstandings. To ensure this, we chose to always offer video translations of these sentences.

The second category, avatar-only, consisted of sentences with many variations differing by only one word or phrase, indicating for instance the time of day or a number of weeks. It would not have been feasible to record a video translation for each version of these sentences.

The third category, hybrid, consisted of sentences that do not fall into one of the other two categories. For these, the system offers both a video translation and an avatar translation. In some cases, the avatar translation is slightly simplified compared to the video translation.

After categorising all of the sentences, those from the first and third category were translated into NGT and recorded by a team consisting of a sign language interpreter and a native Deaf signer. The Deaf signer that is visible in the videos was chosen due to her neutral reputation within the Deaf community, and her clear signing style without a specific dialect. The level of complexity of the sentences involved was not deemed to require validation through backwards translation [14]. Translations were checked by a sign linguist who is also a medical doctor. This resulted in a corpus of 139 video translations. The sentences from the second and third category together comprise 7720 sentences for avatar translation.

3.2 Encoding signs

We determined which signs needed to be encoded for the avatar to be able to correctly translate all of the sentences. For some of these signs, an encoding was taken from an already existing corpus of SiGML representations constructed in previous research [15–17]. Based on the video corpus, we encoded the manual signs and non-manual markers for the additional signs in SiGML. If a sign was not present in the video corpus, we utilised the online dictionary of the Dutch Sign Language Centre[4], which contains videos of native signers signing individual words in NGT.

For illustration, the SiGML encoding of the sign WAT ('what') is given in Figure 2. As indicated, both manual components (handshape, location, movement) and non-manual features (mouth, face, head) are encoded.

[4]https://www.gebarencentrum.nl

3.3 Constructing sentences

In order for the application to operate as fast as possible, we chose to pre-process all sentences and store their translations in a database. At run-time, the application only queries this database and does not compute any translations on the fly.

To construct the translations of full sentences, we created a program that, when given the NGT gloss representation of a sentence, creates a file containing the correct SiGML code for that sentence. Every question was encoded with the 'general interrogative sign' in NGT (palms up, raised eyebrows) at the end of the sentence. Although this sign is not always necessarily used in questions by sign language users, we expect that it increases comprehension in the case of avatar translations. Moreover, in the case of polar (yes/no) questions, the program changes the non-manual component of the last sign before the general interrogative sign to include raised eyebrows, in line with what we observed in the video translations in our corpus.

3.4 User interface

We developed an online user interface for the translation system. The user can choose whether they prefer video translations or avatar translations, and type a sentence they wish to have translated into NGT, or a sequence of search terms. Based on their input they are presented with a list of available sentences from the database. These sentences differ depending on the translation mode chosen (video or avatar). After selecting one of the sentences the translation is then offered in the chosen format.

As mentioned earlier, some of the possible input sentences differ only in one word or phrase. These sentences can be thought of as involving a general template with a variable that can take several values, such as a day of the week, a number (of times, minutes, hours, days, weeks or months), or a time of day. When a user wants to translate such a sentence, they first select the template and then provide the intended value for the variable. For example, they may select the template "I am going to explain more at *time*", and then select a particular time. The time is specified in digital format in order to be able to distinguish between morning, afternoon, and evening. The program translates the digital time to analog, and adds 'morning', 'afternoon', or 'evening' to the sentence if that is deemed necessary given the current time of day. For example, if the current time is 14:00 and the user selects 15:00, 'afternoon' is not added, but if the user selects 03:00, then 'morning' is added for clarification.

While JASigning in principle offers a number of different avatars for sign language animation, there are differences in execution between these avatars. Our user interface therefore only makes use of one of the avatars, and does not allow the user to choose between different options. Optimising the avatar visualisation has not been within the scope of our project thus far but would certainly be desirable, both in the healthcare setting and for further applications.

4 DISCUSSION

This project has lead to an application that can aid communication between healthcare professionals and Deaf patients in situations in which a qualified sign language interpreter is not available. The application offers translations of questions and statements that are frequently used in a healthcare setting, in particular in the diagnosis and treatment of COVID-19. Translations are displayed in video format or by means of an animated avatar, or both.

It is important to consider the respective advantages and limitations of the two translation formats. One advantage of avatar technology is that it provides flexibility and scales up more easily than video translation. Once the sign for a certain word has been encoded in order to translate a given sentence, it can immediately be re-used to translate other sentences including that same word. This makes it particularly straightforward to generate translations of sentences that differ only slightly from each other (e.g., in a phrase indicating the time of day).

The disadvantage, however, of avatar translations is that they are currently less natural and in some cases also more difficult to comprehend. Some signs differ from each other only in very subtle ways, and when produced by an avatar may become indistinguishable. Certain facial expressions and body movements of the avatar are quite unnatural, which can add to the difficulty of understanding the translations. Certainly, given the current state of the art, sign language avatars cannot properly display empathy. This makes it undesirable to use avatar translations in situations where such empathy is required, as is often the case in medical settings.

Video translations, on the other hand, have their own benefits and drawbacks. Evidently, they are better than avatar translations in terms of naturalness and comprehensibility, especially in the case of complex sentences. Moreover, it could well be the case that a patient feels more comfortable watching a video of a human interpreter rather than an animated avatar in a situation in which their physical well-being is at stake.

The main disadvantage of a video translation system is its inflexibility and inability to scale up in an efficient way. All translations have to be recorded separately, even ones that are almost identical. Cutting and pasting video fragments of individual signs to create new sentences does not yield satisfactory results.

Another disadvantage of a video translation system, though possibly less significant, is the difficulty of maintaining consistency among the translations that are offered. Each human interpreter has their own signing 'style'. Therefore, if the interpreter often changes from one video translation to the next, the patient will have to constantly re-adjust to a new signing style. Unfortunately, as the number of sentences in a corpus grows, it becomes less realistic to use the same interpreter for all translations.

A general advantage that a machine translation system (using either pre-recorded videos, or avatar technology, or both) may sometimes have over a human interpreter, especially in the healthcare domain, concerns privacy. A patient may receive sensitive information, and may not want this information to be known to anyone else. In this case, employing a human interpreter has a disadvantage (though this may of course be outweighed by the higher level of translation accuracy and empathy that can be provided by a human interpreter).

Important to note is that constructing sign language translations in either format is a time-consuming affair, though for different reasons. Building a corpus of video translations is time intensive because every translation has to be recorded separately. For avatar translations, it takes time to encode individual signs. These are reusable, however, which becomes especially attractive as the number of required translations grows. However, the overall preference for one method over another is context-dependent: pros and cons should be carefully weighed in each situation.

To conclude, this research has aimed to investigate the potential of a (semi-)automatic sign language translation system in a healthcare setting, addressing in particular the challenges that the current pandemic implies for the communication between healthcare professionals and Deaf patients. It has brought to light various prospects and limitations of such a system. The application is still under development and the next step will be to have it evaluated by people within the Deaf community and by healthcare professionals.

Other lines of future work include the development of applications in different domains, such as the translation of announcements at airports or railway stations, which has already been explored to some extent for other sign languages [9, 18], and a translation system that can support hearing parents of deaf children, who need to learn sign language as fast as possible but usually have limited access to classes and proficient conversation partners.

ACKNOWLEDGMENTS

We are grateful to Manny Rayner, Marijke Scheffener, Irène Strasly, and Inge Zwitserlood for helpful discussions at various stages of this project. Special thanks to Bastien David, John Glauert, and Richard Kennaway for their very generous help with JASigning and SiGML. Finally, we gratefully acknowledge financial support by the Netherlands Organisation for Innovation in Healthcare (ZonMw) under grant number 10430042010027.

REFERENCES

[1] Johannes Fellinger, Daniel Holzinger, and Robert Pollard. Mental health of deaf people. *The Lancet*, 379(9820):1037–1044, 2012.

[2] Michael McKee, Christa Moran, and Philip Zazove. Overcoming additional barriers to care for deaf and hard of hearing patients during covid-19. *JAMA Otolaryngology–Head & Neck Surgery*, 146(9):781–782, 2020.

[3] Helen Grote and Fizz Izagaren. Covid-19: The communication needs of D/deaf healthcare workers and patients are being forgotten. *British Medical Journal*, 369, 2020.

[4] Jane Coerts. *Nonmanual grammatical markers: an analysis of interrogatives, negations and topicalisations in Sign Language of the Netherlands*. PhD thesis, University of Amsterdam, 1992.

[5] Connie de Vos, Els van der Kooij, and Onno Crasborn. Mixed signals: Combining linguistic and affective functions of eyebrows in questions in Sign Language of the Netherlands. *Language and Speech*, 52(2-3):315–339, 2009.

[6] Mary Jo Davidson. Paula: A computer-based sign language tutor for hearing adults. In *Intelligent Tutoring Systems 2006 Workshop on Teaching with Robots, Agents, and Natural Language Processing*, pages 66–72, 2006.

[7] John McDonald, Rosalee Wolfe, Jerry Schnepp, Julie Hochgesang, Diana Gorman Jamrozik, Marie Stumbo, Larwan Berke, Melissa Bialek, and Farah Thomas. An automated technique for real-time production of lifelike animations of American Sign Language. *Universal Access in the Information Society*, 15(4):551–566, 2016.

[8] Vince Jennings, Ralph Elliott, Richard Kennaway, and John Glauert. Requirements for a signing avatar. In *Workshop on Corpora and Sign Language Technologies at the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, pages 33–136, 2010.

[9] Sarah Ebling and John Glauert. Building a Swiss German Sign Language avatar with JASigning and evaluating it among the Deaf community. *Universal Access in the Information Society*, 15(4):577–587, 2016.

[10] Ralph Elliott, John Glauert, Vince Jennings, and Richard Kennaway. An overview of the SiGML notation and SiGML signing software system. In *Workshop on the Representation and Processing of Sign Languages at the Fourth International Conference on Language Resources and Evaluation (LREC)*, pages 98–104, 2004.

[11] John Glauert and Ralph Elliott. Extending the SiGML notation–a progress report. In *Second International Workshop on Sign Language Translation and Avatar Technology (SLTAT)*, 2011.

[12] Martin Prins and Joris B. Janssen. Automated sign language. TNO technical report, 2014.

[13] Anna Middleton, Alagaratnam Niruban, Gill Girling, and Phyo Kyaw Myint. Communicating in a healthcare setting with people who have hearing loss. *Bmj*, 341, 2010.

[14] Anika S. Smeijers, Beppie van den Bogaerde, M. Ens-Dokkum, and A.M. Oudesluys-Murphy. Scientific-Based Translation of Standardized Questionnaires into Sign Language of the Netherlands. *Investigations in healthcare interpreting*, pages 277–301, 2014.

[15] Adriana J. Corsel. Multidimensionality in Sign Language Synthesis: Translation of Dutch into Sign Language of the Netherlands. Bachelor's thesis. University of Amsterdam, https://scripties.uba.uva.nl/search?id=715771, 2020.

[16] Lyke Esselink. Lexical resources for sign language synthesis. Bachelor's thesis. University of Amsterdam, https://scripties.uba.uva.nl/search?id=715792, 2020.

[17] Shani E. Mende-Gillings. The signing space for the synthesis of directional verbs in NGT. Bachelor's thesis. University of Amsterdam, https://scripties.uba.uva.nl/search?id=715794, 2020.

[18] Cristina Battaglino, Carlo Geraci, Vincenzo Lombardo, and Alessandro Mazzei. Prototyping and preliminary evaluation of a sign language translation system in the railway domain. In Margherita Antona and Constantine Stephanidis, editors, *Universal Access in Human-Computer Interaction*, pages 339–350, 2015.