



HAL
open science

The myth of signing avatars

Rosalee Wolfe, John C Mcdonald, Eleni Efthimiou, Evita Fotinea, Frankie Picron, Davy van Landuyt, Tina Sioen, Annelies Braffort, Michael Filhol, Sarah Ebling, et al.

► To cite this version:

Rosalee Wolfe, John C Mcdonald, Eleni Efthimiou, Evita Fotinea, Frankie Picron, et al.. The myth of signing avatars. 1st International Workshop on Automatic Translation for Signed and Spoken Languages, Aug 2021, Online streaming, France. hal-03375968

HAL Id: hal-03375968

<https://hal.science/hal-03375968>

Submitted on 13 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The myth of signing avatars

Rosalee Wolfe Rosalee.Wolfe@athenarc.gr
Institute for Language and Speech Processing, Athena RC, Greece

John C. McDonald jmcdonald@cs.depaul.edu
School of Computing, DePaul University, Chicago, USA

Eleni Efthimiou eleni_e@athenarc.gr
Institute for Language and Speech Processing, Athena RC, Greece

Evita Fontinea evita@athenarc.gr
Institute for Language and Speech Processing, Athena RC, Greece

Frankie Picron frankie.picron@eud.eu
European Union of the Deaf, Brussels, Belgium

Davy Van Landuyt davy.van.landuyt@eud.eu
European Union of the Deaf, Brussels, Belgium

Tina Sioen tina.sioen@eud.eu
European Union of the Deaf, Brussels, Belgium

Annelies Braffort annelies.braffort@lisn.upsaclay.fr
Laboratoire Interdisciplinaire des Sciences du Numérique, Orsay, France

Michael Filhol michael.filhol@lisn.upsaclay.fr
Laboratoire Interdisciplinaire des Sciences du Numérique, Orsay, France

Sarah Ebling ebling@cl.uzh.ch
Department of Computational Linguistics, University of Zurich, Switzerland

Thomas Hanke thomas.hanke@uni-hamburg.de
Institut für Deutsche Gebärdensprache, Universität Hamburg, Germany

Verena Krausneker verena.krausneker@univie.ac.at
Institut für Sprachwissenschaft, Universität Wien, Vienna, Austria

Abstract

Development of automatic translation between signed and spoken languages has lagged behind the development of automatic translation between spoken languages, but it is a common misperception that extending machine translation techniques to include signed languages should be a straightforward process. A contributing factor is the lack of an acceptable method for displaying sign language apart from interpreters on video. This position paper examines the challenges of displaying a signed language as a target in automatic translation, analyses the underlying causes and suggests strategies to develop display technologies that are acceptable to sign language communities.

1. Introduction

Deaf sign language users around the world face continual challenges in daily interaction with hearing, non-signing populations. The gold standard for translating between signed and spoken

languages¹ are certified sign language interpreters who are essential to facilitating communication for education, healthcare, and legal consultation among other situations. However, many transactions in daily living consist of short conversations over a hotel desk, at a store counter or in an office foyer. These interactions are so limited in scope and duration that hiring a qualified interpreter would be prohibitively expensive or quite unnecessary, or even impossible because in most countries there is a shortage of qualified interpreters. In such situations, an automatic translation system between spoken and signed language would ease communication barriers and improve inclusivity. For technology of this sort to be useful, it must display sign language in a way that is acceptable to members of the sign language community.

To be effective, an automated translation system or machine translation system must be able to produce legible, grammatically, and phonologically and phonetically correct, acceptable utterances in a desired target language with minimal or no human involvement. Researchers have made significant progress in translating between high-resource languages that have a written form and some have suggested that automatic translation has achieved human parity in some domains (Hassan, et al., 2018).

Progress in translating between signed and spoken languages has lagged significantly in comparison. Traditionally, this task has been conceived as one of text-to-text translation, involving written representations of sign languages. Since sign languages have no widely accepted written form, an additional required step in going from a spoken language to a sign language is that of displaying signed languages in their natural moving form, in the visual modality (Ebling, 2016). This position paper examines the challenges of displaying signed language as a target in automatic translation, analyses the underlying impediments and suggests strategies to develop display technologies that are acceptable to deaf sign language users.

2. Background

Sign languages are distinct from their surrounding spoken languages. For example, in France, many deaf persons have Langue des Signes Française (LSF), not French, as their preferred language. Since French is a second language to them, even its written form poses a barrier. Many researchers have noted that written language poses barriers to members of the Deaf communities (Traxler, 2000; Gutjahr, 2006; Hennies, 2010; Konrad, 2011).

Deaf sign language users consider themselves members of a minority group, with a distinct language, culture, and shared experiences, rather than as simply persons with a disability (De Meulder, Krausneker, Turner, & Conama, 2019). They continually struggle with the reality that policy makers in governmental departments, educational institutions and health care agencies consist primarily of hearing people who are not familiar with the values, goals and concerns of sign language communities (Branson & Miller, 1998). As a result, there is a history of disenfranchisement which adds a barrier of distrust to the barrier of language that exists between deaf and hearing communities. At present, current technology claiming to translate between spoken and signed languages are not viewed favourably by sign language communities. Rather, the technology is often perceived as a ploy to replace human interpreters (World Federation of the Deaf, 2018; European Union of the Deaf, 2018), or even as cultural appropriation by predominantly hearing researchers, who do not always have linguistic knowledge of these languages, and often have little connection with sign language communities (Erard, 2017).

¹ The term *spoken language* refers to any language that is not signed, whether represented as speech or as text.

Linguists have noted that as long as avatars are only capable of artificial and flawed language, they are very likely to be counterproductive. (Austrian Association of Applied Linguistics, 2019).

This scepticism and often downright hostility towards automatic translation systems is exacerbated by the generally poor quality of their sign language (Sayers, et al., 2021). To date these have exhibited robotic movement and are mostly unable to reproduce all of the multi-modal articulation mechanisms necessary to be legible. They are comparable to early speech synthesis systems which featured robotic-sounding voices that chained words together with little regard to coarticulation and no attention to prosody.

3. Quality of the target language

Just as with text-to-text translation applications, users will judge the quality of the application by the quality of its output to the target language. The same is true when the target language is signed. Poor-quality signing is difficult to understand, just as poor-quality speech synthesis or egregious misspellings are difficult to understand. It undermines the viewer's confidence in the quality of the translation. Worse, poor quality signing alienates the sign language community. Being forced to struggle with the poor signing is no better than being forced to lip read or use captions in the second language.

This is simply more evidence that reconfirms a continuing disenfranchisement. For these reasons, quality of the ultimate signed language display must be given highest priority in a spoken to signed translation system. The motion should be indistinguishable from that of a human signing the same utterance. This visual Turing Test should be the ultimate goal of any sign language display.

4. Sign language in automatic translation services

Among the challenges to acceptable sign language display as part of an automatic translation system, three issues stand out. These are 1) the difference of modalities between signed and spoken languages 2) the representation used to characterize sign languages and 3) the development of the technology required to display sign languages.

4.1. Modality

The modality of sign languages differs markedly from that of spoken languages, which utilize the vocal apparatus for production, and hearing for reception. Spoken languages use visible communicative behaviours like gestures as well, but listeners can comprehend audio-only sources. In contrast, signed languages use only visible actions for production, and vision for reception. Whereas speech utilizes a single vibrating column of air for producing utterances, signed languages use the configuration and movement of multiple body parts concurrently, including hands with all the fingers, head, face, eyes, and torso.

All sign languages have linguistic processes that are not linearly ordered. For example, in American Sign Language (ASL) the appearance of pursed lips in conjunction with the sign SMOOTH intensifies the degree of smoothness. In signed language, layers of processes ranging from the phonological to the prosodic can co-occur (Crasborn, 2006). Co-occurrence is a more general term than synchronized or simultaneous, as co-occurring events do not necessarily start or end at the same time, but they overlap in their duration.

Although there are many discrete lexical items in signed languages, much information is conveyed through forms with infinite variability and depiction, unlike fixed dictionary signs. A case in point are classifiers, which represent general categories or "classes" of objects. They

can be used to describe the size and shape of an object, and they can also represent how an object moves or is utilized. Through the use of classifiers, a signer can describe a scenario with few discrete lexical items. The signer creates an image in space. This is not simply an informal gesture as there are well-documented linguistic rules governing classifier usage (Lepic & Occhino, 2018). These are evocative, not necessarily iconic, and are extremely powerful. In a story about a motorcycle ride (Dudis, 2004), a signer can use an instrument classifier to indicate that the rider is revving the engine and a vehicle classifier to show the rider driving away on a hilly highway (Figure 1).



Figure 1. Classifier usage (Dudis, 2004).

The presence of multiple articulators that can co-occur and classifier usage are examples of the stark difference between signed and spoken languages. For these reasons, it is essential to avoid the trap of casting the problem of signed/spoken translation as a case of simply retrieving lexical items or phrasal units from a dictionary and concatenating them.

4.2. Representation

The second of the three challenges is the question of representation. Languages commonly processed by automatic translation systems have a written form. Signed languages do not. They are languages and cultures that have been preserved and transmitted from generation to generation by “hand to eye to hand”. Determining a standard transcription/annotation system that can capture all of the linguistic information contained in a signed message is still an open question. A linear stream of glosses, even with accompanying superscript strings to indicate prosody and syntax (Adamo-Villani & Wilbur, 2015), does not contain the entire semantic content of a signed utterance, in particular the depicting and spatialized linguistic structures.

This is not analogous to the difference between reading printed text on a page and witnessing an actor perform the text. Less information is captured in a gloss stream than is conveyed in written text. A hearing person may argue that not all features of articulation are captured in a printed sentence of a spoken language, such as speed of delivery, but in languages where adverbs are not necessarily expressed as separate lexical items, the lack of a speed indication is losing semantic information, not just performance information.

4.3. Sign language display

The third challenge is the display of a sign language when it is the target. The most commonly used strategy for this purpose is avatar technology. Three-dimensional avatars have the

advantages of consistency and flexibility. When recording a human signer with traditional video, special care must be taken to ensure consistency of the studio set up and the appearance of the signer between recording sessions. This requires additional time and money. When using an avatar, the lighting and camera set up can be fixed; the clothing can be chosen by the viewer as can the hair and makeup. No additional resources are required to ensure consistency.

In addition, avatars have the advantage of flexibility through the use of animation techniques. They can display co-occurring linguistic processes. Proper application of coarticulation can provide smooth transitions and can inflect signs according to syntactic rules. These properties are necessary for a translation system to produce novel utterances.

Avatars also have flexibility in appearance. They can be easily adapted to look like the shape of the original speaker/source, like a presenter or a cartoon character or a movie character. This flexibility in appearance can also anonymize a signer, so that the signer's identity will remain hidden.

Another advantage of this type of anonymization of content is that it covers one of the key properties of written language, which is inherently more anonymous than a live performance that is spoken or signed. With an anonymously presented avatar, content can be communicated without knowing the person who expressed it.

5. The promise and mythology of avatars

Given that there is a century's worth of development in animation, and nearly half that supporting video game technology, it would be tempting to dismiss the question of using avatars to display sign languages as a solved problem. However, a closer analysis shows that there are still significant challenges yet to be fully addressed.

Animation, the precursor to avatar technology, is powerfully communicative. Animation artists abstract and emphasize the salient features of a character for greater audience appeal and engagement. Simplification of a character's appearance is vital to maximizing emotional impact. This is the reason that the eyes of Disney cartoon characters are twice the size of those of a human and spaced more widely apart.

However, the requirements for sign language display are different from those for portraying cartoon characters. Beyond communicative power, display of sign language requires precision. It must adhere more closely to physical reality. For example, the hands of animation characters such as Mickey Mouse or Homer Simpson have only three fingers. For a hearing audience, this is perfectly acceptable, but three fingers aren't enough to distinguish between the fingerspelled letter W and the number 4 (Figure 2). Another consideration is that while character animation effectively uses the face and body to express emotion, the facial animation is typically at a lower quality than what would be required to portray a sign language legibly.

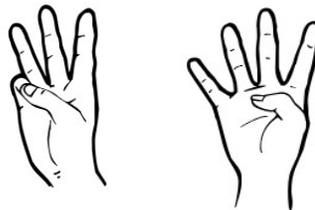


Figure 2. The difference between the letter W and the digit 4 would disappear in a three-fingered character.

Several ground-breaking animations have received attention and praise from sign language communities (Stewart, 2008; Fundación Fesord CV, 2007). These were manually created by artists with the assistance of motion capture. The artists create underlying natural processes of

coordinated muscle action, coarticulation at a biomechanical level and ambient movement. While creating the animation the artists are continually checking whether the animation draft effectively communicates the intended message and editing the draft when there are flaws. However, animations are intended for playback only and are not extensible without manual intervention. Once completed, they are archived, and without additional manual editing cannot be utilized for generating new utterances. In short, animations are not created in real time and are not interactive.

In contrast, video game characters move in response to player input in real time and are highly interactive. Thus, using video game technology might seem like an expedient approach to sign language display for a translation system. However, many game players continue to comment on the poor quality of the game characters. This is due to the effect of the uncanny valley (Tinwell, 2014). If a character appears more human-like, viewers expect the character to behave in a more human-like manner. But because the character's motion cannot be refined and edited by human animators before it is displayed, the results are unsatisfying. As explained by a professional animator (Trentskiroonie, 2015),

For something like film or television, I could create a kickass animation of a monster jumping off a building and landing on the street below, but to do the same thing in a game, the movement has to be broken up into separate parts. This is because he probably won't do the exact same action every time. There may be buildings of different heights in the game, so I can't hard-code the height of the jump into the animation. I have to create an initial jump animation, then an idle hang-time animation to play while he's in the air, and then a landing animation. The programmer then strings the jump, hang-time, and landing together and decides the timing and trajectory of the hang-time part procedurally. ... That takes artistic control away from the animator and can result in some fugly animation.

Unfortunately, a "fugly" motion on a sign language avatar can destroy the legibility and even the meaning of the message, thus making the avatar bothersome or even useless for a deaf sign language user. Finally, the representation of signed languages through avatars will have an effect on the hearing perception of these minority languages. Hearing viewers should not be confronted with "fugly" signed texts and be misled into thinking that it is real sign language in all its beauty and richness.

The analysis of the requirements for a sign language avatar shows that it must have the expressivity of manual animation but the flexibility of a video game character. These two requirements are in conflict. It is still an open question as to how to reconcile these goals.

6. Moving forward

The establishment of a set of best practices would be a substantive step toward the development of better sign language displays in automatic translation systems, but it cannot happen without a mutual collaboration with sign language stakeholders (Tupi, 2019). Deaf leadership is vital for the establishment of a validated methodology for user evaluation of avatar technology. Once created and reviewed, the methodology should be made publicly available to all researchers working in this area. Currently in Austria, there is a small research project aiming to create a Best Practice Protocol for the use of signing avatars (Krausneker, 2021).

This is consistent with the World Federation of the Deaf's position paper on Sign Language Work (World Federation of the Deaf, 2014).

The WFD considers exclusion of Deaf Community and their national organizations from sign language work ... a violation of the linguistic human rights of deaf people. Decisions regarding sign languages should always remain within the linguistic community, in this case deaf people.

Best practice for reviewing research papers would include an awareness of the multidisciplinary qualification required. It is not enough to know about machine translation. Reviewers must also be aware of sign language linguistics, the deaf experience and previous work in sign language machine translation.

When reporting on an advance in sign language avatar technology, researchers should include a sample of the sign language produced by the technique outlined in a paper. Since the sample would necessarily contain motion, it could either take the form of a media file in a commonly available format such as MPEG-4, or a web application available online. Conference organizers and journal editors need to collaborate with academic and professional organizations to archive media accompanying research papers.

7. Conclusion

“Together, we are strong.” -- Lutz König, Hamburg, 14 November 2017

Together, machine translation (MT) researchers, sign language linguists and the deaf sign language community have the potential to form powerful partnerships to educate policy makers (Bragg, et al., 2019). Ideally, Deaf professionals should be educated, supported, and actively sought to include in sign language relevant research projects.

To hearing researchers: Get to know members of sign language communities and learn about deaf culture.

- Take a class in the national sign language of your country. You already know several spoken languages -- why not discover an entirely new world? Or if you don't feel you have time,
- Go to a deaf event -- see a play in sign language, go to a deaf trade show.
- When writing grant proposals that include work relevant for sign languages, include the local and/or national deaf community. Most countries in the world have a National Association of the Deaf. Include budget for interpreters.
- Listen. Just because an idea or a result is incredibly appealing to an MT researcher does not mean that it will be useful or welcomed within the sign language community. Take feedback seriously and act on it.

Through exchange of ideas and concerns, the sign language community can inform MT researchers about their priorities, and MT researchers can clarify the capabilities and limitations of today's technologies. A clear understanding of priorities, expectations, potentials, and limitations will move the state of the art closer to realization of better inclusivity.

Acknowledgments

This work is supported in part by the EASIER (Intelligent Automatic Sign Language Translation) Project. EASIER has received funding from the European Union's Horizon 2020 research and innovation programme, grant agreement n° 101016982. 

Bibliography

- Adamo-Villani, N., & Wilbur, R. B. (2015). ASL-Pro: American sign language animation with prosodic elements. *International Conference on Universal Access in Human-Computer Interaction*, (pp. 307–318).
- Austrian Association of Applied Linguistics. (2019, August). *Position Paper on Automated Translations and Signing Avatars*. Récupéré sur verbal; Verband für Angewandte Linguistik Österreich: https://www.verbal.at/stellungnahmen/Position_Paper-Avatars_verbal_2019.pdf
- Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., . . . others. (2019). Sign language recognition, generation, and translation: An interdisciplinary perspective. *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, (pp. 16–31).
- Branson, J., & Miller, D. (1998). Nationalism and the linguistic rights of Deaf communities: Linguistic imperialism and the recognition and development of sign languages. *Journal of Sociolinguistics*, 2, 3–34.
- Crasborn, O. A. (2006). Nonmanual structures in sign language. Dans K. Brown (Éd.), *Encyclopedia of Language and Linguistics* (éd. 2nd, pp. 668-672). Oxford: Elsevier.
- De Meulder, M., Krausneker, V., Turner, G., & Conama, J. B. (2019). Sign language communities. Dans G. Hogan-Burn, & B. O'Rourke (Éd.), *The Palgrave Handbook of Minority Languages and Communities* (pp. 207-232). London: Palgrave Macmillan.
- Dudis, P. G. (2004). Body partitioning and real-space blends. *Cognitive Linguistics*, 15(2), 223-238.
- Ebling, S. (2016). *Automatic Translation from German to Synthesized Swiss German Sign Language*. Ph.D. dissertation, University of Zurich.
- Erard, M. (2017, November 9). *Why sign-language gloves don't help deaf people*. Récupéré sur The Atlantic: <https://www.theatlantic.com/technology/archive/2017/11/why-sign-language-gloves-dont-help-deaf-people/545441/>
- European Union of the Deaf. (2018, October 26). *Accessibility of information and communication*. Récupéré sur European Union of the Deaf : <https://www.eud.eu/about-us/eud-position-paper/accessibility-information-and-communication/>
- Fundación Fesord CV. (2007, Jan 26). *World Federation of the Deaf 2007*. Récupéré sur youtube: <https://www.youtube.com/watch?v=wW2KBXrPEdM>

- Gutjahr, A. E. (2006). *Lesekompetenz Gehörloser: Ein Forschungsüberblick*. Ph.D. dissertation.
- Hassan, H., Aue, A., Chen, C., Chowdhary, V., Clark, J., Federmann, C., . . . Zhou, M. (2018, Mar 15). *Achieving Human Parity on Automatic Chinese to English News Translation*. Récupéré sur Microsoft.com: <https://www.microsoft.com/en-us/research/uploads/prod/2018/03/final-achieving-human.pdf>
- Hennies, J. (2010). Lesekompetenz gehörloser und schwerhöriger SchülerInnen Ein Beitrag zur empirischen Bildungsforschung in der Hörgeschädigtenpädagogik.
- Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., . . . others. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5, 339–351.
- Konrad, R. (2011). *Die lexikalische Struktur der Deutschen Gebärdensprache im Spiegel empirischer Fachgebärdenlexikographie*. Gunter Narr Verlag.
- Krausneker, V. (2021). *Avatars and sign languages: Developing a best practice protocol on quality in accessibility*. Récupéré sur University of Vienna: <https://avatar-bestpractice.univie.ac.at/>
- Lepic, R., & Occhino, C. (2018). A construction morphology approach to sign language analysis. Dans *The construction of words* (pp. 141–172). Springer.
- Sayers, D., Sousa-Silva, R., Höhn, S., Ahmedí, L., Allkivi-Metsoja, K., Anastasiou, D., . . . others. (2021). *The dawn of the human-machine era: A forecast of new and emerging language technologies*. Récupéré sur LITHME: <https://lithme.eu/wp-content/uploads/2021/05/The-dawn-of-the-human-machine-era-a-forecast-report-2021-final.pdf>
- Stewart, J. (2008, July 21). *The Forest - A story in ASL*. Récupéré sur youtube: <https://www.youtube.com/watch?v=oUclQ10BsH8>
- Tinwell, A. (2014). *The uncanny valley in games and animation*. CRC Press.
- Traxler, C. B. (2000). The Stanford achievement test: National norming and performance standards for deaf and hard-of-hearing students. *Journal of deaf studies and deaf education*, 5, 337–348.
- Trentskiroonie. (2015). *Let's talk about Animation Quality!* Récupéré sur reddit.com: https://www.reddit.com/r/truegaming/comments/2x4fqy/lets_talk_about_animation_quality/

Tupi, E. (2019). Sign language rights in the framework of the Council of Europe and its member states. *Sign language rights in the framework of the Council of Europe and its member states*. Helsinki: Ministry for Foreign Affairs of Finland.

World Federation of the Deaf. (2014, February 19). *WFD statement of sign language work*. Récupéré sur World Federation of the Deaf: <http://wfdeaf.org/wp-content/uploads/2016/11/WFD-statement-sign-language-work.pdf>

World Federation of the Deaf. (2018, March 14). *WFD and WASLI statement of use of signing avatars*. Récupéré sur World Federation of the Deaf: <https://wfdeaf.org/news/resources/wfd-wasli-statement-use-signing-avatars/>