

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Jantunen, Tommi; Rousi, Rebekah; Rainò, Päivi; Turunen, Markku; Moeen Valipoor, Mohammad; García, Narciso

Title: Is There Any Hope for Developing Automated Translation Technology for Sign Languages?

Year: 2021

Version: Published version

Copyright: © Authors, 2021

Rights: CC BY 4.0

Rights url: <https://creativecommons.org/licenses/by/4.0/>

Please cite the original version:

Jantunen, T., Rousi, R., Rainò, P., Turunen, M., Moeen Valipoor, M., & García, N. (2021). Is There Any Hope for Developing Automated Translation Technology for Sign Languages?. In M. Hämäläinen, N. Partanen, & K. Alnajjar (Eds.), *Multilingual Facilitation* (pp. 61-73). Helsingin yliopisto. <https://doi.org/10.31885/9789515150257.7>

Is There Any Hope for Developing Automated Translation Technology for Sign Languages?

Tommi Jantunen¹, Rebekah Rousi², Päivi Rainò³, Markku Turunen⁴, Mohammad Moeen Valipoor⁵ and Narciso García⁶

¹ Sign Language Centre, PO Box 35, 40014 University of Jyväskylä, Finland

² Faculty of Information Technology, PO Box 35, 40014 University of Jyväskylä, Finland

³ School of Interpreting and Linguistic Access, Humak University of Applied Sciences, Ilkantie 4, 00400 Helsinki, Finland

⁴ Faculty of Information Technology and Communication Sciences, Tampere Accessibility Unit, 33014 Tampere University, Finland

⁵ Universidad Politécnica de Madrid, Campus de Montegancedo, 28660 Madrid, Spain

⁶ Universidad Politécnica de Madrid, E.T.S. Ingenieros de Telecomunicación, GTI, 28040 Madrid, Spain

{tommi.j.jantunen, rebekah.rousii}@jyu.fi,
markku.turunen@tuni.fi, paivi.raino@humak.fi,
moeen.valipoor@gofore.com, narciso@gti.ssr.upm.es

Abstract. This article discusses the prerequisites for the machine translation of sign languages. The topic is complex, including questions relating to technology, interaction design, linguistics and culture. At the moment, despite the affordances provided by the technology, automated translation between signed and spoken languages – or between sign languages – is not possible. The very need of such translation and its associated technology can also be questioned. Yet, we believe that contributing to the improvement of sign language detection, processing and even sign language translation to spoken languages in the future is a matter that should not be abandoned. However, we argue that this work should focus on all necessary aspects of sign languages and sign language user communities. Thus, a more diverse and critical perspective towards these issues is needed in order to avoid generalisations and bias that is often manifested within dominant research paradigms particularly in the fields of spoken language research and speech community.

Keywords: Sign Language, Automated Sign Language Translation, Machine Translation, Human Computer Interaction, Interaction Design.

1 Introduction

We have never really met Jack Rueter. Some of us share an auditive memory from seminars on Finno-Ugric languages and general linguistics when we misheard, misspelled, or pronounced his surname *Reuter. Decades after, we find ourselves in a larger Linguistic Universe, now acknowledging him as a distinguished researcher of a vast range of Uralic languages.

Now having this Graphemic Space at our disposal, we would like to draw Rueter's attention from Ural to another issue in which his interests are also invested – namely linguistics and computation. In this article we outline the prerequisites for the machine translation of sign languages (cf. [1]), with the modest hope that Rueter would dedicate part of his enthusiasm to furthering the development of computational infrastructure for sign languages in general. In particular, we would like to draw his attention to Finnish Sign Language.

Similarly to Jack Rueter, we have likewise worked with minority languages during those decades in between, yet from a differing standpoint. We too have explored the challenges and possibilities of developing automated translation technology – for languages that are signed. Our interest in writing this article centres on the complexity of the topic: sign languages and their diversity; and the promise that embodied expression and incorporation of non-conventional signs offers to furthering technological development in this area, and how it contributes to deeper scholarly understanding. To date, many challenges can be observed in attempts to develop this technology. These drawbacks can be seen on technical levels as much as cultural-linguistic levels.

Not only do these challenges and complexities exist on practical technical and cultural-linguistic levels, but aspects surrounding culture itself, marginalisation (inclusion-exclusion depending on language, ethnicity and varying abilities, i.e., literacies), and socio-economic factors enter into the equation. It only takes one to take a quick glimpse at the types of Sign Language Translation technology developments that have occurred thus far, to understand the nature and concentration of the in-built systemic biases that have to this point manifested, and no doubt, will take a long time to address. Similar issues are being faced across many fields of technological development, particularly in areas of intelligent systems. We see it as our responsibility to address these challenges and contribute to devising solutions that will not only rectify imbalance in the logic of the systems and their designs, but also provide better technological design and access for everyone.

The issues we discuss in the paper concern all sign languages. This is not meant to imply that all sign languages share the same resources – or are fully similar. Established, standardised sign languages (e.g. French, Swedish, American, Estonian, Danish, Finnish, Russian Sign Languages and so forth) are – despite their relatively short history – independent languages. Yet, the differences between these languages are primarily lexical, not grammatical. This means that from a general technological perspective they can be approached as an independent typological language class (for a recent discussion regarding the similarities and differences between sign languages, see [2]).

2 State-of-the-Art in Gesture Recognition and Automated Sign Language Translation

In theory, today's technological capabilities enable automated translation of sign languages. Various types of devices incorporate high standard hardware (e.g. cameras and microphones, accelerometers and gyroscopes) that can capture rich data from users. Computing resources and network connections are reasonable, also in mobile devices.

This enables the intensive processing of large quantities of data, even by cloud services. Interactive technologies have evolved and the basic technological components needed for further development of sign language translation technology already exist. These can be seen in the forms of video processing libraries and algorithms, an array of machine learning frameworks and architectures and to an extent, relevant corpora.

Technology facilitating gesture-based interfaces, relevant also to the automated translation of sign languages, can be split into two overall groupings: those relying primarily on wearables such as gloves, rings, accelerometers, etc., and those relying primarily on a camera or sensor-based tracking, where users' hand gestures are recorded at a distance [3]. In the first group, there are solutions such as the one presented by Cheng and colleagues [4], where an accelerometer is used to obtain 3D information of the hand movement with the aim of controlling an entertainment robot. Hand gesture recognition can also be used in rehabilitation processes by utilising resistive sensors that are placed on fingers in order to simulate the pressure distributions on the forearm, and reproduce the movement on an arm articulated model [5]. A continuous hand gesture recognition system, suitable for a low-cost and energy-efficient Human-Computer/Technology Interface, is presented by Gupta et al. [6]. The gesture characterization is made with the data from the accelerometer and gyroscope sensors. Moreover, a Dynamic Time Warping solution is used for gesture classification, and a gesture spotting process is used to distinguish between a database gesture and an involuntary gesture. Kiliboz and Gdkbay [7] utilise a magnetic 3D position tracker attached to the user's hand to collect trajectory data and then represent gestures as an ordered sequence of directional movements in 2D.

In the second group, different kinds of visual information can be used to solve the recognition problem. Visible imagery is employed by Molchanov and associates [8], whereby the hand localization was jointly performed with the hand gesture recognition task, using a multi-resolution sliding window that densely selected image regions, and evaluated each selection when a specific hand gesture was performed. Visible imagery is also used in Lee and Park's [9] study for developing a universal remote-control system using hand gestures, where the hands are segmented from the background using motion and colour skin clues. Other works use a Convolutional Neural Network (CNN) to recognize gestures. This is the case of the work presented by Bao et al. [10], which does not require any previous segmentation process. Similarly, another CNN is proposed in Liang et al. [11] for gesture recognition that also locates the hand. An alternative to visible imagery is near-infrared (NIR) imagery, which is more robust for adverse illumination conditions, favouring its deployment. In Mantecn et al. [12], a NIR-based hand gesture recognition system is proposed to interact within a complex immersive environment, which improves real-time performance of previous works, such as Mantecn et al. [13]. Lastly, 3D or depth information is also used to improve the gesture localization, and thus the general performance. In Cui et al. [14], a 3D CNN solution recognizes different dynamic hand gestures for sign language. In Mantecn et al. [15], hand crafted features based on a variation of Local Binary Pattern for depth images are used for operating Unmanned Aerial Vehicles with gestures.

Effective and functional sign language translation solutions based on the technology outlined above are few and far between. An example of one representing the first group

is the technology presented in Zhou et al. [16] which is basically a wearable tech-glove augmented with an add-on software capable of recognizing dozen types of simple hand configurations involved in finger alphabets and numerical American Sign Language signs. An example of the second group, on the other hand, is the SignAll¹ translation system that exploits AI and computer vision in order to recognize and translate American Sign Language in business and educational settings. Both solutions are campaigned and marketed as state-of-the-art in the field of sign language translation technology. However, although they undeniably include some cutting edge technology, they still are limited, for example, in their mobility (meaning that communicators need to remain in the same rigid positions), size of the corpus, and focus on conventional signs only – all factors affecting, in the end, their general usability and acceptance within the actual sign language users community.

3 Challenges of Automated Sign Language Translation

In reality, sign languages pose many challenges for automated translation. In the analogue sense, the possibility of translating over approximately 300 – some even say nearly 6000 – sign languages into various spoken languages is still highly debated among scholars (see [17], [18], [19]). The conversion of this task to information technology poses even more complications. Danielle Bragg and colleagues [20] highlight the dominance of studies that attempt to achieve this feat through working in knowledge silos. That is, more often than not, development teams represent experts of the same if not similar disciplines. This reduces the team’s capacity for understanding various nuances, interactions and interrelations that can only be seen through diverse lenses, i.e. multidisciplinary. Thus, the current state-of-the-art gesture recognition used for sign language recognition does not thoroughly take the complexity of sign language communication and particularly its translation into account. Nor, does the current state-of-the-art account for the complexity and multidimensionality of the socio-cultural context in question, and its role in conveying meaning. Current solutions focus on algorithms for detecting fingers, hands, and human gestures. The computer recognition of units comprising these elements occurs in relative isolation – without consideration for the role of other factors (bodily, gestural, contextual, environmental, interpersonal etc.) in meaning conveyance and subsequently, meaning-making.

3.1 Datasets – Corpora

From the technological perspective, and according to Bragg et al. [20], some of the main current challenges in automated sign language translation can be classified as pertaining to: datasets; recognition and computer vision; and modeling and natural language processing. From the point-of-view of datasets, the problems are as follows: a) the *size* of the datasets are limited, affecting generalizability in machine learning; b) *signer variety* – the contribution of novices and non-native signers means decreased

¹ <http://www.signall.us>

quality and accuracy, and to increase validity and recognition systems should be trained with real (native or near native) signers in realistic scenarios and datasets should include representative, generalizable samples from diverse age groups, gender, culture, various ethnicities, varying body types and physical traits, clothes, lighting conditions and more; and c) *continuous signing* – most existing datasets contain individual signs, which as a starting point can aid in the development of, for example, sign language dictionaries, it is not sufficient for recognising and interpreting real-world conversations that involve sentences and long utterances.

Many of the problems associated with datasets could be overcome by the use of sign language corpora in the development of automated sign language translation technology. Sign language corpora are large collections of systematized and annotated video materials, created for linguistic purposes, representing different types of sign language use by a large variety of sign language users. However, to date only a few sign language corpora have been created and those that exist are still relatively small in size when compared to machine learning corpora used in the development of spoken and written language translation technology. For example, the recently published first version of Corpus FinSL² (see [21]) – a state of the art corpus on Finnish Sign Language from the linguistic perspective – “only” has 107784 sign tokens produced by “only” 21 Finnish Sign Language users. The plan is to update this corpus in the future. Yet, due to the manual annotation work required this endeavour is extremely time consuming. Moreover, even the most updated version of Corpus FinSL would still be small from the machine learning perspective as the total number of Finnish Sign Language signers recorded for the material is only 80. A similar situation is faced regarding other sign language corpora.

3.2 Recognition and Computer Vision – Depiction

In terms of recognition and computer vision, challenges include *depiction* – the visual representation or enactment of linguistic content (e.g. [22], [23]). In the investigation of corpora of Western sign languages, it has been estimated ([24], [25]) that approximately 30 percent of sign tokens represent depictive content that cannot be unambiguously translated even in the traditional sense (see Fig. 1). The percentage in itself reveals the situation. However, this is made even more confusing by an estimation that the group of 70 percent of non-depictive signs comprises only approximately 3000 lexeme types. If we consider all of this, from the point of view of the current Corpus FinSL data (see above) for instance, the end result is that we can perhaps fairly easily translate (in the traditional sense) approximately 75500 sign tokens with the help of 3000 lexeme types. Yet, more challenges will undoubtedly be encountered regarding the estimated 33500 sign tokens that cannot be reduced into lexeme types. This is due to the unconventionality – or context dependence – of their form and semantic value. Another issue that needs to be considered is that depictive meanings are expressed not only with manual signs (i.e. with hands) but also non-manually, with facial expressions, bodily postures and movements (e.g. [26]).

² <http://urn.fi/urn:nbn:fi:lb-2019012321>

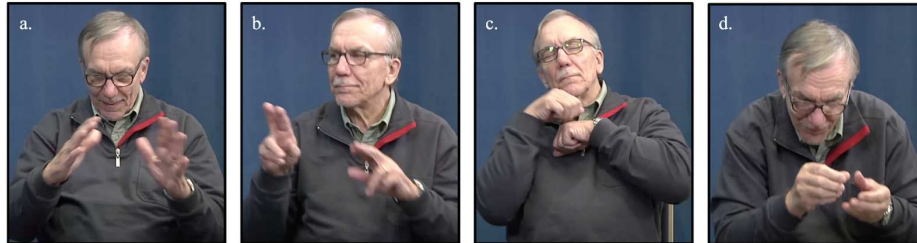


Fig. 1. Examples of depictive signs and enactments that pose problems to all types of translation.

Fig. 1 gives examples of four depicting expressions from Corpus FinSL. In (a.) the signer is depicting an *ad hoc* cylinder shape of an object which, depending on the context could be translated, for example, as a ‘snowman’ or a ‘pile of rocks’. In (b.) the signer is producing an actual phrase ‘boy and dog turn and look around’. However, in another context this utterance could refer to some other characters and targets. In (c.) and (d.), the signer is using his whole body in order to enact the activity of a character in the story he is representing. The translation of these utterances depend on the context as well as on the (human) translators own world knowledge.

To comprehend depiction in sign languages, it is also necessary to draw on the knowledge of deaf culture. Thus, genuine involvement with deaf communities is paramount in order to ascertain semantic value in conjunction with relevant syntactic devices, interactions and relations. It is not sufficient to base sign recognition on speech recognition as this cannot deal with depictions that are uncommon or insignificant in speech. The same concept may have numerous depictions (see Fig. 1), yet annotation systems do not have standards and conventions to account for this. For example, in most sign language corpora (including Corpus FinSL), depicting units have been annotated only with macro-level class notions such as a Size and Shape Specifier (Fig. 1a) or a Handling signs (Fig. 1d). This has been undertaken without any reference to the token level meanings of the units.

The lack of valid and reliable, large-scale *annotations* creates obstacles for applying machine translation and natural language processing (NLP) techniques to sign languages. Annotations in the form of text are generally applied as input. However, this in itself skews the authenticity of sign language expression as often sign languages do not have a standard written form. On the other hand, using more technical transcription or notation methods such as HamNoSys or SignWriting is too time consuming. The consequence of all this is that datasets are and will remain limited in terms of the number of samples and signers (not always reflecting the signing population).

Finally, from the perspective of recognition and computer vision, *generalization* is difficult in relation to unseen factors in communication contexts. These unseen elements (important in depictions) include individuals and situations that inform the meaning of what is being communicated.

3.3 On Modelling, NLP and Human-Computer/Technology Interaction

Modeling and NLP are difficult in the sign language translation technology sense for several reasons. Two of these reasons can be understood as: a) *structural complexity* – most MT and NLP methods have been established for verbal and written languages; and b) *annotations* – as with depiction, annotations in modeling and NLP cause problems due to the absence of reliable, valid and mass annotations. Furthermore, due to the incompatibility between sign language and standard written forms of language traditional approaches to translation such as sign or speech to text present syntactic and semantic inadequacies. Yet, in addition to these linguistic challenges, there are also highly specialised considerations that need to be given to the interaction design of the translation systems and devices.

Human Computer Interaction (HCI) is a constantly evolving area of design and scholarship that in its relatively short lifespan has revealed a myriad of issues, dimensions, challenges and enablers. What must be remembered from the perspective of sign language translation technology development is that the language level of design and logic possesses one layer of issues, while the HCI, or Human Technology Interaction (HTI), dimension introduces a compelling list of other matters. These matters relate equally to the technology itself, as they do to users and indeed designers, environments and contexts within which they are utilised and experienced. From the experiential (emotional and psychological) perspective we may see the technology as possessing its own significance – projecting values, including or excluding populations on the basis of socio-economic conditions, education, accessibility etc., and posing other obstacles such as technophobia and social resistance. In terms of usability, the above mentioned examples of current sign language translation technologies, may be cumbersome. They may also be limited in terms of their allowance for naturalistic movement, confusing to operate (interface with), and distracting in terms of human-to-human social interaction. Functionality to a great extent taps into the challenges that have also been listed above – limited corpus and non-inclusion of unconventional signs. For these reasons we consider co-design, or indeed co-engineering, with people from deaf cultures to be a critical component when considering effective and affective interaction design of sign language translation technologies.

One specific challenge for developing user interfaces for assistive purposes, such as sign language translation, concerns how they are integrated into existing tools. Separate interfaces or applications are often challenging to maintain and develop further if they are not integrated into existing tools and applications. Even though there are some success stories, such as how screen readers and speech synthesis technology in general has been integrated in both desktop and mobile user interfaces, most of the assistive technology solutions are still separate from the interfaces used by most of the users. In order to develop integrated interfaces compromises are often needed, which creates another challenge for end user requirements. Thus, finding a good balance between user requirements and other factors is often challenging.

3.4 Co-Engineering, Participation and Culture

While on the one hand there are current problems with the technological development on the linguistic and machine learning levels, much more consideration also must be given to the interaction design of the devices and their software. In fact, Korte, Potter and Nielsen [27] note the importance of including considerations for deaf culture, when attempting to engage in interaction design for people belonging to deaf culture. In their study with Deaf Australians, Korte et al. found that deaf, as with any culture, maintains its own conventions, customs, etiquette and linguistic differences. As with any other group of target users, particular cultural considerations need to be made a priority, as this subsequently affects not simply the willingness to adopt, engage with and sustain use of a device or application (e.g. [28], [29]), but also the integrity of the system and its relevance [30]. Understanding culture is not simply about understanding norms, but also understanding values [31]. When considering design for accessibility for instance, it is wrong to approach the development work and user involvement in terms of working with people with ‘special needs’. For, within the deaf culture, those with special needs are those who possess a health problem or condition other than that of no hearing [27].

Customs and cultural cues affect the interaction design on a very practical level. For instance, in the case of Australian Deaf, a light touch on the arm is used to gain attention. When exceeding arms reach, a wave is also acceptable. Each interacting actor must wait until both have visual attention. If someone’s attention is diverted (gaze moves elsewhere), then the signing individual should stop and wait until their interaction partner renews gaze on the signing [27]. Thus, while we understand that the ability for technology to detect, collect and translate continuous signing is ideal, there is also the need for understanding that continuous signing does not mean non-stop signing. Rather, it means communication in rhythms that are dependent on attention, attention seeking and waiting for social, intentional and visual cues. In other words, for instance, while the computer vision should capture all the meaningful communication gestures, it should not capture *everything*.

The nature of these conventions, and indeed unfamiliarity to hearing communicators, means that the best communicators of sign, and particularly of sign with deaf children (as in Korte et al.’s study case [27]), are native deaf adults. This is an important implication for the interaction design of sign language translation technology, as the designers themselves (or members of the design and development team) should preferably not only be fluent in sign, but also be members of the deaf cultural community itself. Native deaf adults are skilled at attracting and maintaining attention, particularly of other deaf individuals. These types of skills are highly valuable in the context of interaction design, as they afford cues for understanding how the logic of the interface and communication system should work. Furthermore, it should also be noted that deaf individuals and those who live with them (i.e., parents) have a heightened awareness of non-verbal communication that potentially may provide worthy insight in relation to understanding the unconventional gestures and signs mentioned above.

4 The Need for Automated Sign Language Translation?

Despite the affordances provided by existing technology (Section 2), the challenges in developing a robust and cost-effective sign language translation tool are still great (Section 3). In addition, we also need to ask ourselves one more important question: How needed is such tool in the first place?

Let us imagine an international conference on Uralic and Baltic languages, organised by deaf linguists for an audience composed of deaf scientists and deaf laymen. Hearing researchers in their conferences would rely on English, German or Russian or another language shared by the colleagues of their discipline. Skilled interpreters would convey aloud the contents of keynotes and papers to the local audience. But at the “deaf-only conferences”, interpreters are needed solely for non-signers, i.e. “hearing” participants. No (electronic, mobile) dictionaries would otherwise be consulted during vivid corridor discussions or opening and closing events. Even if all attendees would come from different parts of the world, no electronic technology would be utilised. For a hearing linguist, this is an inconceivable scene. Regardless of their doubts, this linguistic miracle is materialized every time deaf people who do not share each other’s native signed languages meet. Deaf people find a way of being able to communicate.

In research literature, the aforementioned linguistic behaviour is called International Sign (IS) or cross-signing (see [32], [33]). It emerges from “shared deaf cognition”, visual conceptualizations of the world, created by deaf people whose language is visual and gestural. It is full of elements containing visual cues, gestures, pointing and acting out in the tridimensional space around the signer. When abstract or concrete things are not present or visible, deaf signers animate and materialize them with their hands and body in this very space, be they mathematical or physical, cosmological or microbiological entities or processes (see [34], [35]).

It is widely assumed that the existing conventional sign languages have their origins in similar moments of interaction between deaf people. Due to their fairly long developmental history, most European sign languages (SL) nowadays are well-established independent languages (Swedish SL, British SL, French SL, Finnish SL, Estonian SL etc.). They contain a nationally agreed conventional lexicon, i.e., precise signs, both motivated and abstract, with well-defined semantic meanings (see Section 3). But as all signed languages are produced by visual articulators of the human body (i.e. hands, torso, head, mouth along with facial expressions), they are also able to convey more fluid, unfixated meanings, using iconic, pantomime-like segments, that have nevertheless, properties with a more or less determined meaning. It is these depicting elements that are constantly exploited by all established sign languages and all sign language users when, for instance, the local established lexicon is insufficient or non-existent in a determined context. For centuries, deaf people have also made use of the aforementioned pantomimic gestures in contexts or locations where a signed language has not been able to emerge or when people born deaf have not been able to acquire a conventional sign language. In these contexts the signed expressions are called “home signs”.

The bi- and multilingualism of deaf sign language users with some spoken or written language should not be forgotten either. This use of majority language alongside sign

languages means, at least in modern Western cultures, that very often *ad hoc* communication between signers and non-signers could be organised simply with writing some shared language to paper or to a digital device such as a smartphone. There is no real need for an automated translation solution in this sense. For more complex communicative situations, and for instance circumstances characterised by illiteracy as an example, there is always the traditional interpreting service with human interpreters. As far as we can see, this is likely to remain the best way to translate between sign languages and spoken languages in many everyday contexts (doctors appointments, schools etc) for years to come.

But developing automated translation technology for sign languages is not insignificant either. Reasons for developing such technology can be classified into both concrete and ideological ones. An example of a very concrete motivation might be that which was set up by the EU web accessibility directive, related guidelines (WCAG) and national legislations that requires all web content to be accessible to all EU citizens. For spoken language users there already exists several speech to text translators which can produce either closed captioning or subtitling, for example, to videos including speech. This is not the case with sign language videos, which obviously creates a gap in accessibility to people not familiar with sign languages.

On an ideological level, the desire to develop sign language translation technology can be argued to be of value in itself. Again, to use an analogy from the spoken language research, technology development has undeniably contributed to the deeper understanding of spoken languages and this dimension of progress should not be denied from sign languages. By understanding sign languages better in terms of translation technology we will eventually gain a better understanding of human language in general. It is also possible to argue that by developing automated translation technology for sign languages we will also eventually leverage the status of sign languages. Sign languages are minority languages, even endangered ones, and paying attention to them in this way is likely to result in positive consequences on socio-political level at some point.

5 Conclusion

Delivering a universal tool for sign language translation that is robust and cost-effective, and that would operate between a large number of spoken languages and sign languages is still not realistic. It can also be asked if such a tool, or a smaller scale version, is even needed. However, we believe contributing to the improvement of sign language detection, processing and even translation to spoken languages in the future is something that should not be abandoned. However, the tasks in this should be assigned realistically, and should focus on all necessary aspects of sign languages and sign language users' community, not only those we are accustomed to in the field of spoken language research and speech community.

A feature we want to emphasize in particular is unconventionality. Full conventionality is not something that characterizes languages in general. Instead, all languages include utterances which do not belong to conventional types but are rather unconventional manifestations that exist on a token level. These unconventional utterances are

used for depictive purposes and can thus refer to a variety of *ad hoc* meanings. It is only by taking non-conventional, depictive units into account that the development of automated sign language translation technologies can truly progress. Moreover, we believe that by focusing on these unconventional units, sign language translation technology development can eventually give something back to spoken language research as well.

In the end, the development of automated sign language translation technology may require us to redefine what is considered as translation. In order for any translation to be feasible, the unconventional aspect of language needs to be taken into account. Eventually, this may mean for example, that manual, gestural and vocal depictions (depictive signs and gestures, and enactment) are accepted to be converted also into visualizations and images – and vice versa.

Acknowledgements

This article stems from an interdisciplinary collaboration and discussions between numerous incredible people. In addition to an extended team representing the authors' institutions (Sign Language Centre and Faculty of Information Technology, University of Jyväskylä; School of Interpreting and Linguistic Access, Humak University of Applied Sciences; and TAUCHI, Tampere University), the authors wish to acknowledge the crucial role of the representatives from the following institutions: Europa University Viadrina, Finnish Association of the Deaf, Kosovar Association of the Deaf, and Fundación CNSE Para La Supresión De Barreras De Comunicación.

References

1. Rueter, J., Hämäläinen, M.: Prerequisites For Shallow-Transfer Machine Translation Of Mordvin Languages: Language Documentation With A Purpose. In *Материалы Международного образовательного салона*, pp. 18-29. Ижевск: Институт компьютерных исследований (2020).
2. Fischer, S.: Sign languages in their historical context. In: C. Bowern & B. Evans (Eds.), *The Routledge Handbook of Historical Linguistics*, pp. 442-465. Abingdon: Routledge (2015).
3. Vuletic, T., Duffy, A., Hay, L., McTeague, C., Campbell, G., Greal, M.: Systematic literature review of hand gestures used in human computer interaction interfaces. *International Journal of Human-Computer Studies* 129, 74-94 (2019).
4. Cheng, J., Bian, W., Tao, D.: Locally regularized sliced inverse regression based 3D hand gesture recognition on a dance robot. *Information Sciences* 221, 274-283 (2013).
5. Nan, L., Shourong, W., Minfeng, W., Bo, L., Hong, H., Li, J.: Hand motion recognition based on pressure distribution maps and LS-SVM. In: *International Conference on Mechatronics and Control (ICMC)*, pp. 1027-1031 (2014, July).
6. Gupta, H. P., Chudgar, H. S., Mukherjee, S., Dutta, T., Sharma, K.: A Continuous Hand Gestures Recognition Technique for Human-Machine Interaction Using Accelerometer and Gyroscope Sensors. *IEEE Sensors Journal* 16(16), 6425-6432 (2016).
7. Kiliboz, N. Ç., Güdükbay, U.: A hand gesture recognition technique for human-computer interaction. *Journal of Visual Communication and Image Representation* 28, 97-104 (2015).
8. Molchanov, P., Gupta, S., Kim, K., Kautz, J.: Hand gesture recognition with 3D convolutional neural networks. In: *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1-7, Boston, MA (2015).

9. Lee, D., Park, Y.: Vision-based remote control system by motion detection and open finger counting. *IEEE Transactions on Consumer Electronics* 55(4), 2308-2313 (2009).
10. Bao, P., Maqueda, A. I., del-Blanco, C. R., García, N.: Tiny hand gesture recognition without localization via a deep convolutional network. *IEEE Transactions on Consumer Electronics* 63(3), 251-257 (2017).
11. Liang, Z., Liao, S., Hu, B.: 3D Convolutional Neural Networks for Dynamic Sign Language Recognition. *The Computer Journal* 61(11), 1724-1736 (2018).
12. Mantecón, T., del-Blanco, C. R., Jaureguizar, F., N. García, N.: A real-time gesture recognition system using near-infrared imagery. *PLOS ONE* 14, 1-17 (2019).
13. Mantecón, T., del Blanco, C.R., Jaureguizar, F., García, N.: Hand Gesture Recognition using Infrared Imagery Provided by Leap Motion Controller. In: *International Conference on Advanced Concepts for Intelligent Vision Systems, ACIVS 2016*, pp. 47-57. Lecce, Italy, (2016, October 24-27).
14. Cui, R., Liu, H., Zhang, C.: A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training. *IEEE Transactions on Multimedia* 21(7), 1880-1891 (2019).
15. Mantecón, T., del Blanco, C.R., Jaureguizar, F., García, N.: New generation of human machine interfaces for controlling UAV through depth-based gesture recognition. In: *SPIE International Conference on Unmanned Systems Technology XVI*, SPIE 9084, pp. 90840C-1-11. Baltimore, MA, USA (2014, May 5-9).
16. Zhou, Z., Chen, K., Li, X. et al.: Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays. *Nature Electronics* 3, 571-578 (2020).
17. Metzger, M.: *Sign language interpreting: Deconstructing the myth of neutrality*. Washington, DC: Gallaudet University Press (1999).
18. McDermott, M. A. N., Palchanes, K.: A literature review of the critical elements in translation theory. *Image: The Journal of Nursing Scholarship* 26(2), 113-118 (1994).
19. Veale, T., Conway, A., Collins, B.: The challenges of cross-modal translation: English-to-Sign-Language translation in the Zardoz system. *Machine Translation* 13(1), 81-106 (1998).
20. Bragg, D., Koller, O., Bellard, M. et al.: Sign language recognition, generation, and translation: An interdisciplinary perspective. In: *ASSETS 2019 – 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 16-31 (2019).
21. Salonen, J., Kronqvist, A. & Jantunen, T.: The Corpus of Finnish Sign Language. In: Efthimiou, E., F. Stavroula-Evita, T. Hanke, J. Hochgesang, J. Kristoffersen & M. Mesch (eds.) *Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pp. 197-202. European Language Resources Association, Paris (2020).
22. Jantunen, T.: Constructed action, the clause and the nature of syntax in Finnish Sign Language. *Open Linguistics* 3, 65-85 (2017).
23. Ferrara, L., Hodge, G.: Language as Description, Indication, and Depiction. *Front. Psychol.* 9, 716 (2018).
24. Johnston, T.: Lexical Frequency in Sign Languages. *Journal of Deaf Studies and Deaf Education* 17(2), 163-193 (2012).
25. Fenlon, J., Schembri, A., Rentelis, R., Vinson, D., Cormier, K.: Using conversational data to determine lexical frequency in British Sign Language: The influence of text type. *Lingua* 143, 187-202 (2014).
26. Puupponen, A.: Towards understanding nonmanuality: A semiotic treatment of signers' head movements. *Glossa: a journal of general linguistics* 4(1), 1-39 (2019).

27. Korte, J., Potter, L. E., Nielsen, S.: The impacts of Deaf culture on designing with Deaf children. In: Proceedings of the 29th Australian Conference on Computer-Human Interaction, pp. 135-142 (2017, November).
28. Carroll, J. M.: Human-computer interaction: psychology as a science of design. *Annual review of psychology*, 48(1), 61-83 (1997).
29. Pacey, A.: *The culture of technology*. MIT Press (1983).
30. Clemmensen, T.: Whatever happened to the psychology of human-computer interaction?: A biography of the life of a psychological framework within a HCI journal. *Information Technology & People* 19(2), 121-151 (2006).
31. Cockton, G.: Value-centred HCI. In: Proceedings of the third Nordic conference on Human-computer interaction, pp. 149-160 (2004, October).
32. Crasborn, O. A., Hiddinga, A.: The paradox of international sign. The Importance of deaf-hearing encounters for deaf-deaf communication across sign language borders. In M. Friedner & A. Kusters, (Eds.), *It's a small world. International deaf spaces and encounters*, pp. 59-69. Washington, DC: Gallaudet University Press (2015).
33. Hiddinga, A., Crasborn, O.: Signed languages and globalization. *Language in Society* 40(4), 483–505 (2011).
34. Rainò, P., Ahonen, O., Halkosaari, L.: The deaf way of interpreting mathematical concepts. In: C. Stone (Ed.), *Deaf Interpreting in Europe. Exploring best practice in the field*, pp. 10-20. Copenhagen: Danish Deaf Association (2018).
35. Rainò, P., Ahonen, O.: From zero to sign. Transferring food safety from Finnish to Finnish Sign Language. In: C. Stone (Ed.) *Deaf Interpreting in Europe. Exploring best practice in the field*, pp. 21–33. Copenhagen: Danish Deaf Association (2018).