

User Involvement in the Research and Development Life Cycle of Sign Language Machine Translation Systems

Lisa Lepp and Dimitar Shterionov and Mirella De Sisto

Department of Cognitive Science and Artificial Intelligence,
Tilburg University, 5037 AB Tilburg, The Netherlands

Abstract

Machine translation (MT) has evolved rapidly over the last 70 years, thanks to the advances in processing technology, methodologies, and the ever-increasing volumes of data. This trend is observed in the context of MT for spoken languages. However, when it comes to sign language (SL) translation technologies, the progress is much slower; sign language machine translation (SLMT) is still in its infancy with limited applications. One of the main factors for this set back is the lack of effective, respectful and fair user involvement across the different phases of the research and development of SLMT.

We present a meta-review of 111 articles on SLMT from the perspective of user involvement. Our analysis investigates which users are involved, and what tasks they assume in the first four phrases of MT research: (i) Problem and definition, (ii) Dataset construction, (iii) Model Design and Training, (iv) Model Validation and Evaluation. We find out that users have primarily been involved as data creators and monitors as well as evaluators. We assess that effective co-creation, as defined in (Lepp et al., 2025), has not been performed and conclude with recommendations for improving the MT research and development landscape from a co-creative perspective.

1 Introduction, Motivation and Related Work

Machine translation (MT) has evolved rapidly over the last 70 years. The first MT systems, i.e., rule-based MT, built around human-crafted rules and dictionaries, followed a very human-intensive process. With the shift towards data-driven MT, the MT development process became structured

around the collection and processing of large volumes of data with the use of powerful computational tools. This process was distributed over distinct human-intensive (e.g. data collection) as well as computationally-heavy tasks (e.g. training a word-alignment with a tool such as giza++ (Och and Ney, 2003) or training an encoder-decoder neural network (Bahdanau et al., 2015)), aiming to reduce human efforts in quickly delivering effective and efficient MT systems. Along the way, it aligned with the generic machine learning (ML) and deep learning (DL) practices, and as such can be divided into six key phases: (i) Problem and use-case definition and solution ideation, (ii) Dataset construction, (iii) Model development, (iv) Quality assessment (automatic and / or human), (v) Model deployment and (vi) Monitoring and maintenance.¹ In the context of MT, humans with different expertise are involved in these stages, e.g. native speakers generate new data; native speakers and professional translators evaluate MT output; linguists participate in the data processing and preparation; engineers and computer scientists develop model architectures and train models.

MT primarily addresses text-to-text, text-to-speech, and speech-to-text use-cases, which pertain to Spoken Languages (SpLs), where substantial progress and qualities matching human standards are now observed. When it comes to user involvement in MT projects, users may take part in the data collection and in the evaluation stages, but are rarely involved in the other stages.

Translation technologies for SLs, however, have not progressed as quickly as SpL MT. Challenges related to data, modeling and the complexity of processing are significant contributors to this slow

¹These phases or stages may vary depending on the granularity or the grouping of sub stages. E.g., another 5-phase formulation is: (i) Problem definition, (ii) Dataset collection and processing, (iii) Model Design and Training, (iv) Model Validation and Evaluation and (v) Model deployment and maintenance.

progress. An equally important consideration is the role of humans. In SLMT and Natural Sign Language Processing (NSLP), SL data is typically collected in the form of video recordings of signing individuals. The typical SLMT process still involves distinct transformation phases (Shterionov et al., 2024), which require human intervention. In these fields, humans are crucial not just as data creators, evaluators, or monitors, but also as active partners in developing practical and socially impactful SLMT systems.

Caselli et al. (2021) acknowledge the need to better include users in the research process and advocate for a more user-involving natural language processing (NLP) research. They wrote 9 guidelines for participatory design in NLP (we summarize these in Appendix B). Harder et al. (2013) analyses user involvement in different fields and propose a participation typology which ranges over various degrees of user involvement. However, in the context of SpL MT and NLP research there are no clear indications for the increase in user-involvement over the whole life-cycle. As advocated by (Caselli et al., 2021), engaging user communities in NLP projects is essential. However, in this field, the term *co-creation* typically refers to human-AI collaboration for content generation, as explored in recent studies (Sharma et al., 2024; Konen et al., 2024; Ding et al., 2023) on optimizing interactions between large language models (LLMs) and human creators. Co-creative methods have also been applied to tasks like poetry generation (Gonçalo Oliveira et al., 2017), literature synthesis (Manjavacas et al., 2017), and interpreting (Nakaguchi et al., 2016), and there no significant work demonstrating effective user involvement for MT.² While this may not be significantly problematic for the adoption of translation technologies for SpLs at present,³ it has a negative impact on SLMT research progress and adoption due to issues related to lack of expectation management through unacceptable and impractical outputs (e.g. SL translation gloves) to unethical research (e.g. involving non-signers in the data creation process). The recent work of Lepp et al. (2025) proposes a formal definition of co-creation for SLMT accord-

ing to which users contribute as equal partners⁴ in the SLMT project as well as a participatory typology to aid the assessment of SL user and SL community (SLC) members involvement in such projects. We must emphasize the following points. First, not every SL user is inherently part of a SLC. The term "SLC" may suggest that it encompasses all variations in SL fluency, equal access to a visual language, and educational opportunities, but this is not the case, as highlighted by the American National Association of the Deaf⁵. Therefore, we draw a clear distinction between SLCs and SL users, ensuring that SL users outside of an SLC — for example, Hard-of-Hearing (HoH) signing individuals⁶—are also acknowledged, along with individuals who are part of an SLC.⁷

Reflecting on co-creative and user-involving practices, they analyses 111 articles to identify the degree of involvement of SL users and SLC members. They use an adaptation of (Harder et al., 2013)'s typology for the specific case of SLMT and NSLP,⁸. Their analysis provides a generic overview of these articles. However, the necessity and feasibility of co-creation may differ significantly across the phases of the ML life cycle (or MT life cycle). For instance, dataset construction is an area where user involvement is often critical, while model design and training may present practical challenges in integrating co-creation effectively.

We take the work of (Lepp et al., 2025) one step further and conduct a deeper review of the 111 articles on SLMT based on (1) the involvement of the SL user per *research and development phase*, and (2) the roles of the involved SL user and / or

²An ACL Anthology search from Oct 7, 2024 found 146 relevant works overall.

³Current language technologies have evolved and spread to an extent that they have become indispensable part of professional and non-professional translation activities.

⁴With equal partners, we propose involving SL users and the SLC as essential collaborators during the MT phases.

⁵<https://www.nad.org/resources/american-sign-language/community-and-culture-frequently-asked-questions/>

⁶We drew the distinction between Hard of Hearing (HoH) and Deaf from the literature review in 2; however, this does not imply that HoH individuals are inherently less fluent in sign language than Deaf individuals. Fluency and authenticity in sign language are not determined by medical hearing status, but rather by individual preferences and choices in language use.

⁷Similar to SpLs, SLs are dynamic languages with dialects and regional variations, particularly in vocabulary. Furthermore, there are home-signs, family-signs, village-signs, and individual signs, among others. Most existing SLMT models, methods, and databases rely on standardized SL, which makes it challenging to accommodate these variations. Since we advocate for co-creation to leverage diverse perspectives, it is essential to consider these inter-signer variations when developing SLMT models, methods, and databases.

⁸We summarize their typology in Appendix A.

the kind of functions or task they have fulfilled (if can be derived from the reviewed article).

The granular assessment of user involvement we provide in this work, leads to better insights into where participation is needed and what efforts should be focused on better involving SL users and SLC members. We identify where and to what extent projects have been co-created and where there is room for improvement. Doing so, we contribute to the literature gap of to what extent the SLC has been involved throughout the different phases in the research life cycle of a technical project.

This paper is structured as follows: in Section 2 we provide our meta-review of the 111 articles referenced in (Núñez-Marcos et al., 2023; Coster et al., 2024). Next, we present with the assessment of user involvement per MT phase in Section 2.2. Section 3 presents a discussion and a critical reflection follows in Section 4. We conclude in Section 5 with general remarks and directions for future work.

2 Meta-review of user involvement and co-creation

In 2023 and 2024, two review articles of SLMT were published by Coster et al. (2024) and Núñez-Marcos et al. (2023). These articles contain an overview of a large volume of literature on SLMT, focusing on the technological solutions, different approaches and historical advancements. To the best of our knowledge, these are the most complete and recent historical overviews of work in the field of SLMT and NSLP. We conducted a meta-review, i.e., a manual analysis, of the SLMT-related papers reviewed in (Coster et al., 2024) and (Núñez-Marcos et al., 2023) from the perspective of (1) SL users involvement per ML research phase and (2) the roles that the SL users had in these phases. We rely on their work for our meta-review because in addition to their recency as literature review works our insights can directly complement their findings. That is, we believe our work fills in the societal gap of these works. We do acknowledge the fact that our work does not involve the articles beyond 2023 and we leave this for future work where a new study should look into both technological (theirs) as well societal (ours) aspect of SLMT and NSLP work.

2.1 Selection and filtering criteria

To align our analysis with the work of (Lepp et al., 2025), we follow their selection criteria. These are:

- The article needs to be mentioned in (Coster et al., 2024) or (Núñez-Marcos et al., 2023);
- The article needs to have open access;⁹
- The study should focus on SLMT, or on one of the phases;
- The study should focus on translation of SLs or between SLs and SpLs in either direction (but not only on SpLs);

After these exclusion steps, the remaining **111 articles**, (56,9 % of the original 193) were considered in the following discussion. For completeness, we list these 111 papers in Appendix C.¹⁰

2.2 User involvement in SLMT-research

To gain insights on the user involvement in the SLMT research and development projects covered by the aforementioned articles, we decompose these projects according to the typical machine learning (ML) phases most-commonly adopted in MT (noted in Section 1) – (i) Problem and use-case definition and solution ideation, (ii) Dataset construction, (iii) Model development, (iv) Quality assessment (automatic and / or human), (v) Model deployment and (vi) Monitoring and maintenance – and consider the first four phases. We did not look beyond the phase of quality assessment, i.e. Phase (iv), as the reviewed articles do not cover phases (v) and (vi). We categorize these papers according to the phases of the MT research life cycle to assess the extent of user involvement as a proxy to co-creation implementation and identify areas for improvement.

We also analyzed the kind of roles the SL user may have had, and —if it is clear— what kind of tasks they worked on during the different MT research life cycle phases.

Table 1 presents a summary of the overall distribution of articles—both with and without user involvement classified across the four different phases of the research life cycle in an MT project.

The results in Table 1 shows that the user has been involved in 11% of the reviewed articles across one of the research phases: less than 1% of the articles follow co-creation practices in phases

⁹As one of the reviewers rightly mentioned, some works were excluded that might have been more co-creation based. However, since we decided to align with (Lepp et al., 2025), we applied the same selection criteria.

¹⁰These are the same articles that are listed in the work of Lepp et al. (2025) but for completeness we add them to this paper.

(i) and (iii); 5% – in phase (ii) and 6.9% in phase (iv).

2.3 Roles, tasks and functions

As outlined in (Lepp et al., 2025), (Harder et al., 2013), and (Caselli et al., 2021), co-creation involves a diverse group of actors (users, researchers, etc.) with different roles. We further analyze the articles within the 11% in the “With user involvement” column in Table 1, which follow some form of co-creation. We examined who the user was carried out, the tasks assigned to the SL users (if specified), and in which phases they were involved. This data is presented in Table 2.

The roles and user types listed in Table 2 are presented as they appear in the reviewed articles. Although some of these roles might seem to fit together in one group, we maintain them separately due to the additional information or uncertainty they carry. For instance, “Deaf and Hearing” may or may not include team members, whereas “Deaf and Hearing team members” explicitly indicates that the individuals are part of the development team (as reflected in their assigned roles). Based on this, we offer the following general observations:

1. User types span over 24 different user type-role combinations. We distinguish 8 user types: Hearing, Deaf –across different regions–, HoH, CoDa¹¹, as well as Experts¹², Interpreters, Linguists, and Teachers (with or without an indication whether these are hearing, deaf or HoH individuals); and 4 tasks (which determine the role these individuals “play”): data recording, data annotation, data collection¹³ and participating as a member of the research and development team. We recognize that row 6 (e.g., deaf experts in Table 2) is marked as 0, 0, 0, 0 for the analyzed phases. However, as noted by Coster et al. (2019), Desai et al. (2024), and Marshall and Sáfár (2002) that discusses the importance of ‘co-creation with the DHH community’ and ‘the

¹¹i.e. Children of Deaf Adults

¹²For the definition of *Expert*, we adopt the following description: an individual who possesses a comprehensive and profound understanding, along with competence in knowledge, skills, and experience, acquired through practice and education in a specific field or area of study https://en.wikipedia.org/wiki/Expert?utm_source=chatgpt.com

¹³We distinguish between data recording and data collection with the former involving the user in the recording of SL data, while the latter may imply that the user is tasked to collect existing (already recorded) data.

value of feedback and guidance from Deaf users’, we have categorized this subgroup as ‘Deaf experts’.

2. When comparing these roles with the advanced typology of relationships (see Appendix A from (Lepp et al., 2025)), we observe that there is a clear gap of engagement between researchers and SL users and/or SLCs. This aligns with Level -1 (Denigration –direct or indirect impact), Level 0 (Neglect) or Level 1 (Learning From). In the article of López-Ludeña et al. (2012), there is a combination of Deaf and Hearing participants. In another paper, that of Ebling and Huenerfauth (2015), both Deaf and Hearing team members were involved, but despite the promising nature of their involvement, it was limited to the data collection phase. We can state, based on these findings, that the researcher holds the *power*, particularly in phases (i) and (iii), and is the *sole decision-maker* throughout *all* phases of the MT project.
3. When comparing Table 2 with the guidelines from (Caselli et al., 2021), we observe that these distinct roles do not align with several of the guidelines: Principle 1 – there is no discussion leading to consensus; Principle 2 – the process is not reflexive but limited to one or two phrases; Principle 3 and Principle 4 – SL users are predetermined and treated as data; Principle 5 – wider communities are not involved, and Principle 7 – language has been seen as an end, rather than a means. Additionally, Principle 8 (consent versus intrusion) and Principle 9 (considering the dynamics) are debatable, as there is a lack of meta-data regarding the appropriateness of the involvement or the application for grants in these MT processes.

3 Discussion of user involvement per phase

We delve into these articles further and hints to several tendencies of positive and negative practices broken down per phase.

1. **Phase (i): Problem and use-case definition, and solution ideation**

Research phases	No user involvement	User involvement
(i)Problem and definition	110	1
(ii)Data Construction	89	22
(iii)Model Design Training	110	1
(iv)Model Validation and Evaluation	94	17
Total	403	41

Table 1: The amount of reviewed articles per MT research phase over two categories: with or without user involvement.

Actor category and task	Phases				Total
	(i)	(ii)	(iii)	(iv)	
1. Hearing only	0	1	0	1	2
2. HoH only	0	1	0	0	1
3. Deaf only	0	2	1	5	8
4. CoDa only	0	0	0	0	0
5. Deaf and hearing	0	0	0	2	2
6. Deaf experts	0	0	0	0	0
7. Hearing and Deaf experts	0	1	0	0	1
8. Deaf signers across different regions	0	1	0	0	1
9. Deaf for data recordings	0	2	0	0	2
10. Deaf for data annotations	0	1	0	0	1
11. Deaf for data recordings and annotations	0	1	0	0	1
12. Deaf for data-collection	0	1	0	0	1
13. Deaf via Video-channel	0	1	0	0	1
14. Interpreters	0	1	0	1	2
15. Interpreters and CoDa’s	0	1	0	0	1
16. Interpreters and Deaf	0	1	0	1	2
17. Deaf and Hearing team-members	0	1	0	0	1
18. Linguists	0	0	0	2	2
19. Teachers	0	1	0	1	2
20. Expertise unclear	1	0	0	2	3
21. Linguists and teachers	0	1	0	0	1
22. Teachers + Deaf across different regions	0	2	0	0	2
23. Deaf experts and interpreters	0	1	0	0	1
24. Deaf and CoDa for video recordings	0	1	0	0	1
Total:	1	22	1	17	41

Table 2: The amount of articles which include co-creative practices and the roles of participants they mention, categorized over the four research phases of MT.

There is only one article –i.e. [Morrissey and Way \(2007\)](#), that notes involvement in phase (i). This observation contradicts Principles 1, 2, 6 and 9 of ([Caselli et al., 2021](#)). That is, in an effective co-creative project, users and developers should be in agreement early on of its development (Principle 1); as recommended in Principles 2 and 6, the design, which is encapsulated in phase (i) as well as in (iii), should be a continuous process; Prin-

ciple 9 suggests the involvement of the users and the community at a stage where goals are discussed and decided, which, typically, takes place in phase (i) of an SLMT project.

2. **Phase (ii): Dataset construction** We observe that in the current MT landscape, as shown in [Table 2](#), the SL user is primarily involved in data recording, data collection, or annotation tasks. This observation conflicts with

recommendations of the most relevant Principles of (Caselli et al., 2021) – Principle 7 (Language is a means rather than an end) and Principle 4 (Data and communities are not separate). Following the typology of Lepp et al. (2025), work that only includes SL users in data collection and processing tasks would be classified as Level 0 or Level 1. Furthermore, despite involvement of SL users in this phase (the largest number among all phases), there are considerations that need to be taken into account. These include what data will be created, who will create it, how many signers are involved, and whether they are representative of the population from which the data is gathered. For example, the work of Vandeghinste et al. (2024); Sisto et al. (2022); De Meulder (2021) differentiate between SL data as a source (original data) and SL data as a target (translated from SpL data), with the latter case potentially leading to MT producing less natural translations; their research also underlines that non-native signers produce SL data that is impacted by their first language. Another issue relate to the collection method and technical setup for best human-computer interaction. For example, Jedlička et al. (2020) note how certain aspects of motion capturing (MoCap) environments can lead to user discomfort and propose a lightweight marker setup, at the expense of a large number of cameras.

3. (iii) Model development

As shown in Table 2, only one article described the involvement of deaf users in phase (iii). While practical challenges related to co-creation arise in this MT phase, the solution may lay within Principle 2 (the concept of a continuous, reflexive, and ongoing design process) and typological levels 4a, 4b, and 4c (such as the exchange of knowledge encompassing a wide range of expertise, and expanding the community as the project progresses), as well as Principle 9 (the complex dynamics of funding, formulating research goals, and community involvement) from Caselli et al. (2021). Through continuous dialogue about the needs of the SL user, the requirements of the model, and the technical possibilities, a consensus can be reached. This approach helps address the challenge of implementing co-creation in MT phase (iii), –the Model De-

velopment phase–, but also the communicative aspect in between SL user and/or the SLC, and academics. We ought to point out that the typical MT/ML model development requires the efforts of an expert – someone who is familiar with using computational tools and methods for the design, development and (hyperparameter) optimization of such models. The role of the *expert* and the *user* are distinct and it is therefore difficult to integrate the user in this phase. However, the work of (Fails and Olsen, 2003; Amershi et al., 2014) offers an alternative modeling strategy which involves the user more actively – interactive MLT development. Perhaps this phase should be decomposes into smaller, more regular training / validation cycles in which the users are involved.

4. (iv) Quality assessment (automatic and / or human)

As shown in Table 2, we observe that in the current MT landscape, a significant amount of work (17 articles) involves users during evaluation, i.e. phase (iv). In these articles, the authors seek feedback primarily from deaf users (5 articles), linguists (2 articles), experts of unclear designation (2 articles), and a combination of Deaf and Hearing individuals (2 articles), as well as hearing, interpreters, interpreters and deaf and teachers (1 article each). These users have been asked to provide feedback on the outcomes and/ or results of the translation such as (Al-Khalifa, 2010; Chiu et al., 2007; Stein et al., 2006; Wu et al., 2007), with some being involved in multiple phases, particularly in phase (ii), in for example, (Khan et al., 2020; Luqman and Mahmoud, 2019; Müller et al., 2022; Rodríguez et al., 2020)). Thus, users appear to be involved in overlapping roles, such as data collection and/ or preparation, and MT system evaluation. While this is a solid starting point, we would like to highlight that continuous assessment across multiple MT phases can be beneficial for managing expectations and aligning participants and goals, i.e. following Principle 8 (The thin red line between consent and intrusion) and Principle 1 (Consensus and conflict) of (Caselli et al., 2021).

We ought to note that, as can be inferred from

Table 2, there exist collaboration and knowledge exchange between researchers, interpreters, experts (sometimes undefined), linguists, and teachers, while individuals who are Hard of Hearing (HoH) and Deaf have not been recognized and involved as language experts. However, historically and still to day, these collaborations are limited to certain distinct roles that do not align with Caselli et al. (2021)’s guidelines nor with Lepp et al. (2025)’s recommendations for effective co-creation. Recognizing that HoH and Deaf individuals are not only native¹⁴ in their language, and therefore possess the most hands-on knowledge and experience, but are also the main end-users of SLMT systems who can steer the development of such technology, as well as expanding their involvement in future projects would allow for a more socially relevant translation technology.

4 Critical Reflection on User Involvement in SLMT Phases

The analysis of user involvement in the different phases of Sign Language Machine Translation (SLMT) reveals significant gaps in alignment with co-creation and participatory design best practices. While there has been progress in areas like data collection and evaluation, key phases such as problem definition and model development lack sufficient user input. This under representation not only affects the quality of the technology but raises ethical concerns about the exclusion of the very users SLMT is designed to serve.

Phase (i): Problem and Use-Case Definition

User involvement in the early phase of problem definition and solution ideation is crucial to ensure that SLMT technologies address real-world needs. However, only one study (Morrissey and Way, 2007) reports any user involvement in this phase. This omission contradicts several co-creation principles (Caselli et al., 2021), such as the importance of early collaboration and continuous design iteration. Without user input at this stage, there is a risk that the technologies developed may not adequately reflect the needs and experiences of the Deaf and Hard of Hearing (HoH) communities, leading to solutions that are out of touch with

¹⁴As Vandeghinste et al. (2024) indicate, the term “native” signer is an “ill-fitting label” and instead, the term “authentic” signers should be used. However, for parallels with “native” or L1 speakers, we chose to use this widely accepted formulation.

real-world challenges.

Phase (ii): Dataset Construction SL users play a more substantial role in dataset construction, primarily through data collection, annotation, and recording. However, these tasks often place users in passive roles, with little involvement in designing the data collection process itself. This limits the potential for authentic and representative data. Research (e.g., (Vandeghinste et al., 2024)) shows that non-native signers or those unfamiliar with specific sign languages can distort data, leading to less accurate translations. A more participatory approach where users are actively involved in shaping data collection methods would help ensure the quality and authenticity of the datasets used for SLMT.

Phase (iii): Model Development Model development is a technically challenging phase, and only one study reports SL user involvement in this area. While experts are essential for model design and optimization, the limited involvement of SL users in this phase perpetuates a divide between technical expertise and user experience. Emerging approaches like interactive machine learning (Fails and Olsen, 2003; Amershi et al., 2014) suggest that more iterative, user-driven cycles of model development could better integrate user insights and improve the relevance of the technology.

Phase (iv): Quality Assessment In the quality assessment phase, SL users are most actively involved, with 17 studies seeking user feedback on translation outputs. However, this involvement is often limited to evaluation, without clear recognition of Deaf and HoH individuals as primary language experts. This oversight reduces the potential of user-driven insights. Additionally, feedback from users is often not integrated into earlier phases, preventing a continuous, cross-phase dialogue that could better align expectations and outcomes. A more integrated approach, where users contribute to evaluation across multiple phases, would ensure that SLMT systems better meet their needs.

5 Conclusion

Since its inception in the 1950s, a significant progress has been made in the field of Machine

Translation (MT) for Spoken Languages reaching to human-like quality. However, the evolution of Sign Language Machine Translation (SLMT) has been slower due to a variety of challenges, including the complexity of data collection, modeling, as well as the intricate nature of human involvement. In contrast to SpL MT, where nowadays user involvement is often limited to data collection and evaluation, the inclusion of users of SLMT technology in all phases of SLMT research and development is quite important, although often overlooked. For example, in phase (i) users should be involved in order to identify the right use cases, and aid the ideation of a societal-relevant solution; in phase (ii) users should be involved to work on the data (record, annotate, guide); in phase (iv) users should be involved in evaluating the model and validating the solution. Phase (iii) requires expert knowledge to design, develop and validate a model; however, in line with the work on iterative ML (Fails and Olsen, 2003; Amershi et al., 2014), perhaps this phase should be broken down into smaller, more regular training/ validation cycles in which the users are involved. We leave testing this idea for future work.

To assess the current state of user involvement in SLMT, we analyzed 111 articles that were previously reviewed in (Núñez-Marcos et al., 2023; Coster et al., 2024). Our analysis reveals that user involvement in SLMT is still largely limited, with substantial participation in phases (ii) and (iv), that is – as content creators, monitors and evaluators but with minimal participation during early phases such as problem definition and model design. This lack of engagement can result in poorly aligned expectations, suboptimal outputs, and ethical concerns, particularly when non-signers are involved in data creation (as indicated, among others, by Buchan et al. (2017); Caselli et al. (2021); Morley et al. (2023)). To address these issues, we advocate for a more participatory approach, where SL users are integral collaborators, not just data providers or evaluators. This requires adjusting the roles of SL users to better reflect their expertise and ensuring that they are involved throughout the entire MT life cycle.

Overall, this work calls for a shift in how SLMT projects are approached, emphasizing the importance of co-creation and partnership with the SL user and SLCs to ensure that translation technologies are developed in a way that is both technically sound and socially responsible.

6 Author contributions

Conceptualization: L. Lepp;
 Methodology: L. Lepp;
 Validation: D. Shterionov;
 Formal analysis: L. Lepp;
 Investigation: L. Lepp;
 Resources: L. Lepp, M. De Sisto, D. Shterionov;
 Data curation: L. Lepp, M. De Sisto, D. Shterionov;
 Writing – original draft: L. Lepp;
 Writing – review and editing: L. Lepp, M. De Sisto, D. Shterionov;
 Visualization: L. Lepp;
 Supervision: M. De Sisto, D. Shterionov, G. Chrupała;
 Project administration: M. De Sisto, D. Shterionov, G. Chrupała;
 Funding acquisition: D. Shterionov and M. De Sisto.

References

- Hend S. Al-Khalifa. 2010. [Introducing arabic sign language for mobile phones](#). In *Computers Helping People with Special Needs*, pages 213–220, Berlin, Heidelberg. Springer.
- Saleema Amershi, Maya Cakmak, W. Bradley Knox, and Todd Kulesza. 2014. [Power to the people: The role of humans in interactive machine learning](#). *AI Mag.*, 35(4):105–120.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. [Neural machine translation by jointly learning to align and translate](#). In *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, CA, USA. 15pp.
- Jim Buchan, Muneera Bano, Didar Zowghi, Stephen MacDonell, and Amrita Shinde. 2017. [Alignment of stakeholder expectations about user involvement in agile software development](#). In *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering, EASE '17*, page 334–343, New York, NY, USA. Association for Computing Machinery.
- Tommaso Caselli, Roberto Cibin, Costanza Conforti, Enrique Encinas, and Maurizio Teli. 2021. [Guiding principles for participatory design-inspired natural language processing](#). In *Proceedings of the 1st Workshop on NLP for Positive Impact*, pages 27–35, Online. Association for Computational Linguistics.
- Yu-Hsien Chiu, Chung-Hsien Wu, Hung-Yu Su, and Chih-Jen Cheng. 2007. [Joint optimization of word alignment and epenthesis generation for chinese to taiwanese sign synthesis](#). *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(1):28–39.

- Mathieu De Coster, Mieke Van Herreweghe, and Joni Dambre. 2019. Towards automatic sign language corpus annotation using deep learning. In *6th Workshop on Sign Language Translation and Avatar Technology (SLTAT)*, page 3.
- Mathieu De Coster, Dimitar Shterionov, Mieke Van Herreweghe, and Joni Dambre. 2024. [Machine translation from signed to spoken languages: state of the art and challenges](#). *Univers. Access Inf. Soc.*, 23(3):1305–1331.
- Maartje De Meulder. 2021. [Is “good enough” good enough? ethical and responsible development of sign language technologies](#). In *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*, pages 12–22, Virtual. Association for Machine Translation in the Americas.
- Aashaka Desai, Maartje de Meulder, Julie A. Hochgesang, Annemarie Kocab, and Alex X. Lu. 2024. [Systemic biases in sign language AI research: A deaf-led call to reevaluate research agendas](#). *CoRR*, abs/2403.02563.
- Zijian Ding, Alison Smith-Renner, Wenjuan Zhang, Joel Tetreault, and Alejandro Jaimes. 2023. [Harnessing the power of LLMs: Evaluating human-AI text co-creation through the lens of news headline generation](#). In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 3321–3339, Singapore. Association for Computational Linguistics.
- Sarah Ebling and Matt Huenerfauth. 2015. [Bridging the gap between sign language machine translation and sign language animation using sequence classification](#). pages 2–9.
- Jerry Alan Fails and Dan R. Olsen. 2003. [Interactive machine learning](#). In *Proceedings of the 8th International Conference on Intelligent User Interfaces, UI ’03*, page 39–45, New York, NY, USA. Association for Computing Machinery.
- Hugo Gonalo Oliveira, Tiago Mendes, and Ana Boavida. 2017. [Co-PoeTryMe: a co-creative interface for the composition of poetry](#). In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 70–71, Santiago de Compostela, Spain. Association for Computational Linguistics.
- Marie K. Harder, Gemma Burford, and Elona Hoover. 2013. [What Is Participation? Design Leads the Way to a Cross-Disciplinary Framework](#). *Design Issues*, 29(4):41–57.
- Pavel Jedlička, Zdeněk Krňoul, Jakub Kanis, and Miloš Železný. 2020. [Sign language motion capture dataset for data-driven synthesis](#). In *Proceedings of the LREC2020 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*, pages 101–106, Marseille, France. European Language Resources Association (ELRA).
- Nabeel Sabir Khan, Adnan Abid, and Kamran Abid. 2020. [A novel natural language processing \(nlp\)-based machine translation model for english to pakistan sign language translation](#). *Cognitive Computation*, 12(4):748–765.
- Kai Konen, Sophie Jentzsch, Diaoulé Diallo, Peer Schütt, Oliver Bensch, Roxanne El Baff, Dominik Opitz, and Tobias Hecking. 2024. [Style vectors for steering generative large language models](#). In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 782–802, St. Julian’s, Malta. Association for Computational Linguistics.
- Lisa Lepp, Dimitar Shterionov, Mirella De Sisto, and Grzegorz Chrupała. 2025. [Co-Creation for Sign Language Processing and Translation Technology](#). *Information*, 16(4).
- Verónica López-Ludeña, Rubén San Segundo, Juan Manuel Montero, Ricardo de Córdoba, Javier Ferreiros, and José Manuel Pardo. 2012. [Automatic categorization for improving spanish into spanish sign language machine translation](#). *Comput. Speech Lang.*, 26(3):149–167.
- Hamzah Luqman and Sabri A. Mahmoud. 2019. [Automatic translation of Arabic text-to-Arabic sign language](#). *Universal Access in the Information Society*, 18(4):939–951.
- Enrique Manjavacas, Folgert Karsdorp, Ben Burtenshaw, and Mike Kestemont. 2017. [Synthetic literature: Writing science fiction in a co-creative process](#). In *Proceedings of the Workshop on Computational Creativity in Natural Language Generation (CC-NLG 2017)*, pages 29–37, Santiago de Compostela, Spain. Association for Computational Linguistics.
- Ian Marshall and Éva Sáfár. 2002. [Sign language generation using HPSG](#). In *Proceedings of the 9th Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages: Papers*, Keihanna, Japan.
- Jessica Morley, Libby Kinsey, Anat Elhalal, Francesca Garcia, Marta Ziosi, and Luciano Floridi. 2023. [Operationalising AI ethics: barriers, enablers and next steps](#). *AI Soc.*, 38(1):411–423.
- Sara Morrissey and Andy Way. 2007. [Joining hands: Developing A sign language machine translation system with and for the deaf community](#). In *Proceedings of the Conference and Workshop on Assistive Technologies for People with Vision and Hearing Impairments: Assistive Technology for All Ages (CVHI-2007)*, Granada, Spain, 28th - 31th August, 2007, volume 415 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Mathias Müller, Sarah Ebling, Eleftherios Avramidis, Alessia Battisti, Michèle Berger, Richard Bowden, Annelies Braffort, Necati Cihan Camgöz, Cristina España-Bonet, Roman Grundkiewicz, Zifan Jiang,

- Oscar Koller, Amit Moryossef, Regula Perrollaz, Sabine Reinhard, Annette Rios, Dimitar Shterionov, Sandra Sidler-Miserez, and Katja Tissi. 2022. [Findings of the first WMT shared task on sign language translation \(WMT-SLT22\)](#). In *Proceedings of the Seventh Conference on Machine Translation, WMT 2022, Abu Dhabi, United Arab Emirates (Hybrid), December 7-8, 2022*, pages 744–772. Association for Computational Linguistics.
- Takao Nakaguchi, Masayuki Otani, Toshiyuki Takasaki, and Toru Ishida. 2016. [Combining human inputters and language services to provide multi-language support system for international symposiums](#). In *Proceedings of the Third International Workshop on Worldwide Language Service Infrastructure and Second Workshop on Open Infrastructures and Analysis Frameworks for Human Language Technologies (WLSI/OIAF4HLT2016)*, pages 28–35, Osaka, Japan. The COLING 2016 Organizing Committee.
- Adrián Núñez-Marcos, Olatz Perez-de-Viñaspre, and Gorka Labaka. 2023. [A survey on Sign Language Machine Translation](#). *Expert Syst. Appl.*, 213(Part):118993.
- Franz Josef Och and Hermann Ney. 2003. [A systematic comparison of various statistical alignment models](#). *Computational Linguistics*, 29(1):19–51.
- Jefferson Rodríguez, Juan Chacón, Edgar Rangel, Luis Guayacán, Claudia Hernández, Luisa Hernández, and Fabio Martínez. 2020. [Understanding motion in sign language: A new structured translation dataset](#). In *Computer Vision - ACCV 2020 - 15th Asian Conference on Computer Vision, Kyoto, Japan, November 30 - December 4, 2020, Revised Selected Papers, Part VI*, volume 12627 of *Lecture Notes in Computer Science*, pages 669–684. Springer.
- Ashish Sharma, Sudha Rao, Chris Brockett, Akanksha Malhotra, Nebojsa Jojic, and Bill Dolan. 2024. [Investigating agency of LLMs in human-AI collaboration tasks](#). In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1968–1987, St. Julian’s, Malta. Association for Computational Linguistics.
- Dimitar Shterionov, Lorraine Leeson, and Andy Way. 2024. [The Pipeline of Sign Language Machine Translation](#), pages 1–25. Springer Nature Switzerland, Cham.
- Mirella De Sisto, Vincent Vandeghinste, Santiago Egea Gómez, Mathieu De Coster, Dimitar Shterionov, and Horacio Saggion. 2022. [Challenges with sign language datasets for sign language recognition and translation](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference, LREC 2022, Marseille, France, 20-25 June 2022*, pages 2478–2487. European Language Resources Association.
- Daniel Stein, Jan Bungeroth, and Hermann Ney. 2006. [Morpho-syntax based statistical methods for automatic sign language translation](#). In *European Association for Machine Translation Conferences/Workshops*.
- Vincent Vandeghinste, Mirella De Sisto, Santiago Egea Gómez, and Mathieu De Coster. 2024. [Challenges with Sign Language Datasets](#), pages 117–139. Springer Nature Switzerland, Cham.
- Chung-Hsien Wu, Hung-Yu Su, Yu-Hsien Chiu, and Chia-Hung Lin. 2007. [Transfer-based statistical translation of taiwanese sign language using PCFG](#). *ACM Trans. Asian Lang. Inf. Process.*, 6(1):1.

A Lepp et al.'s typology

The typology of Lepp et al. (2025) is shown in Table 3.¹⁵

B Caselli et al.'s principles

1. **PD is about consensus and conflict.** The design of co-creation should be conducted in discussion and alignment between the involved parties.
2. **Design is an inherently disordered and unfinished process.** The design should be a continuous, reflexive and ongoing process (principle 2 and 6 of (Caselli et al., 2021) and level 4c of our proposed typology in Table 3. (Caselli et al., 2021) mention that the term *community* needs to be defined in a reflexive and adaptable manner, with its continuous changes. (Harder et al., 2013) assume that this definition is a fixed format, based on the amount of *power* of different researchers (i.e. hearing, HoH or deaf) to define the SLC.
3. Communities are often **not determined a priori**.
4. **Data and communities are not separate things** Principle 4 of (Caselli et al., 2021) contain the assumption that we expect that communities have a prominent role in the development of NLP-systems, but that the communities until now most often only function as language data providers. This assumption raises the question where the separation line between SL-user and researchers is, or in which cases the SL-user indeed only provides data. In the last case we can categorize this on level 2 of (Harder et al., 2013).
5. **Community involvement is not scraping** In principle 5, the social interactions are described as necessary for the creation or development of a tool for a specific community, wherein also the ethical engagements, equity, reciprocity, and respect should be discussed. As Level 4.b. and level 4.c assume that working together in equality, with clear ethical practices are already described, this principle is also hard to divide to one level. Ideally suited -yes- working on equal level is

the highest possible achievement, although in most of the current SLMT projects this step is not implemented or discussed. The development of the expectations/ ethical engagement should be on level 3 (as this part is meant as learning from each others needs) or level 4 (in discussion with each other), and if this is already discussed and decided, then this principle can be divided into level 4b or level 4c for the execution. But also in this case, a reciprocity attitude is needed for reflection and adaption of execution.

6. **Never stop designing** Principle 6 states out that when a NLP-tool is based on PD, there should be awareness about the needs of the SLC and include them into the design stage. By including them, technical and resource issues can be decreased, and participants effort can be recognized as labor.
7. **Language¹⁶ is a means rather than an end.** Principle 7 refers to switch the perspective from *language as data* to *language as people*, wherein the main focus should be to serve people's needs instead of trying to copy people's language use. This principle can ideally be compared with level 4b (Growing as one) or level 4c (Working as one), but in most of the current SLMT this principle is comparable with level 2 -as the researchers need the SLC for this perspective-switch- or level 3, wherein both parties have a discussion and consensus about which perspective is followed.
8. **The thin red line between consent and intrusion** Principle 8 can be part of some of the lower levels already - as soon as some form of recognition of language as people is formed, so this principle can be seen as 'learning about' (level 1) or 'Learning from (Level 2).
9. **The need to combine research goals, funding and societal political dynamics.** The last principle - principle 9 - refers to the complex dynamics of funding (for projects that support co-creation with the community), goals of the research projects, and the community itself. As the most SLMT-projects are not supported

¹⁵The table is added in this article with the agreement of all authors of Lepp et al. (2025).

¹⁶Please be aware that in the article of (Caselli et al., 2021) the original principle is *Text is a means rather than an end*, that we have more specified in this article to language.

Level (-1)		Level (0)	Level (1)	Level (2)	Level (3)	Level (4)		
<i>Denigration direct impact</i>	<i>Denigration indirect impact</i>	<i>Neglect</i>	<i>Learning about</i>	<i>Learning from</i>	<i>Learning together</i>	<i>Learning as one</i>	<i>Growing as one</i>	<i>Working as one</i>
Hearing researchers make decisions without the SLC (neither HoH or deaf researchers) involvement, contrary to the SLCs interests, producing outputs with direct impact on the SLC.	Hearing researchers make decisions without the SLC (neither HoH or deaf researchers) involvement, contrary to or unaware of the SLCs interests, producing outputs with no direct impact on the SLC.	Hearing researchers make decisions without the SLC (neither HoH or Dear researchers) involvement, ignorant or dismissive of the SLCs interests.	Hearing researchers ask the SLCs and the users (and/or HoH or deaf researchers) opinions, but do not necessarily take them into account: the hearing researchers make the final decisions.	Hearing researchers ask the SLCs and the users opinions and consider the SLCs and users seriously. Hearing researchers still makes the final decision based on the information, HoH and deaf researchers are asked for evaluation, but not included in the process.	Major objectives and issues are discussed / negotiated jointly involving hearing, HoH and deaf researchers, and SL users. Most decisions are made jointly, e.g. by consensus-building.	A consortium that includes hearing, HoH and deaf researchers, and SLC members, jointly built, discuss relevant issues by having knowledge exchange (e.g. seminars on different topics from all involved communities).	Hearing, HoH and deaf researchers, and SL users work together on equal basis, are all integrated into the scope of the research cycle, but the SL user is not involved in the execution of each step and / or the societal diversity is not representative.	Hearing, HoH and deaf researchers, and SL users have a full consensus about the practices, the design is a continuous process and both the hearing researchers as well as the SL users are equally integrated into the scope, depth and breadth of the research project.

Table 3: Advanced typology of participation relationships of Lepp et al. (2025).

by a grand for the above needed adaptations, this principle can be compared to level 1 or level 2.

C Paper reviews

1. Angelova, G., Avramidis, E., Möller, S.: Using neural machine translation methods for sign language translation. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pp. 273–284 (2022) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
2. Arvanitis, N., Constantinopoulos, C., Kosmopoulos, D.: Translation of sign language glosses to text using sequence-to-sequence attention models. In: 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), pp. 296–302 (2019). <https://doi.org/10.1109/SITIS.2019.00056>. IEEE Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
3. Barberis, D., Garazzino, N., Prinetto, P., Tiotto, G., Savino, A., Shoab, U., et al. (2011). Language resources for computer assisted translation from italian to italian sign language of deaf people. In Proceedings of accessibility reaching everywhere AEGIS workshop and international conference (pp. 96–104). Deaf involvement: Yes, an interpret that helped in the production of signs Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
4. Bauer, B., Nießen, S., & Hienz, H. (1999). Towards an automatic sign language translation system. In In 1st international. Citeseer. Deaf involvement: Yes, 1 DGS interpreter Evaluation: there are no deaf people involved, only one hearing interpreter for data recordings/ collection Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
5. Brouer, M., & Benabbou, A. (2019). ATLASLang MTS 1: Arabic text language into Arabic Sign Language machine translation system. *Procedia Computer Science*, 148, 236–245. Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
6. Bungeroth, J., Ney, H.: Statistical sign language translation. In: Workshop on Representation and Processing of Sign Languages, LREC, vol. 4, pp. 105–108 (2004). Citeseer Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process.

- Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
7. Camgoz, N. C., Koller, O., Hadfield, S., & Bowden, R. (2020a). Multi-channel trans- formers for multi-articulatory sign language translation. In European conference on computer vision (pp. 301–319). Springer. Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
 8. Camgoz, N.C., Koller, O., Hadfield, S., Bowden, R.: Sign lan- guage transformers: Joint end-to-end sign language recognition and translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10023–10033 (2020) Deaf involvement: Yes, the existing 9 DGS-signers of the dataset Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
 9. Cao, Y., Li, W., Li, X., Chen, M., Chen, G., Hu, L., et al. (2022). Explore more guidance: A task-aware instruction network for sign language translation enhanced with data augmentation. arXiv preprint arXiv:2204.05953. Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
 10. Chaudhary, L., Ananthanarayana, T., Hoq, E., Nwogu, I.: Signnet ii: A transformer-based two-way sign language translation model. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
 11. Chen, Y., Wei, F., Sun, X., Wu, Z., & Lin, S. (2022). A simple multi-modality transfer learning baseline for sign language translation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5120–5130). Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
 12. Chen, Y., Zuo, R., Wei, F., Wu, Y., Liu, S., Mak, B.: Two-stream network for sign language recognition and translation. arXiv pre- print arXiv: 2211. 01367 (2022) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
 13. De Meulder, Bert and Van Landuyt, Marleen and Omardeen, Sadiq: Systemic Biases in Sign Language AI Research: A Deaf-Led Call to Reevaluate Research Agendas. arXiv preprint arXiv:2408.13171 (2024)
 14. D’Haro, L. F., San-Segundo, R., Cordoba, R. d., Bungeroth, J., Stein, D., & Ney, H. (2008). Language model adaptation for a speech to sign language translation system using web frequencies and a map framework. In Ninth annual conference of the international speech communication association. Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
 15. Dasgupta, T., & Basu, A. (2008). Prototype machine translation system from text-to- Indian sign language. In Proceedings of the 13th international conference on intelligent user interfaces (pp. 313–316). Deaf involvement: No Evaluation: we have evaluated the sys?tem based on the feedbacks of the ISL experts Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
 16. Davydov, M., & Lozynska, O. (2017a). Information system for translation into Ukrainian sign language on mobile devices. In 2017 12th international scientific and technical conference on computer sciences and information technologies, Vol. 1 CSIT, (pp. 48–51). IEEE. Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
 17. De Coster, M., D’Oosterlinck, K., Pizurica, M., Rabaey, P., Ver- linden, S., Van Herreweghe, M., Dambre, J.: Frozen pretrained transformers for neural sign language translation. In: Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL), pp. 88–97. Association for Machine Translation in the Americas, Virtual (2021). Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1, but the authors mention that ‘co-creation with the DHH community members is the key’. Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
 18. De Coster, M., Dambre, J.: Leveraging frozen pretrained written language models for neural sign language translation. Informa- tion 13(5), 220 (2022) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
 19. Dey, S., Pal, A., Chaabani, C., Koller, O.: Clean text and full- body transformer: Microsoft’s submission to the wmt22 shared task on sign language translation. In: Proceedings of the Seventh Conference on Machine Translation, pp. 969–976. Association for Computational Linguistics, Abu Dhabi (2022). <https://aclanthology.org/2022.wmt-1.93> Deaf involvement: No (at least not clear mentioned: the authors mention something about human evaluation, but it seems that that is out of the scope of this article). Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps

20. Dreuw, P., Stein, D., Deselaers, T., Rybach, D., Zahedi, M., Bungeroth, J., Ney, H.: Spoken language processing techniques for sign language recognition and translation. *Technol. Disabil.* 20(2), 121–133 (2008) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
21. Dreuw, P., Stein, D., Deselaers, T., Rybach, D., Zahedi, M., Bungeroth, J., et al. (2008). Spoken language processing techniques for sign language recognition and translation. *Technology and Disability*, 20(2), 121–133. Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
22. Dreuw, P., Stein, D., Ney, H.: Enhancing a sign language translation system with vision-based features. In: *International Gesture Workshop*, pp. 108–113 (2007). Springer Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
23. Egea, S., McGill, E., & Saggion, H. (2021). Syntax-aware transformers for neural machine translation: The case of text to sign gloss translation. In *Proceedings of the 14th workshop on building and using comparable corpora*. Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
24. Fang, B., Co, J., & Zhang, M. (2017). DeepASL: Enabling ubiquitous and non-intrusive word and sentence-level sign language translation. In *Proceedings of the 15th ACM conference on embedded network sensor systems* (pp. 1–13). Deaf involvement: 11 hearing participants who learned ASL via 3-hours tutorials Evaluation: level -1: contrary to the SLCs interests) 11 hearing participants who learned ASL via 3-hours tutorials Level: -1 a Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
25. Foong, O. M., Low, T. J., & La, W. W. (2009). V2s: Voice to sign language translation system for malaysian deaf people. In *International visual informatics conference* (pp. 868–876). Springer. Deaf involvement: Yes, 100 people (groups of children, male, female, young and older), but no deaf. Evaluation: It is not focused on SL, but on SpLs Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups / variations are involved Scope: In none of the research life cycle steps
26. Forster, J., Schmidt, C., Hoyoux, T., Koller, O., Zelle, U., Piater, J.H., Ney, H.: *Rwth-phoenix-weather: A large vocabulary sign language recognition and translation corpus*. In: *LREC*, vol. 9, pp. 3785–3789 (2012) Deaf involvement: It was not implemented Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 a Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
27. Forster, J., Schmidt, C., Koller, O., Bellgardt, M., Ney, H.: *Extensions of the sign language recognition and translation corpus rwth-phoenix-weather*. In: *LREC*, pp. 1911–1916 (2014) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
28. Fu, B., Ye, P., Zhang, L., Yu, P., Hu, C., Chen, Y., et al. (2022). ConSLT: A token-level contrastive framework for sign language translation. *arXiv preprint arXiv:2204.04916*. Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
29. Gan, S., Yin, Y., Jiang, Z., Xie, L., Lu, S.: Skeleton-aware neural sign language translation. In: *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 4353–4361 (2021) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
30. Grieve-Smith, A. B. (1999). English to American Sign Language machine translation of weather reports. In *Proceedings of the second high desert student conference in linguistics (HDSL2)*, Albuquerque, NM (pp. 23–30). Deaf involvement: No, although the author mention in future work that the output needs to be cross-checked with a native signer Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
31. Grif, M. G., Korolkova, O. O., Demyanenko, Y. A., & Tsoy, Y. B. (2011). Development of computer sign language translation technology for deaf people. In *Proceedings of 2011 6th international forum on strategic technology*, Vol. 2 (pp. 674–677). IEEE. Deaf involvement: not clear Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
32. Guo, D., Zhou, W., Li, A., Li, H., & Wang, M. (2019). Hierarchical recurrent deep fusion using adaptive clip summarization for sign language translation. *IEEE Transactions on Image Processing*, 29, 1575–1590. Deaf involvement: Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
33. Halawani, S. M. (2008). Arabic sign language translation system on mobile devices. *IJCSNS International Journal of Computer Science and Network Security*, 8(1), 251–256. Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps

34. Hoque, M. T., Rifat-Ut-Tauwab, M., Kabir, M. F., Sarker, F., Huda, M. N., & Abdullah-Al- Mamun, K. (2016). Automated bangla sign language translation system: Prospects, limitations and applications. In 2016 5th international conference on informatics, electronics and vision ICIEV, (pp. 856–862). IEEE. Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 a Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
35. Huang, J., Zhou, W., Zhang, Q., Li, H., Li, W.: Video-based sign language recognition without temporal segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
36. Huenerfauth, M. (2004). A multi-path architecture for machine translation of english text into American Sign language animation. In Proceedings of the student research workshop at HLT-NAACL 2004 (pp. 25–30). Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
37. Jin, T., Zhao, Z., Zhang, M., Zeng, X.: Mc-slt: Towards low- resource signer-adaptive sign language translation. In: Proceedings of the 30th ACM International Conference on Multimedia, pp. 4939–4947 (2022) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
38. Jin, T., Zhao, Z., Zhang, M., Zeng, X.: Prior knowledge and memory enriched transformer for sign language translation. In: Findings of the Association for Computational Linguistics: ACL 2022, pp. 3766–3775 (2022) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
39. Kamata, K., Yoshida, T., Watanabe, M., & Usui, Y. (1989). An approach to Japanese-sign language translation system. In Conference proceedings., IEEE international conference on systems, man and cybernetics (pp. 1089–1090). IEEE Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
40. Kan, J., Hu, K., Hagenbuchner, M., Tsoi, A.C., Benamoun, M., Wang, Z.: Sign language translation with hierarchical spatio-temporal graph neural network. In: Proceedings of the IEEE/ CVF Winter Conference on Applications of Computer Vision, pp. 3367–3376 (2022) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
41. Kim, S., Kim, C.J., Park, H.-M., Jeong, Y., Jang, J.Y., Jung, H.: Robust keypoint normalization method for korean sign language translation using transformer. In: 2020 International Conference on Information and Communication Technology Convergence (ICTC), pp. 1303–1305 (2020). <https://doi.org/10.1109/ICTC49870.2020.9289551>. IEEE Deaf involvement: Yes, for training (16 signers) and testing of data (4) Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
42. Kouremenos, D., Ntalianis, K., & Kollias, S. 2018. A novel rule based machine translation scheme from Greek to Greek sign language: Production of different types of large corpora and language models evaluation. 51, 110–135, Deaf involvement: No Evaluation: level -1. A translator, Human evaluation is fundamental and remains of crucial importance to proper assessment of the quality of MT systems. When the output of an MT system is evaluated, however, the whole process is taken into account. In our case, different aspects of the proposed RBMT system are evaluated such as: (a) all stages of development of the transfer rules, (b) accuracy of translation and (c) complexity. Thus, it cannot be understood by deaf people, who cannot read the Greek language. A complete MT system for the GSL should produce animations, while a genuine and proper evaluation should involve deaf people, measuring comprehension regarding the animated output Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
43. Kumar, S.S., Wangyal, T., Saboo, V., Srinath, R.: Time series neural networks for real time sign language translation. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 243–248 (2018). <https://doi.org/10.1109/ICMLA.2018.00043>. IEEE Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
44. Li, D., Xu, C., Yu, X., Zhang, K., Swift, B., Suominen, H., Li, H.: Tspnet: Hierarchical feature learning via temporal semantic pyramid for sign language translation. Adv. Neural. Inf. Process. Syst. 33, 12034–12045 (2020) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
45. Li, R., Meng, L.: Sign language recognition and translation network based on multi-view data. Appl. Intell. 52(13), 14624–14638 (2022) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
46. López-Ludeña, V., San-Segundo, R., Morcillo, C. G., López, J. C., & Muñoz, J. M. P. (2013). Increasing

- adaptability of a speech into sign language translation system. *Expert Systems with Applications*, 40(4), 1312–1322. Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
47. Luqman, H., Mahmoud, S.A.: A machine translation system from arabic sign language to arabic. *Univ. Access Inf. Soc.* 19(4), 891–904 (2020). <https://doi.org/10.1007/s10209-019-00695-6> Deaf involvement: no Evaluation: There are no deaf people involved: level -1. Evaluation by hearing Arab speakers for translation-evaluation (as the output is Arabic) Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
48. Marshall, I., & Sáfár, É. (2002). Sign language generation using HPSG. In *Proceedings of the 9th Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages: Papers*. Deaf involvement: No Evaluation: There are no deaf people involved (level -1) but the authors are aware of co-creation: Sign research has frequently been carried out by hearing people using deaf informants and hence insights are typically second-hand. Additionally, the status of deaf informants themselves within the deaf community raises a significant issue. Typically only 5-10% of deaf people are born to deaf parents and thus are viewed as the genuine native signers who should act as informants and who should be asked to identify the preferred manner of signing a proposition rather than merely acceptable signing (Neidle et al. 2000). Deaf informants with hearing researchers and initial review by hearing signers are used to establish initial hypotheses. More extensive review by deaf users of the generated signing provides detailed feedback and guides revision. Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
49. Marshall, I., & Sáfár, É. (2003). A prototype text to British Sign Language (BSL) translation system. In *The companion volume to the proceedings of 41st annual meeting of the association for computational linguistics* (pp. 113–116). Deaf involvement: Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
50. Miranda, P.B., Casadei, V., Silva, E., Silva, J., Alves, M., Severo, M., Freitas, J.P.: Tspnet-hf: A hand/face tspnet method for sign language translation. In: *Ibero-American Conference on Artificial Intelligence*, pp. 305–316 (2022). Springer Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
51. Mohamed, A., Hefny, H., et al.: A deep learning approach for gloss sign language translation using transformer. *Journal of Computing and Communication* 1(2), 1–8 (2022) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
52. Morrissey, S. (2008). Assistive translation technology for deaf people: translating into and animating Irish sign language. Deaf involvement: Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: Only (hearing) researchers Scope: Only (hearing) researchers
53. Morrissey, S., & Way, A. (2005). An example-based approach to translating sign language. Deaf involvement: Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
54. Morrissey, S., & Way, A. (2006). Lost in translation: the problems of using mainstream MT evaluation metrics for sign language translation Deaf involvement: The authors mention: Clearly, in addition, human evaluation remains crucial for all such approaches. Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
55. Morrissey, S., Way, A., Stein, D., Bungeroth, J., Ney, H.: Combining data-driven mt systems for improved sign language translation. In: *European Association for Machine Translation (2007)* Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
56. Moryossef, A., Yin, K., Neubig, G., Goldberg, Y.: Data augmentation for sign language gloss translation. In: *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*, pp. 1–11. Association for Machine Translation in the Americas, Virtual (2021). <https://aclanthology.org/2021.mtsummit-at4ssl.1> Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
57. Nießen, S., & Ney, H. (2004). Statistical machine translation with scarce resources using morpho-syntactic information. *Computational Linguistics*, 30(2), 181–204. Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
58. Orbay, A., Akarun, L.: Neural sign language translation by learning tokenization. In: *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pp. 222–228 (2020). IEEE Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps

59. Othman, A., Jemni, M.: English-asl gloss parallel corpus 2012: Aslg-pc12. In: 5th Workshop on the Representation and Processing of Sign Languages: Interactions Between Corpus and Lexicon LREC (2012) Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
60. Partaourides, H., Voskou, A., Kosmopoulos, D., Chatzis, S., Metaxas, D.N.: Variational bayesian sequence-to-sequence networks for memory-efficient sign language translation. In: International Symposium on Visual Computing, pp. 251–262 (2020). Springer Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
61. Porta, J., López-Colino, F., Tejedor, J., & Colás, J. (2014). A rule-based translation from written Spanish to Spanish Sign Language glosses. *Computer Speech and Language*, 28(3), 788–811. Deaf involvement: no Evaluation: Level -1 A parallel Spanish-LSE corpus has been created by two hearing interpreters (one of them was CODA) Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
62. Sáfár, É., & Marshall, I. (2001). The architecture of an english-text-to-sign-languages translation system. In *Recent advances in natural language processing RANLP*, (pp. 223–228). Tzigov Chark Bulgaria. Deaf involvement: Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
63. Sáfár, É., & Marshall, I. (2002). Sign Language Translation via DRT and HPSG. *Conference on Intelligent Text Processing and Computational Linguistics*. Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
64. San Segundo, R., Pérez, A., Ortiz, D., Luis Fernando, D., Torres, M. I., & Casacuberta, F. (2007). Evaluation of alternatives on speech to sign language translation. In *INTERSPEECH* (pp. 2529–2532). Citeseer. Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
65. San-Segundo, R., Barra, R., Córdoba, R., D’Haro, L. F., Fernández, F., Ferreiros, J., et al. (2008). Speech to sign language translation system for Spanish. *Speech Communication*, 50(11–12), 1009–1020. Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
66. San-Segundo, R., Barra, R., D’Haro, L., Montero, J. M., Córdoba, R., & Ferreiros, J. (2006). A spanish speech to sign language translation system for assisting deaf-mute people. In *Ninth international conference on spoken language processing*. Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
67. Saunders, B., Camgoz, N. C., & Bowden, R. (2020b). Progressive transformers for end-to-end sign language production. In *European conference on computer vision* (pp. 687–705). Springer. Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
68. Schmidt, C., Koller, O., Ney, H., Hoyoux, T., Piater, J.: Using viseme recognition to improve a sign language translation system. In: *International Workshop on Spoken Language Translation*, pp. 197–203 (2013). Citeseer Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
69. Stein, D., Dreuw, P., Ney, H., Morrissey, S., Way, A.: Hand in hand: automatic sign language to English translation. In: *Proceedings of the 11th Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages: Papers, Skövde, Sweden (2007)*. <https://aclanthology.org/2007.tmi-papers.26> Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
70. Stein, D., Schmidt, C., Ney, H.: Analysis, preparation, and optimization of statistical sign language machine translation. *Mach. Transl.* 26(4), 325–357 (2012) Deaf involvement: Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
71. Stein, D., Schmidt, C., Ney, H.: Sign language machine translation overkill. In: *International Workshop on Spoken Language Translation (IWSLT) 2010* (2010) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
72. Stoll, S., Camgöz, N. C., Hadfield, S., & Bowden, R. (2018). Sign language production using neural machine translation and generative adversarial networks. In *Proceedings of the 29th British machine vision conference (BMVC 2018)*. University of Surrey. Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps

73. Tarres, L., Gállego, G.I., Giro-i-Nieto, X., Torres, J.: Tackling low-resourced sign language translation: Upc at wmt-slt 22. In: Proceedings of the Seventh Conference on Machine Translation, pp. 994–1000. Association for Computational Linguistics, Abu Dhabi (2022). <https://aclanthology.org/2022.wmt-1.97> Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
74. Tokuda, M., & Okumura, M. (1998). Towards automatic translation from Japanese into Japanese sign language. In Assistive technology and artificial intelligence (pp. 97–108). Springer. Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
75. Wazalwar, S. S., & Shrawankar, U. (2017). Interpretation of sign language into English using NLP techniques. *Journal of Information and Optimization Sciences*, 38(6), 895–910. Deaf involvement: no Evaluation: The videos were interpreted by hearing teachers of school for the deaf Level -1 Level: -1 a Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
76. Yin, A., Zhao, Z., Jin, W., Zhang, M., Zeng, X., He, X.: Mslst: Towards multilingual sign language translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5109–5119 (2022) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
77. Yin, A., Zhao, Z., Liu, J., Jin, W., Zhang, M., Zeng, X., He, X.: Simulst: End-to-end simultaneous sign language translation. In: Proceedings of the 29th ACM International Conference on Multimedia, pp. 4118–4127 (2021) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
78. Yin, K., Read, J.: Better sign language translation with stmc-transformer. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 5975–5989 (2020). <https://doi.org/10.18653/v1/2020.coling-main.525> Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
79. Zhang, X., Duh, K.: Approaching sign language gloss translation as a low-resource machine translation task. In: Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL), pp. 60–70. Association for Machine Translation in the Americas, Virtual (2021). <https://aclanthology.org/2021.mtsummit-at4ssl.7> Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
80. Zhao, J., Qi, W., Zhou, W., Duan, N., Zhou, M., & Li, H. (2021). Conditional sentence generation and cross-modal reranking for sign language translation. *IEEE Transactions on Multimedia*, 24, 2662–2672. Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
81. Zhao, L., Kipper, K., Schuler, W., Vogler, C., Badler, N., & Palmer, M. (2000). A machine translation system from English to American sign language. In Conference of the association for machine translation in the Americas (pp. 54–67). Springer. Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
82. Zheng, J., Chen, Y., Wu, C., Shi, X., & Kamal, S. M. (2021). Enhancing neural sign language translation by highlighting the facial expression information. *Neurocomputing*, 464, 462–472. Deaf involvement: No Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
83. Zheng, J., Zhao, Z., Chen, M., Chen, J., Wu, C., Chen, Y., Shi, X., Tong, Y.: An improved sign language translation model with explainable adaptations for processing long sign sentences. *Computational Intelligence and Neuroscience 2020* (2020) Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
84. Zhou, H., Zhou, W., Zhou, Y., Li, H.: Spatial-temporal multi-cue network for sign language recognition and translation. *IEEE Trans. Multimedia* (2021). <https://doi.org/10.1109/TMM.2021.3059098> Deaf involvement: no Evaluation: there are no deaf people involved: only focus on the MT-process. Level -1 Level: -1 b Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
85. Baldassarri, S., Cerezo, E., & Royo-Santas, F. (2009). Automatic translation system to Spanish sign language with a virtual interpreter. In IFIP conference on human-computer interaction (pp. 196–199). Springer. Deaf involvement: Yes, two teachers of a school for interpreters Evaluation: Level 0? Or level 1? Assessment was done by two teachers of a school of interpreters considering the accuracy of two aspects: the translation and the synthesis of the signs by the virtual interpreter. Level: 0 Depth: (hearing) researchers, SL-interpreters, but not the SL-user Breadth: little variation (not the SL-user involved) Scope: In the evaluation/ reflection phrase

86. Camgoz, N.C., Hadfield, S., Koller, O., Ney, H., Bowden, R.: Neural sign language translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7784–7793 (2018). <https://doi.org/10.1109/CVPR.2018.00812> Deaf involvement: no Evaluation: Level 0. For the corpus, they used 9 different signers. Furthermore, the corpus annotations are made by SL-interpreters and deaf specialists. Level: 0 Depth: Breadth: Scope:
87. Camgöz, N.C., Saunders, B., Rochette, G., Giovanelli, M., Inches, G., Nachtrab-Ribback, R., Bowden, R.: Content4all open research sign language translation datasets. In: 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), pp. 1–5 (2021). <https://doi.org/10.1109/FG52635.2021.9667087> Deaf involvement: Evaluation: Level 0. There were deaf experts and SL interpreters for the matching of the SPLs text with the corresponding SL-video pairs, and annotation process Level: 0 Depth: Breadth: Scope:
88. Dal Bianco, P., Ríos, G., Ronchetti, F., Quiroga, F., Stanchi, O., Hasperué, W., Rosete, A.: Lsa-t: The first continuous argentinian sign language dataset for sign language translation. In: Ibero- American Conference on Artificial Intelligence, pp. 293–304 (2022). Springer Deaf involvement: no, a generated corpus from videos of YouTube Evaluation: Level 0. Videos of channel CN Sordos, a news channel created by deaf people and deaf people's relatives. 103 deaf signers as guests Level: 0 Depth: Breadth: Scope:
89. Ebling, S., & Huenerfauth, M. (2015). Bridging the gap between sign language machine translation and sign language animation using sequence classification. In Proceedings of SLPAT 2015: 6th workshop on speech and language processing for assistive technologies (pp. 2–9). Deaf involvement: Yes, deaf and hearing team members (translating) Evaluation: Yes, deaf and hearing team members (translating), but not clear to what extent (level 0) Level: 0 Depth: Breadth: Scope:
90. Ko, S.-K., Kim, C.J., Jung, H., Cho, C.: Neural sign language translation based on human keypoint estimation. *Appl. Sci.* 9(13), 2683 (2019) Deaf involvement: no Evaluation: level 0. 14 hearing-impaired for recordings (a copy of an 'expert' signing the requested signs, which the signers needed to copy) Level: 0 Depth: Breadth: Scope:
91. Krňoul, Z., Kanis, J., Železny, M., & Müller, L. (2007). Czech text-to-sign speech synthesizer. In International workshop on machine learning for multimodal interaction (pp. 180–191). Springer. Deaf involvement: two participants for the evaluation of the Sign Speech synthesizer Evaluation: Level 0. two experts in SignSpeech for the evaluation of the Sign Speech synthesizer Level: 0 Depth: Breadth: Scope:
92. Massó, G., & Badia, T. (2010). Dealing with sign language morphemes in statistical machine translation. In 4th workshop on the representation and processing of sign languages: Corpora and sign language technologies, Valletta, Malta (pp. 154–157). Matthes, S., Hanke, T., Regen, A., Storz, J., Wörseck, S., Efthimiou, E., et al. (2012). Deaf involvement: No, the authors mention: Unfortunately, it was not possible to conduct a human evaluation by native deaf signers Evaluation: (level 0) The authors created a corpus based on Catalan Weather texts which were translated by a native deaf signer Level: 0 Depth: Breadth: Scope:
93. Moe, S.Z., Thu, Y.K., Thant, H.A., Min, N.W., Supnithi, T.: Unsupervised Neural Machine Translation between Myanmar Sign Language and Myanmar Language. *TIC* 14(15), 16 (2020) Deaf involvement: Yes, for data collection Evaluation: Level 0. 30 SL trainers and deaf people from different MSL dialects for data collection Level: 0 Depth: Breadth: Scope:
94. Moe, S.Z., Thu, Y.K., Thant, H.A., Min, N.W.: Neural Machine Translation between Myanmar Sign Language and Myanmar Written Text. In: the Second Regional Conference on Optical Character Recognition and Natural Language Processing Technologies for ASEAN Languages, pp. 13–14 (2018) Deaf involvement: no Evaluation: Level 0. Yes, data collection of 22 SL-trainers, and deaf people with different MSL dialects and different ages Level: 0 Depth: Breadth: Scope:
95. Morrissey, S. (2011). Assessing three representation methods for sign language machine translation and evaluation. In Proceedings of the 15th annual meeting of the European association for machine translation (EAMT 2011), Leuven, Belgium (pp. 137–144). Citeseer. Deaf involvement: However, the authors pointed out that, given the automatic evaluation used, it was not clear which was the best format and that experiments should be accompanied by human evaluation to ascertain the translation quality Evaluation: Level 0. A native ISL signer manually translated and signed the dialogue in ISL Level: 0 Depth: Breadth: Scope:
96. Müller, M., Ebling, S., Avramidis, E., Battisti, A., Berger, M., Bowden, R., Brafort, A., Cihan Camgöz, N., España-Bonet, C., Grundkiewicz, R., Jiang, Z., Koller, O., Moryossef, A., Perrollaz, R., Reinhard, S., Rios, A., Shterionov, D., Sidler-Miserez, S., Tissi, K., Van Landuyt, D.: Findings of the first wmt shared task on sign language translation (wmt-slt22). In: Proceedings of the Seventh Conference on Machine Translation, pp. 744–772. Association for Computational Linguistics, Abu Dhabi (2022). <https://aclanthology.org/2022.wmt-1.71> Deaf involvement: Seven teams participated, four native German speakers who were educated interpreters Evaluation: Level 0. Manually correction of subtitles by deaf signers, evaluators trained DGSG interpreters Level: 0 Depth: Breadth: Scope:
97. Rodriguez, J., Martinez, F.: How important is motion in sign language translation? *IET Comput. Vision* 15(3), 224–234 (2021) Deaf involvement: No Evaluation: Level 0. 9 deaf signers and 2 CODAs for the recordings of the dataset Level: 0 Depth: Breadth: Scope:
98. Al-Khalifa, H. S. (2010). Introducing Arabic sign language for mobile phones. In International conference on computers for handicapped persons (pp. 213–220). Springer. Deaf involvement: Evaluating of the system, not clear if the group of users were deaf. Evaluation: Five participants: 3 deaf and 2 non-deaf people answered a survey (level 1) Level: 1 Depth: Breadth: Scope:
99. Chiu, Y.-H., Wu, C.-H., Su, H.-Y., & Cheng, C.-J. (2006). Joint optimization of word alignment and epenthesis generation for Chinese to Taiwanese sign synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1), 28–39. Deaf involvement:

- Subjective evaluation (with missing how many, who are the subjects etc) Evaluation: Level 1 Five profoundly deaf students in the sixth grade evaluated the utility of the proposed approach as practical learning aid Level: 1 Depth: 5 SL users + (hearing) researchers Breadth: Little variation Scope: In the evaluation/ reflection phrase
100. Hilzensauer, M., Krammer, K.: A multilingual dictionary for sign languages: “spreadthesign”. ICERI2015 Proceedings, 7826–7834 (2015) Deaf involvement: no Evaluation: Level 1. Fifteen partner countries, according to a list, which were discussed with deaf collaborators - who then chose the signs / or sign dialects Level: 1 Depth: Breadth: Scope:
 101. Khan, N. S., Abid, A., & Abid, K. (2020). A novel natural language processing (NLP)- based machine translation model for English to Pakistan sign language translation. *Cognitive Computation*, 12, 748–765. Deaf involvement: Yes, deaf scholars for evaluation (amount is not mentioned) Evaluation: Level 1, Translating English sentences into PSL sentences with the help of SL interpreters and three deaf subjects for recordings, also used for evaluation. Level: 1 Depth: Breadth: Scope:
 102. López-Ludeña, V., San-Segundo, R., Montero, J. M., Córdoba, R., Ferreiros, J., & Pardo, J. M. (2012). Automatic categorization for improving Spanish into Spanish Sign Language machine translation. *Computer Speech and Language*, 26(3), 149–167. Deaf involvement: Two experts in LSE, who were also involved into the corpus generation, but the authors acknowledged that deaf people also should be evaluate how the avatar represents these signs Evaluation: Level 1: These sentences were translated into LSE, both in text (sequence of signs) and in video, and compiled in an excel file. The translation was carried out by two LSE experts in parallel. When there was any discrepancy between them, a committee of four people (one Spanish linguist, 2 deaf LSE experts, and a Spanish linguistic expert on LSE) who knew LSE took the decision: select one of the LSE expert proposals, propose a new one translation alternative, or considering both proposals as alternative translations. Level: 1 Depth: Two LSE experts (for translation) , Spanish linguist, 2 deaf LSE experts and a Spanish-LSE experts Breadth: little variation but the SL-user is involved Scope: implementation, reflection
 103. Luqman, H., & Mahmoud, S. A. (2019). Automatic translation of Arabic text-to-Arabic sign language. *Universal Access in the Information Society*, 18(4), 939–951. Deaf involvement: Yes, evaluation by 1 deaf person and 1 translator Evaluation: level 1: based on wordlist 2 native signers for translating Arabic into ArSL, evaluation by 1 deaf person and 1 expert bilingual translator Level: 1 Depth: (hearing) researchers, SL-interpreters, three deaf subjects Breadth: Little variation in the groups Scope: implementation, reflection
 104. Rodriguez, J., Chacon, J., Rangel, E., Guayacan, L., Hernandez, C., Hernandez, L., Martinez, F.: Understanding motion in sign language: A new structured translation dataset. In: Proceedings of the Asian Conference on Computer Vision (2020) Deaf involvement: Yes, for training and testing of data Evaluation: Level 1. Five deaf signers out of different regions has been recorded, 10 signers for training and testing evaluation. Level: 1 Depth: Breadth: Scope:
 105. Sagawa, H., Ohki, M., Sakiyama, T., Oohira, E., Ikeda, H., & Fujisawa, H. (1996). Pattern recognition and synthesis for a sign language translation system. *Journal of Visual Languages and Computing*, 7(1), 109–127. Deaf involvement: Yes, 1 deaf person for data-collection (level -1 or level 0) Evaluation: Four hearing-impaired and two interpreters evaluated the SL sentences (level 1) Level: 1 Depth: Only (hearing) researchers Breadth: 1 deaf person, four HoH persons, two interpreters Scope: In data-collection and evaluation
 106. San-Segundo, R., López, V., Martín, R., Sánchez, D., García, A.: Language resources for Spanish–Spanish sign language (lse) translation. In: Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies at LREC, pp. 208–211 (2010) Deaf involvement: Yes, ten deaf signers tested the system in a real-life situation Evaluation: Level 1. the first day was an information day about the project and the evaluation, the second day within 6 different scenarios was tested. Level: 1 Depth: Only (hearing) researchers Breadth: No different groups/ variations are involved Scope: In none of the research life cycle steps
 107. Stein, D., Bungeroth, J., & Ney, H. (2006). Morpho-syntax based statistical methods for automatic sign language translation. In Proceedings of the 11th annual conference of the European association for machine translation. Deaf involvement: Yes, for evaluation (2 deaf people) Evaluation: Yes. For the rating of the coherence of a German sentence to the avatar output (level 1) Level: 1 Depth: 2 SL users + (hearing) researchers Breadth: Little variation Scope: In evaluation/ reflection phrase
 108. Su, H.-Y., & Wu, C.-H. (2009). Improving structural statistical machine translation for sign language with small corpus using thematic role templates as translation memory. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(7), 1305–1315. Deaf involvement: 10 deaf students (divided into control and test group) Evaluation: Level 1. The developed parallel bilingual corpus has been annotated and verified by 3 TSL linguists Level: 1 Depth: 10 deaf students divided over 2 groups, and 3 TSL linguists Breadth: Variation by two control and test groups, check by TSL linguists Scope: implementation, reflection
 109. Wu, C.-H., Su, H.-Y., Chiu, Y.-H., & Lin, C.-H. (2007). Transfer-based statistical translation of Taiwanese sign language using PCFG. *ACM Transactions on Asian Language Information Processing (TALIP)*, 6(1), 1–es. Deaf involvement: Subjective evaluation Evaluation: Level 1: group 1: 10 hearing people who used TSL for years, group 2: 10 native TSL signers evaluated the translated sentences Level: 1 Depth: 10 hearing people who used TSL for years + 10 native TSL signers + (hearing) researchers Breadth: Variation by two groups (native and non-native signers) Scope: In the evaluation/ reflection phrase
 110. Zhou, H., Zhou, W., Qi, W., Pu, J., & Li, H. (2021). Improving sign language translation with monolingual data by sign back-translation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1316–1325). Deaf involvement: no Evaluation: level 1. SL linguistic experts, several SL teachers for design of the specific content, 10 native signers for video recording Level: 1 Depth: Breadth: Scope:

111. Jantunen, T., Rousi, R., Rainò, P., Turunen, M., Moeen Valipour, M., & García, N. (2021). Is there any hope for developing automated translation technology for sign languages. *Multilingual Facilitation*, 61–73. Deaf involvement: Evaluation: Level 2. There are NADs included, and also a paragraph about Co-Engineering, Participation and Culture Level: 2 Depth: Breadth: Scope:
112. Morrissey, S., & Way, A. (2007). *Joining hands: Developing a sign language machine translation system with and for the deaf community*. Deaf involvement: two deaf signers for translation work and consultation work + data-collection Evaluation: Level 2? Yes, the involvement of deaf colleagues, members of the deaf community within the choice of a domain for SLT (by asking the Centre for Deaf Studies), the human translation, advice on the SL grammar and linguistics, manual evaluators of the translated output Level: 2 Depth: deaf colleagues + (hearing) researchers Breadth: Deaf Studies, deaf colleagues (in team) and SLC Scope: Initiation, planning, implementation, reflection