Assisting Sign Language Translation: What Interface Given the Lack of Written Form and Spatial Grammar?

Marion Kaczmarek
LIMSI - CNRS, Université Paris Saclay
kaczmarek@limsi.fr

Michael Filhol
LIMSI - CNRS, Université Paris Saclay
michael.filhol@limsi.fr

Abstract

Computer-assisted translation (CAT) software offers tools for the translators to ease their tasks, and gain time as well as comfort. However, despite the growing need for Sign Language content, there has been no effort to equip Sign Language translation with CAT software. The problem we address here is the specification of such software. Sign Languages are visual and iconic, with grammar and discourse organisation, but also no written form. This is problematic when it comes to CAT, for it relies on editable written structures and the fact that the concatenation of the translated segments will result in the translation of the concatenated source segments (we call it the linearity assumption). In this paper, we explain that Sign Language cannot follow those rules. We address those differences by means of new adapted modules which would be more flexible, and by considering new tools based on professionals' feedback towards their actual practice as well as the problems they encounter during the translation process. We will detail those results along with the presentation of how we envisage a sign language concordancer, and its database.

1 Introduction

In our computer-assisted society, translation is no exception. For the past decades, computer-assisted translation (CAT) software has contributed to ease the translators' job. It allows them to work faster and increase their productivity, and also their comfort. The big innovation resides not only in providing tools such as glossaries or dictionaries, but an entire integrated working environment, including the translation memory (TM) which allows to build on previous work and share resources, hence the gain of time. Translation professionals are not the only ones to benefit from CAT. It has permitted to make more documents and information accessible to a broader audience. In 2006, the United Nations adopted the Convention on the Rights of Persons with Disabilities (CRPD). The Convention entered into force in 2008, providing a framework to extend the rights of people with disabilities, including the right to full access to information and communications. However, deaf people are often deprived of this right due to the lack of communications in Sign Language (SL).

Since CRPD, the need for translated SL content kept growing, whether concerning public services, broadcasting, public transportation... or any real-time information. In France, the number of professional SL translators is very low. Only one translator is registered to the AFILS (French Sign Language translators and interpreters association), against a total of 144 interpreters. Despite the high needs for translated content, and the scarcity of professional translators, none of the CAT software currently on the market can support SL. We wish to suggest that as well as their vocal language counterparts, SL translators could benefit from such software equipment. Not only would they improve their working conditions, but allowing to create more SL content would also be a step forward towards SL accessibility. The problem we address here is the specification of such software: what should a CAT program offer for SL, and what should its interface look like?

2 **Sign Language Translation**

SLs are natural oral languages using the visual-manual modality to convey meaning through manual articulations, body gestures, facial expressions, etc. Contrary to audio-vocal languages which are linearly constrained (pitch and volume can vary but only one sound can be produced at a time), SLs express multiple information at a time. They also make heavy use of persistent spatial references which can be reused when needed throughout the discourse. Here is an example of a French text from our data, which we have in its translated form in French Sign Language (LSF):

L'État de Sao Paulo a vécu une deuxième vague d'attaques du crime organisécontre la police, avec un bilan de 52 morts dont 35 policiers en deux jours et la multiplication des mutineries dans les prisons, a-t-on appris dimanche de source officielle.

In English: The State of Sao Paulo experienced a second wave of organized crime, with attacks against the police resulting in 52 deaths including 35 police officers in two days, and the multiplication of mutinies in prisons, as reported on Sunday by official sources.

Relevant screenshots of the LSF translation are given in fig.1, in chronological order of the utterance. Here is the global meaning of what is signed as translation: in Sao Paulo which is located around here in Brazil, criminals already assaulted police forces, and now official sources tell us that it happened again. They announce the number of victims, and the growing mutinies in prisons.

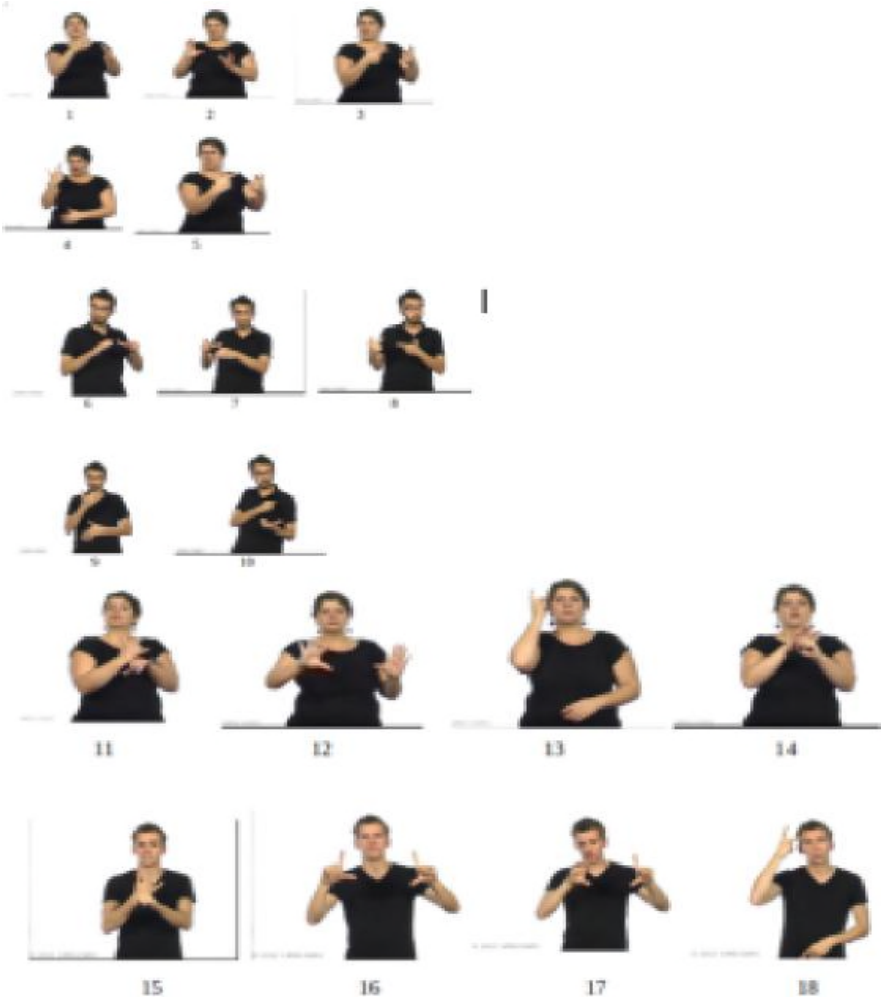

Figure 1: Sign Language example screenshots

In the LSF translation, we have 3 different versions from 3 different signers, signer 1 and 2 set up the two groups involved in the events: the police on the right (1,6), and criminals on the left (2,7).

Then they sign the action: the shooting, which is oriented from the criminals toward the police forces (3,8). More is signed afterwards on the topic of the announcement made by the official sources. But when she comes back to the shooting topic to inform there has been a second attack (4,9), she does not set up the groups a second time. Rather, she uses the previous spatial references (5,10), still active and semantically relevant in her signing space. This illustrates the heavy part played by the 3d signing space in SL, from which ensue other specificities detailed below.

SLs are also iconic languages, meaning that the signs are inspired by reality (what linguists call depicting shapes, Cuxac Sallandre). To some extent, they are sometimes constrained by reality, especially when it comes to locations.

In our example, we can observe two strategies used by the translators. The first one consists in spelling the proper name and produce a generic localization. The second one, used in frames 11 to 18 in fig.1, is about showing where it is. We can observe that both of them first locate Brazil, with the lexical sign for the country. In a second time, they identify a more precise portion of this country, the state of São Paulo, which they locate on the east coast of Brazil (on the left of the screen, whereas the first translator generically located it on the right). This illustrates the need for encyclopedic resources.

Even if some systems exist to describe SL in graphic forms (such as SignWriting, or HamNoSys), they do not allow to treat more than isolated signs, and are not widely used by the deaf community. SL does not have a formal written form as there is for English or French. Videos are currently the best way to keep traces of SL**,** but they are not as easily editable or searchable as text is. Contrary to text-to-text translation where the translation is built incrementally, the text-to-sign translation is prepared, then filmed in one shot.

Also, an important part of SL translation is the ordering of the various discourse components. In SL, the context of an event is always signed first and the event action second, which generally forces to set up the scene before signing anything happening in it. A typical order observed for news items such as the example above is: date, place, characters, object, action. In our example, place is the first information to be signed, but it is also the first information given in the source text. However, the "reported by official sources" part, which is the last information given in the source text, can be found near the beginning in the signed translation. This illustrates the work done on the order to translate to SL**.**

We can also observe that signers 1 and 2 seem to be adding context: it is the "first attack" part. Nothing in the source text states about the first attack, it only states about the second one, that implies the existence of a first one. But to make it clearer, those two translators set-up the entire context and scene as described below the figures. This is the way they chose to translate "nouvelle vague", or in English, "a second wave". Like their spoken language counterparts, SL translators are subject to deliberate decision making, and need to access resources to find which way to translate suits best the given situation.

3 State of the art

3.1 Sign Language

Although rather recent, some studies are interested in SL translation, but only for full automation, focusing on capturing SL or generating it. The first one is about sign recognition and SL translation based on sensory gloves. Bukhari *et al.* (2016) proposed a sign-to-speech translation system for Android phones, based on the use of their one sensory glove. Apart from being an invasive and constraining system to use, it only focuses on the manual articulators without taking facial expressions into account. Also, the system has been trained with only 20 signs, thus focusing on lexical items. Sensory gloves systems work best with finger spelling recognition, but cannot be identified as translation systems because SL does not come down to spelling.

Text-to-sign translation studies are often paired with avatar technologies issues, as in Halawani's Arabic SL Translation system for mobile devices study. It combines a text input, for which corresponding signs are then animated by an avatar, based on a database of motion captured signs. Even if the avatar technology is getting better, such text-to-sign translation systems still focus on lexical elements, with a strict grammar and barely no iconic output.

SL resources being rare resources, Barberis *et al.* (2010) and Bertoldi *et al*. (2010) both proposed works based on Italian Sign Language. The first one is about MT for Italian Sign Language, from text to animated avatar. They mention statistical translators such as MOSES which they trained for SL. The second work is about the creation of a parallel corpus between Italian and Italian Sign Language, within the ATLAS project. This corpus is meant to train a virtual interpreter rendered as an animated signer.

Yet, we did not find any mention of SL CAT software where the human user is involved. MT could be used in CAT, but we would like to focus on the entire translation environment rather on just the MT output. Our aim is to equip professional translators with suitable tools for their tasks. Let us look at what text-to-text CAT is made of.

3.2 Text-to-text CAT

Text-to-text CAT software is now more what we call Translation Environment Tools (TEnT). They are integrated suites of tools, and serve as a workstation for the translator. The previously mentioned tools can differ from one TEnT to another, but they also do include (and work around) a TM (Bowker Fisher).

TM is the great innovation brought by CAT software. First commercialized in the '90s, it has been broadly used since then. The translator is equipped with a database of all prior translations, which he can look up either to find a perfect match and insert it in its translation, which can be automated, or just to inspire new translations (which we can call concordancer look up). It has transformed the translator's job, in terms of productivity, cost and consistency of the translation output (Lagoudaki, 2006; O'Hagan, 2009). TM skills are now an entire part of the translation industry.

Koehn (2009) investigated a sentence-completion prediction system to assist human translators. The evaluation is based on user activity data, such as keystroke, deletions, cursor movements and timing. He found out that in most cases, CAT software does benefit the user. Translators do work faster, and achieve better translation quality using the offered assistance (quality is judged by target language experts). Only 2 out of the 10 translators he tested did not improve with the computer assistance he provided (post-editing machine translation, prediction, and options to select from a translation table). Even if the literature CAT evaluating benefits is short, the positive insights expressed by the professional, paired with those results seem to support the idea that CAT software does make the human translator more efficient.

4 Observations

As observed in practice as well as stated by the scarce studies cited above, CAT software relies on three key language or interface features. The first one is the use of text thus the need for a written form. Intended for text-to-text translation, the input data to translate is in written form. Whether the translator does so *ab nihilo* or using the integrated tools, he uses text. Likewise, the tools provide assistance in written form: glossaries, terminology assistants, and the TM gives access to written-only content. Each action thus implies keyboard editing: writing translation, modifications, or searching with the tools.

The second is what we called the "principle of linearity". CAT software automatically processes the source text in smaller units, the size of which can be parametrised. Paired with each smaller unit, which we call source segments, is an empty corresponding target segment. It is the translator's job to fill in those target segments with the proper translation. The source text is segmented, but the

order is not changed and cannot be changed in the integrated environment. It assumes that the concatenation of the translated segments leads to the translation of the concatenated source segments, in the same order. In our example, the original order is modified in all of the three translations. We demonstrated in a previous work (KaczmarekFilhol, 2019) that this is not true when it comes to SL. We filmed SL translators at work, and labelled their translation process into tasks. One of those tasks was segmenting and ordering the source text. We also observed that the translated result always presented information in an order different from the source one. Not only was it the only systematic task out of eight, but also the most time consuming (about a third of the process' total duration).

The third is the importance of the TM. It stores alignments (i.e. each source segment paired with the associated translated segment) and allows their reuse later. If the program matches a previously stored source segment, then the TM suggests the target segment as a possible translation. The translator is free to accept, decline, or accept with modifications. TM is also usually a collaborative tool, meaning it can be shared with colleagues, among a translation department, and even be provided by the clients. Generally speaking, it is a way to bring more consistency between the translations either in time or between the persons working on a same project, therefore gaining consistency. The alignments are automatically produced and stored along the process of translation.

5 Limits

Based on the previous sections, we saw that the two basics of CAT software interface are the written form requirement and the principle of linearity. These specificities make it difficult to reuse CAT interface in a straightforward manner. The entire interface has to be redesigned to overcome those two obstacles. It implies to think about a way to work around the lack of written form, as well as a way to display the target without constraining the order. Videos are not easily editable, so that only adapting the software to support videos will not be enough.

In addition, we talked about the importance of TM and the assistance it provides in the translation process. In SL, the target content in its final version takes the form of videos, which are not queryable. Storing short fragments of videos to edit them together is nonsense: the signers are not the same. The TM therefore cannot be reused in the same way it exists for text-to-text translation. However, as it is the main concept of CAT software, we should find a way to provide the same kind of assistance for SL**.** Such a tool should support video, store previous translations in the form of multiple alignments with various spans, and present the alignments.

Finally, encyclopedic content is necessary for SL translators to produce a higher quality translation. Encyclopedic content here includes maps, pictures, biographies or general knowledge. Current software does provide some assistance for lexical search (glossaries, terminology tools) yet nothing for any encyclopedic content. Hence, SL CAT software should be able to support with a larger range of searches, proposing more knowledge-search friendly tools.

The next paragraphs discuss each of the features in turn.

6 Interface solutions

6.1 No assumed linearity

Based on our observations, we have shown the non-application of the principle of linearity in SL. Plus, organizing is a task that uses time as we observed and filmed. Therefore, we should assist it. We first suggest that the SL CAT software should not automatically segment the source text. Instead, the translator can generate empty blocks, which are freely movable. The user is free to generate as many blocks as he wants to, create links between them or not, and organize them in any way he wants. As shown in the figure below, the blocks can work either vertically or horizontally. Each column depends on the previous one, but each line is independent from the

others. The blocks 2a' and 2a" can be switched without impacting the others. However, switching the blocks 2a and 2b would result in moving 2a' and 2a" a line under as well. It is a hierarchical structure.

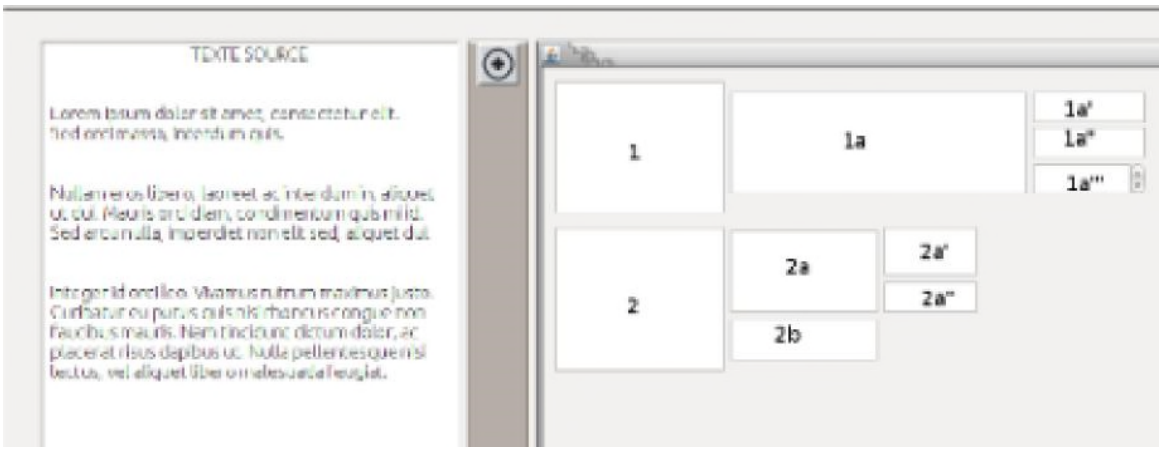

Figure 2: An example of blocks interface

6.2 No written form

To deal with the lack of written form, there should be no constraint on what kind of content the translator can fill the blocks with. For example, text notes if the user is comfortable with writing, input from keyboard or selected from the source text. Pictures and maps could be uploaded either from files on the computer or directly from the web. Drawings could be either uploaded or generated in the software using provided tool. Also, videos have to be supported, either uploaded as well or directly filmed with a video tool included.

Each block is seen as a thinking space, where the user can keep ideas about his translation in any form he wants. Those five propositions come from both observed common practices and needs expressed by the professionals we work with.

These blocks are not parts of the final translation, but serve more like a detailed outline. When done with his preparation, the user has to produce a video to deliver the translation. The blocks can be used as a prompter while the translator is filming himself for the final product.

7 Tools

7.1 Translation Memory

An alignment is a pair composed of two text segments, where one is the translation of the other. In our particular case, one of those segments has to be a portion of a video, identified using appropriate time tags.

Our suggestion here is to adapt a concordancer to SL to serve as a TM. Concordancers provide, upon query, examples of expressions translated in context. It relies on a database filled with alignments which in our case can be previous translations. In text-to-text translation, the alignments are created and stored automatically along the process. With SL, it requires the translator's direct involvement to create them manually. Once the user has filmed the final translation, he can associate selected segments of the source text with its SL translation, by marking the corresponding time tags from the video. Such alignments are stored as a database which is built by the users themselves. The entire source text is paired with the final translation video to create at least one alignment per translation. If the expression queried has been aligned, the concordancer answers with the smallest span found. If it has not been aligned but still

previously translated, the concordancer answers with the video of the entire source text that contains the query. If it has never been translated before, the concordancer cannot answer the query.

We needed a small database to start our testing with the concordancer, but aligned SL resources are rare, so we chose to build one ourselves. We used a French–LSF parallel corpus of forty short news texts, each translated in SL by three professional translators, and filmed, resulting in 120 videos of an average 30-second duration. We manually aligned segments of various spans. The source text was processed using the Brat software, from which we extracted data as so:

```
T22     Alignement 1199 1263    une deuxième vague d'attaques du crime organisé contre la police
T23     Alignement 1270 1322    un bilan de 52 morts dont 35 policiers en deux jours
T24     Alignement 1326 1376    la multiplication des mutineries dans les prisons,
```

Figure 3: An example data from Brat

Each source text is given an identification code, and the segments are identified using the number of their first letter and their length.

The videos were processed using the Elan software. Each translated segment was identified between the corresponding time tags using the annotation feature. We extracted data as so:

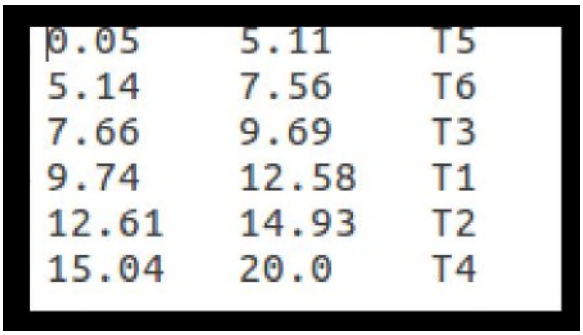

Figure 4: An example data from Elan

Each video is given an identification code, and the segments are identified with the time at which they start and their duration.

The user's query can be directly input from the keyboard, or a selected portion of the text he is working on. The expression is then compared to the source texts in the database, and upon match found, the results are expressed as follows: TxtID Starting position Length VidID Starting time Duration Opt.Tag1 Opt.Tag2

TxtID is the identification code of the text, starting position is the position of the first letter of the segment in the source text. VidID is the identification code for the video, starting time is the time tag corresponding to the beginning of the segment's translation and Tag 1 2 are specific information about the alignment. This information is added by the user who aligns, and should be characteristics to serve as search filter. For example, it could be the name of the aligner, the name of the signer, the topic of the translation, its duration etc. At the moment, we have not implemented query filters yet, but we plan to so that the results brought could be better suited to the user's expectations.

The matching results are displayed as they would be in a text-to-text bilingual concordancer, except that it includes videos. It could be displayed like the following:

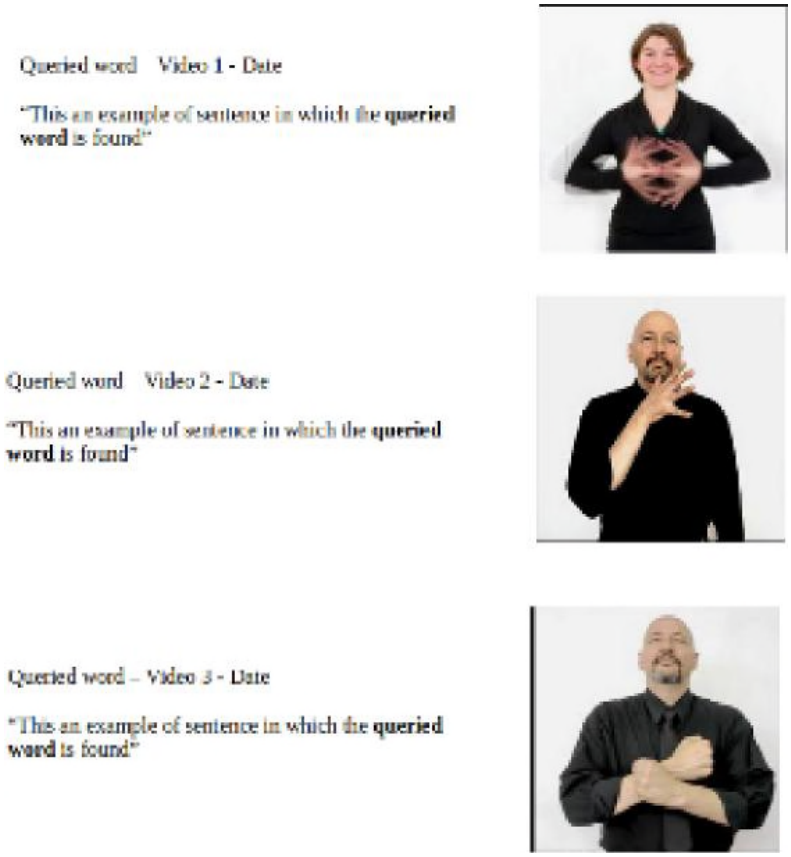

Figure 5: An example of display for the concordancer

7.2 **Encyclopedic assistance**

To provide the right encyclopedic assistance, we organized a brainstorming session with interpreters during a previous study (Kaczmarek-Filhol 2019). They were asked about their everyday practices, and to express their insights concerning their jobs. We ended up with a list of expressed needs and common problems encountered, which we did compare with the observations made in the filmed translation sessions. We identified 5 types of search. Two of them are already treated by text-to-text Cat software: lexical search, and definition look-up.

From our SL perspective, lexical search will mean soliciting various resources in order to find the adequate sign for concepts, including place names or proper names. Definition look-up will mean searching for definitions of source words or concepts which are not clear to the translator's mind. It also may help to find a way of signing it if no sign is known or found.

The other 3 types of search identified are, as shown earlier in this paper, more specific to SL. It concerns map search which means, when required to depict relative geographical locations or to sign a place when no specific sign is known or found, translators search for maps and plans. In a similar way, we identified picture search: to identify protagonists cited in the source text or to find a suitable periphrasis. Also to describe things that needs to be, or to assign a pertinent temporary sign to someone or something for the duration of the translation. And last
but not least, encyclopedic look up for context, which is observed when the source text refers to previous events or links between people that are unknown to the translators, they collect background about it.

Back to our example: Looking up if Sao Paulo has a defined sign would be lexical search, as well as searching a sign for mutiny. Looking up where Sao Paulo is however, is considered map search. If "organized crime" is not a clear notion for the translator, he may need to search for a definition. He may also need to learn more about the implied first attack to spatially organize his way of signing. These tasks can easily be assisted with Natural Language Processing features. Dealing with named entities such as proper names or locations, they can be detected in the source text. Those highlighted entities could then serve as an input into the encyclopedic search tool. Currently, translators do those search tasks using diverse website, opened in multiple tabs in a single web browser. To gain in comfort and readability, a first step would be to aggregate the results from all these searches in a single window within the software, or in different windows regarding the type of search.

8 Evaluation

A first prototype following the suggestions made above is currently under development. At the moment, it includes a global interface based on the blocks interface as well as a first try of SL concordancer. It will be available online for professionals to test. As it must be based on their needs and insights, we plan on an iterative process with the users to converge on the most adequate features and interface.

To evaluate our work, and how it benefits the users, we will measure four variables. The first one is obviously time, for the software should shorten the time spent on one translation. The second one will be comfort, as it is shown for text-to-text translators, does it affect SL translation the same way? The third variable will be confidence. Do they feel more confident about their translations when assisted by the software? And the last will be consistency, either between the multiple translations of a same translator, or in a group of translators.

9 Conclusion

This being the first work concerning CAT for SL, we hope to draw interest to the particular case of SL translation. Even if some SLs are legally recognized as official languages in their respective countries, most of them still have no status at all. Encouraging and easing SL translation is a first step to make SL more visible, and hopefully enhance accessibility on a more global scale.

References

Davide Barberis, Nicola Garazzino, Paolo Prinetto, Gabriele Tiotto, Alessandro Savino, Umar Shoaib, Nadeem Ahmad 2011. Language resources for Computer Assisted Translation from Italian to Italian Sign Language of Deaf people In *Proceedings of the AEGIS Workshop and International Conference,*

Nicola Bertoldi et all. 2010. On the creation and the annotation of a large-scale Italian-LIS parallel corpus *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, p.* 19-22.

Lynne Bowker Desmond Fisher 2010. Computer-aided Translation *Handbook of translation studies, Volume 1.* p.60-65

JanFizza Bukhari, Maryam Rehman, Saman Ishtiaq Malik, Awais M. Kambohand, Ahmad Salman 2016 American Sign Language Translationthrough Sensory Glove; SignSpeak *International Journal of u-and e-Service, Science and TechnologyVol.8, No.1 (2015), pp.131-142*

Cuxac C. et Sallandre M. 2007. Iconicity and arbitrariness in French Sign Language: Highly iconic structures, degenerated iconicity and diagrammatic iconicity *Empirical approaches to language typology, 36. p.13*

Dr.Sami M.Halawani 2008. Arabic Sign Language Translation System On Mobile Devices *International Journal of Computer Science and Network Security, VOL.8 No.1, January 2008.*

Kaczmarek M. Filhol M. 2019. Computer-assisted Sign Language translation: a study of translators' practice to specify CAT software *6th Workshop on Sign Language Translation and Avatar Technology Hamburg, Germany, Sept 29, 2019.*

Koehn P. 2009. A Process Study of Computed Aided Translation *Kluwer Academic Publisher.*

Lagoudaki E. 2006. Translation Memories Survey *Translation Memory systems: Enlightening users' perspective. Imperial College London.*

O'Hagan M. 2009. Computer-aided Translation (CAT) *In Baker, Mona/Saldanha, Gabriela (eds.), Routledge Encyclopedia of Translation Studies. London and New York: Routledge,.* p.48-51