



HAL
open science

Computer-assisted Sign Language translation: a study of translators' practice to specify CAT software

Marion Kaczmarek, Michael Filhol

► To cite this version:

Marion Kaczmarek, Michael Filhol. Computer-assisted Sign Language translation: a study of translators' practice to specify CAT software. Workshop on Sign Language Translation and Avatar Technology (SLTAT), Sep 2019, Hamburg, Germany. hal-02923914

HAL Id: hal-02923914

<https://hal.science/hal-02923914>

Submitted on 27 Nov 2020

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

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

Computer-assisted Sign Language translation: a study of translators' practice to specify CAT software

Marion Kaczmarek
Michael Filhol
marion.kaczmarek@limsi.fr
michael.filhol@limsi.fr
LIMSI
Orsay, France

ABSTRACT

Sign Language professional translators, unlike their text-to-text counterparts, are not equipped with computer-assisted translation software. No prior study has been conducted on this topic, and we aim at specifying such a software. To do so, we based our study on the professionals Sign Language translators' practices and needs. The aim of this paper is to identify the necessary steps in the text-to-sign translation process. By filming and interviewing professionals, we build a list of tasks and see if they are systematic and performed in a definite order. Finally, we reflect on how CAT tools could assist those tasks, and how to adapt their interface to Sign Language.

KEYWORDS

Sign Language, translation, CAT, computer-assisted translation

1 INTRODUCTION

Text-to-text translation tasks have been assisted by software for the past decades. Computer-assisted translation (CAT) offers tools that allow translators to gain time and efficiency (Koehn, 2009 [2]) by automating repetitive tasks and storing prior work in order to reuse it later. They can support various written languages but none of them are adapted for Sign Language translation. CAT relies on two fundamental things. The first is an editable written form for both source and target languages, which does not exist for Sign Language. The second is what we call the *principle of linearity*, which assumes that the concatenation of the translated segments will result in the translation of the concatenated source segments. A problem is that it has not been established for Sign Language, and we even produce evidence against it in this paper.

There has been no work concerning Sign Language CAT software, despite the growing need for translated Sign Language content and the very low number of available translators. Before asking ourselves how to adapt present CAT software to Sign Language translation, the first step is to identify which tasks are involved in the text-to-sign translation process. Next, we can think about the way in which those tasks can be assisted and/or automated, by discussing with involved professionals and reviewing the needs we collected from them.

To do so, we worked with two groups of professional Sign Language translators. We aimed for both objective and subjective data because people can struggle to verbalize their activities in detail, and tend to omit things they actually do. On the contrary, when asked, they can express things that we do not observe in their everyday practice. The two groups are disjoint, meaning that each



Figure 1: A still of the set-up, showing one of the groups working.

participant only took part in one of the experiments. In the first one, we collected objective data by assigning a translation task to the translators, which they execute without further interaction with us. In the second experiment, we collected subjective data by collecting participants' insights: they verbalize needs, practices and problems without being assigned any task.

The aim of the two studies we present is to identify the steps of the translation process, if they are systematic or not, and also if they are performed in a certain order.

2 FIRST EXPERIMENT: OBSERVING TRANSLATORS AT WORK

The first part of this work concerns objective data. To analyse translators' actions, we filmed them in their work in their working place, from reading text for the first time to the delivery of a translated result.

2.1 Methodology

The translation exercise was performed in pairs, each composed of one deaf professional Sign Language translator and one hearing professional Sign Language interpreter, both working on the same text together. This is the usual company's choice, though the participants were not yet used to it. The benefit for us is that the problems they would encounter would be discussed, and issues verbalized, which allows us to annotate the process in tasks. By task, we mean a specific action taken in the translation process. We organized the filming in sessions, each time submitting a selection

of journalistic texts to translate. The first session was a break-in session with three short three-line news items, dealt with one by one, so that the participants would get used to the set-up. The second session assigned twenty items of the same kind, and allowed the participants to organize their work freely. Finally, the third and last session proposed three long texts (about half a page). The recordings took place over five days, and we collected a total of 16 hours of footage, with one camera view on each translator's table and computer. When they were ready to deliver a translation, we filmed the result with the same cameras. They could split the performance in several takes, but no effort was put later in editing since the final result was not our focus.

We annotated this data by labelling tasks on a time line, for each text of each pair. We then compared the annotations between pairs and between texts, to detect repeated and shared practices, and a total or partial order constraining the tasks. At the time of writing this abstract, three texts for both groups, or three hours of video, are annotated.

2.2 Results

This presentation will report on our annotations and observations. We identified the following tasks :

- lexical search: soliciting various resources in order to find the adequate sign for concepts, including place names or proper names.
- discussing signs: for concepts with slight meaning differences influenced by context, the translators often discuss which signs suit the best the situation.
- map search: when it is 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.
- definition look-up: 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.
- encyclopedic look-up for context: when the source text refers to previous events or links between people that are unknown to the translators, they collect background about it.
- 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.
- segmenting the text and ordering segments: this includes dividing the source text in smaller units, arranging them in an appropriate order for Sign Language, one translator's rehearsing the result and discussing with the other to rework the discourse organisation.
- memorizing: the translator signs for himself, without interacting or even looking at his colleague. It is a very personal and informal production of the intended result, repeated as many times as needed to be completely memorised.

These tasks can be completed in a different order, with different resources and during various periods of time depending on the text or the pair. But most importantly, they remain the same. In our presentation, we provide examples of each, and further explain how CAT software could be useful in those cases.

Besides, the task of segmenting and ordering (bullet 7) is observed in every one of the six analysed translations. This shows that translators always spend time to work out an order for the contents of their translation. The delivered signed results also show that this order is always different to that of the source texts. This is evidence that the *principle of linearity* is neither assumed when translating nor verified in the resulting translations. This is a major challenge in adapting the typical CAT interface to Sign. We will present specifications to address this problem, along with more detailed results.

3 SECOND EXPERIMENT : BRAINSTORMING WITH SIGN LANGUAGE INTERPRETERS

The second part of the work concerns Sign Language interpreters for subjective data. We consulted them for two reasons. On the one hand, there are very few professional Sign Language translators in France, and interpreters are often asked to act as translators. On the other hand, before performing planned major interpretations, they ask for written notes to prepare, and therefore process a source text as a translator would.

3.1 Methodology

We organised a brainstorm session to verbalize common practices regarding translation as well as preparation, commonly encountered problems, current solutions, and concrete needs. The brainstorming session lasted two hours and a half, between six interpreters who work for the same interpretation department. There were six interpreters, with a range in professional experience. The session was under the authors' supervision, one taking notes and the other one moderating the discussions, asking open and general questions about their practice. The main questions were displayed on slides so that everyone could refer to them when needed. Participants were given sticky notes and pens to write down ideas about what tools could a computer software bring to them, without previous knowledge of it. Those notes were to be discussed later, after a short presentation to show them what current CAT is all about. The brainstorming session ended by a free discussion on the topic, where everyone was able to bring more ideas, express views and problems that did not come to mind earlier, and debate with each other.

3.2 Results

Pictures were often mentioned, and are used as a base reference to create signs when finding a periphrasis or a suitable description is necessary (for example in very specific fields, companies' jargon, Harvey, 2002 [1]). The most cited problems are the lack of Sign Language resources, or at least the time spent to look for them, which is related to the lack of sharing. They bring ideas of concordancer, glossaries or the need to access other colleagues prior work to find inspiration. If they do not succeed in their search, they tend to draw inspiration from pictures, dictionary definitions, encyclopedias and video examples when available. When asked what kind of assistance CAT software could provide, the need of free segmentation is cited, along with video tools to record and edit, drawing tools, search engines for encyclopedic knowledge, automated source text

information extraction (named entities, numbers, dates etc.) and translation memory.

4 DISCUSSION

The results gathered in both experiments are consistent and coincide for most parts. But the two of them also brought specific results. The first one identified practices that were not referenced in the second one, such as map search, encyclopedic search for context, rehearsing and memorizing the translation. On the other hand, the second one brought light on problems we did not clearly identify in the observations: the lack of shared Sign Language resources, the need to access prior work, whether done by others or by oneself.

This particular last point shows interest in a Sign equivalent to translation memory, which text-to-text translators make heavy use of. Integrating such a feature in a CAT software for Sign Language raises a new challenge, as without the *principle of linearity*, the text

and Sign segments do not match and cannot be aligned automatically like they are in text-to-text translation. Alignment may need the translators' direct involvement, which is a new constraint on the interface design.

In this proposal, we address the problems mentioned in this paper by specifying pieces of interface for a Sign Language CAT software tool. In a near future, we wish to develop and test prototypes with the professionals, in an iterative process to converge to the most suitable tool for them. With our work, we hope to draw more interest on this particular topic, and foster Sign Language translation practice.

REFERENCES

- [1] Malcolm Harvey. 2002. Traduire l'intraduisible. *ILCEA (online)* 3 (2002), 39–49. <https://doi.org/10.4000/ilcea.790>
- [2] Philipp Koehn. 2009. A Process Study of Computed Aided Translation. *Machin Translation* 23, 4 (2009), 241–263. <https://doi.org/10.1007/s10590-010-9076-3>