Live Subtitling through Respeaking: a new Discipline in Interpretation?

Rita Geens
University College Ghent

Abstract
Over the last few decade(s) the world of interpreting has been changing rapidly; one of the reasons for this is the emergence of new disciplines. One of the new and expanding fields is that of live subtitling through respeaking, a discipline related to interpretation in various ways.

Live subtitling is crucial for the deaf and hard of hearing, offering them an essential complement to classical subtitling, which can only be realized when a script of the recording is available. Thanks to live subtitling, the hearing impaired are no longer denied full access to live programmes.

VRT, the Flemish public channel, relies on voice recognition to produce live subtitles, which are mainly applied to news reports and major (sports) events. A respeaker rephrases the source text into subtitles, i.e. in short, preferably simple sentence structures.

The first live subtitlers were trained on the spot. The obvious affinity between live subtitling and simultaneous interpreting into a foreign language gave rise to the idea of teaching the live subtitlers a crash course in simultaneous translation. It is my involvement in the delivery of this course that induced me to write this paper. It became clear to me that live subtitling skills are easily acquired by interpreters. Since live subtitling is a rapidly expanding field offering job opportunities, it seems appropriate to train (future) interpreters in live subtitling through respeaking.

From intertitle to live subtitle

Subtitling has come a long way. In the beginning there was the intertitle, used in silent movies to narrate story points and present key dialogue. Then came the soundtrack and thanks to translation, movies and television, programmes could and can travel around the world. The most prevalent methods used to make foreign language television programmes available to a domestic market are dubbing and, of course, subtitling. Subtitling was initiated, among others, in Belgium, the Netherlands and Great Britain.

We are all familiar with prepared subtitling. The translator receives a recording and, for example, a dialogue list of a soap, a drama series. After spotting precise in and out time codes he prepares a list of subtitles correspondingly, mostly in a different language. In other words, prepared subtitling is often an interlingual activity.
In the case of live or real-time subtitling there is no pre-recording or transcript; there is only the oral message of the live interview or live presentation that should be rendered in subtitles immediately. So far live subtitling is basically intralingual. The Flemish public channel VRT and the commercial broadcaster VTM use live subtitling for instance to subtitle live interventions during news broadcasts, live sports events etc.

The decision to rely on speech recognition for live subtitling was made out of necessity.

Over the last few years VRT has tried different techniques to produce live subtitles, all based on keyboards. The first one was the **querty keyboard**; introduced in 1984 it lasted until the beginning of 2001. The subtitler listened to the live comment and tried to type the spoken words as fast as possible, after which the subtitle was broadcast. Here accuracy was high, but speed was far from sufficient. A typing speed of 90 to 100 words/minute is far below the speaking rate, so the delay was inevitably huge.

The second transcription technique tried at VRT was based on the **Velotype keyboard**, a Dutch invention. It attains a higher speed (120 words per minute) producing syllables instead of characters. However, training takes a long time and therefore the system was too expensive for a small broadcasting company such as VRT. Nevertheless, Velotype is still used in some countries, e.g. in the Netherlands and in Sweden. The third method used, was the **stenograph keyboard**, traditionally associated with court reporting. It was designed for verbatim transcription in court rooms. It is a fast and accurate system: the stenographer can produce up to 225 words per minute and the system claims a 98% recognition accuracy. However, the use of stenography was as problematic for VRT as that of Velotype, and for the same reason: broadcast stenography requires a high degree of skill - it takes two or three years of training and experience - so stenographers are in short supply. Steno keyboards are still being used for at least some of the live subtitling in English-speaking countries, where and when there is a ready supply of keyboard operators, which is not the case in Flanders.

Since the keyboard methods proved either too slow, or too costly, VRT had to search for a different solution, especially because of the increasing demand for live subtitling not only in Flanders and in Belgium, but also in other European countries.
This is to be explained by the fact that virtually all TV programmes broadcast by public stations in the EU must eventually be made accessible for the hearing impaired, since partial or complete loss of hearing is a condition that affects more than 83 million people in the EU and, given the aging of the European population, this problem will continue to grow. Moreover the Belgian government has ruled that VRT is to subtitle nearly 100% (i.e.95%) of its programmes by 2010.

It is worth mentioning that subtitling is not restricted to the deaf and hard of hearing, although they do account for some 70% of the users in Flanders. Subtitling is also increasingly used as a tool for foreign language acquisition by migrants, who want to learn Dutch. Furthermore, viewers increasingly rely on it to avoid disturbing family members engaged in activities other than watching television. All these developments reinforce the demand for more and more subtitling.

In 2001, the end of the keyboard era at VRT, some 50% of the VRT programmes were subtitled via Teletext, the Flemish counterpart of Ceefax. Raising the current 65 % of subtitled programs to the required 95 % by 2010 is a real challenge, especially because part of the extra 30% are live TV broadcasts.

VRT’s search for a system that was fit for purpose and dependable enough to subtitle live programs in Dutch in an acceptable way, would eventually lead to Dragon Naturally Speaking, which is a dictation system based on speech recognition in Dutch. Its capacity of 250 words per minute enables it to keep up with a normal speaking rate. At the moment Dragon’s Fifth version, Dragon Naturally Speaking 5, has a recognition rate of 95 or 96 %, which means that an error was produced roughly once in every two subtitles.

In the first year of its introduction the technique of live subtitling was used for off screen live sports commentary, and since 2002 it has also been introduced in the news programmes and for special events, such as Election Broadcasts and Eurosong for children. In this way, VRT became the first broadcasting company in Europe to produce live subtitling through respeaking.

Dragon Naturally Speaking is a speech-to-text system, a system that converts spoken language into written text. In fact, such dictating systems are designed to be used by doctors, lawyers and journalists etc., who dictate their texts instead of writing them or having them
written. In other words, they are not primarily designed to be used for live subtitling.

**Live subtitling**

Since live subtitling is a new technique, it is not always understood how it operates. It is a common mistake to assume that live subtitles are produced by direct recognition from the TV soundtrack. Unfortunately automatic speech to text conversion that is accurate and speaker-independent is still far off. In order to enhance the accuracy of the dictation, the system is first trained to recognize a new voice. During that initial training session a new user reads out a few texts allowing the system to get acquainted with the speaker’s voice and speech characteristics which will help to reduce the number of recognition mistakes.

A second misunderstanding is the assumption that the live subtitler merely repeats the live text literally. In fact he is often obliged to adapt, to respeak the message. He then has to revise the written subtitle, i.e. he reads and – if necessary – corrects it and finally he has to transmit it.

The first step, the respeaking, is necessary for different reasons:

- the available vocabulary
- the need for reduction
- the need for segmentation
- the need for correction.

*The available vocabulary*

As regards the vocabulary available, Dragon has a vocabulary list of some 230,000 words. This implies that this list is far from complete. Technical terms and proper names in particular are often not included. Therefore, when the speaker uses words that are not listed in the vocabulary, the respeaker has to try to avoid using them, as the system will not recognize them and therefore offer the nearest match. Consequently, the previous version of Dragon Naturally Speaking, Dragon 9, turned Condoleezza Rice’s name into *kon de liefste lijst* (translated literally this would be *could the most beloved list*).

If the respeaker knows in advance or suspects that the journalist will use a certain term or name, he can add it to the vocabulary list. Dragon’s own vocabulary list can be complemented with up to 56,000 items.
VRT has developed topic-specific vocabularies for sport events and news programmes. These lexicons are updated on a nearly daily basis, especially for the news casts. But still, when a respeaker comes across a term that he fears is not listed in the lexicon; he cannot afford the hazard of misrecognition. Instead, he will try to find a quick solution by referring to Condoleezza Rice as the Secretary of State.

**Reduction**
Reduction of the original message is necessary for two reasons: to adapt the message to the reading speed of the viewers, and to avoid delay. The average, very regular speaking rate ranges between 180 to 240 words per minute, whereas the average reading speed has been proven to range between 150 and 180 words per minute. If the subtitlers were merely to repeat everything that is being said, the viewers would not be able to keep up.

A second reason why the respeaker has to reduce the message is that he has to reduce the time lag, for delay is inevitable. The respeaker has to listen to the message, to rephrase it and to check the correctness of the written output before the subtitle is broadcast. Recognition may go wrong for different reasons. The phoneme analysis may malfunction with homophones or the subtle distinction between short or long vowels. Moreover, the system sometimes fails to distinguish separate words or compounds and it is not familiar with assimilation rules either. The recognition rate of 95 to 96% of the first Dragon version used at VRT implied, as pointed out before, that statistically an error was produced roughly once in every two subtitles.

Manual keyboard correction resulted in unacceptable delays of 15 seconds or more for live subtitling, which understandably gave rise to quite a lot of complaints from the deaf and hard of hearing. In fact, the subtitles need to be displayed within 2-3 seconds after the audio they represent. Since then, recognition rate has improved. The present Dragon Naturally Speaking version 10 claims the recognition rate is up to 98%. Yet the written output still requires correction, and in order to make up for the time lag due to potential corrections and transmission, the respeaker tries to render the source text in the shortest possible form.

**Segmentation**
The respeaker also has to segment the original message, i.e. he has to split longer source sentences into smaller units, preserving the
logical connections between them. Prepared subtitling aims at two-
line subtitles that do not exceed 64 characters and spaces. The
respeaker, of course, cannot count the number of characters, but as a
rule he tries to avoid complex sentences that require reading on
consecutive screens.

**Correction**

Spoken language as in live interviews is known to contain
grammatical and lexical mistakes on a regular basis. Obviously the
live subtitler is expected to rephrase the imperfect spoken message
into (grammatically and lexically) correct subtitles.

Taking into account the complexity of live subtitling, it will be clear
that these activities cannot always be performed by only one person.
In the case of weather forecasts and live events, such as tennis for
instance, it is possible for one person to cover the live event on his
own.

More difficult live programmes such as interviews and especially
debates with up to six people involved require one or even two
subtitle respeakers and a corrector/transmitter. In that case the
respeaker does not operate in the transmission mode, which enables
him to concentrate fully on his live comment. The correction of the
inevitable mistakes and the transmission are dealt with by the
corrector/transmitter on the spot.

**Respeaking and interpreting**

It will be clear that the skills involved in live subtitling are often closely
linked to those of a conference interpreter:

- Both the live subtitler and the simultaneous interpreter rely on
similar techniques including: *dual tasking* (i.e. listening and
speaking simultaneously) and segmentation (although the
interpreter will not use segmentation systematically).
- Both respeak a message, i.e. they render the contents of an
incoming message in a different form.
- Both are working live.
- Both are expected to restrict the time lag as much as possible.
- Both are used to rephrasing grammatically incorrect source
language correctly.

These similarities between live subtitling and interpreting induced the
public channel VRT to have the live subtitlers trained in interpreting
techniques, a task which they entrusted to me. For obvious reasons it seemed useful to make an inventory of means and devices that can be used systematically for the reduction and segmentation of sentences. Then the subtitlers were trained to use these means and devices live.

At the same time, it is quite obvious that an interpreter can easily learn live subtitling, since he masters most skills that are required. Therefore, it seems quite logical to offer interpreters some training in live subtitling techniques, which is nowadays the case in Belgium.

As from the academic year 2007-2008, the Faculty of Translation Studies, University College Ghent has introduced a course on Audiovisual Translation for Interpreters, including a section on Live Subtitling. The Live Subtitling section is meant to familiarize the students with this technique and to teach them how to rephrase by reducing and segmenting sentences.

Training the interpreter into a subtitler

When studying reduction the student is invited to use a form of “short speak” relying systematically on substitution and omission. Substitution implies rewording in a shorter form, whereas omission involves deleting redundant information. These techniques can be applied in quite a number of cases, but we will confine ourselves to some examples that were easy to translate from Dutch into English. Omission is possible when the audience can be supposed to have prior knowledge. When the speaker mentions e.g. the international Red Cross organisation, the respeaker can make do with the Red Cross, since most television viewers know that the Red Cross is an international organisation.

Omission is e.g. also possible when a speaker uses a redundant sentence opener - as speakers often do, in order to gain time to reflect upon their answer.

It is clear that we should put an end to all speculation concerning the possibility…

can be reduced to:

We should put an end to all speculation concerning the possibility …
In dialogues, very often the beginning of the answer contains part of the question, which makes it possible to reduce the answer. In the following example, the journalist in the studio asks the following question:

Is the meeting still going on, Stef?

And the reporter on the spot who is attending the meeting answers:

Yes Martine, the meeting is still going on

The Dutch respeaker could easily reduce this answer to:

Yes, Martine.

The interpreter will not only be trained to spot these and other forms of redundancy, he will also be trained in substitution. In other words, he will be trained in recognizing the words and word groups that can be substituted systematically by a shorter equivalent. Whenever possible, the respeaker should try and replace a word or group of words by a shorter equivalent.

Attributive adjuncts can, for instance, be quite long, but often a shorter equivalent is available; relative clauses and prepositional adjuncts can often be replaced by adjectives.

The damage that was caused to the environment and the government of Great Britain can be replaced respectively by:

Environmental damage and the British government

The respeaker tries to use one-word prepositions, conjunctions, verbs, etc. The speaker may for instance use the conjunctions *in order to* and *despite the fact that*, while the respeaker will use *to* and *(a)though* and if the verbal periphrasis *to put an end to* can be replaced by *to end*, the respeaker will do so. In this way source language sentences can often be reduced considerably:

It is clear that we should put an end to all speculation concerning this possibility.

*(15 words, 24 syllables 89 characters and spaces)*
We should end all speculation on this possibility.

(8 words, 10 syllables, 51 characters and spaces)

This reduction makes the message shorter, i.e. adapted to the viewers’ reading speed and initial delay can be reduced. The content of the message is preserved and even keywords such as *speculation* and *possibility* are retained. This makes the subtitles easier to read for people with residual hearing: it would be confusing for them not to recognize at least some of the words they hear; lip-synchrony for deaf people is not possible, but for them, too, it would be very confusing not to recognize at least some of the words they lip-read.

The students are also trained in segmentation, i.e. they are trained to create subtitles that can be easily understood in the short time they appear on the screen. As pointed out before, a respeaker tries to avoid complex sentences that require reading on consecutive screens. Therefore complex sentences are split into smaller ones, preserving the logical connections between them as in the following example:

Although he was quite ill during the last few years of his life, he was able to compose several important works.

He was quite ill during the last few years of his life.

Yet he was able to compose several important works.

**Conclusions**

Translation and interpreting are protean activities which change and adapt in response to market demands and technological possibilities. One of these possibilities is speech technology, which gave birth to live subtitling through speech recognition.

At the moment, live subtitling through respeaking is not a perfect system. The results are satisfactory, but some problems, such as misrecognition remain. However, thanks to research, the recognition rate and therefore the output improves with every new version of Dragon.

As regards live subtitling itself, research into this subject is still limited as it is a very young discipline. Research is being conducted into the phenomenon of *delay* at the Artesis University College in Antwerp, and at the University College of Ghent transformations for improving substitution, omission and segmentation are being
explored. The results of these activities may also lead to better subtitles.

As live subtitles are improving anyhow, live subtitling is becoming a rapidly expanding field, both in Belgium and abroad. Not only VRT, but also the BBC, as well as Dutch, French, Spanish, German companies are using it more and more.

Since live subtitling involves skills closely linked to interpreting, interpreters can easily be trained to become live subtitlers. In fact, VRT has started hiring interpreters for this job, and is satisfied with the results. Live subtitling may, therefore, be considered as an interesting new niche market for interpreters.

Bibliography


