

DEPARTMENT OF MANAGEMENT

**Live Subtitling with Speech Recognition
Causes and Consequences of Text Reduction**

**Bieke Luyckx, Tijds Delbeke,
Luuk Van Waes, Mariëlle Leijten & Aline Remael**

UNIVERSITY OF ANTWERP
Faculty of Applied Economics



Stadscampus
Prinsstraat 13, B.213
BE-2000 Antwerpen
Tel. +32 (0)3 265 40 32
Fax +32 (0)3 265 47 99
<http://www.ua.ac.be/tew>

FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF MANAGEMENT

Live Subtitling with Speech Recognition Causes and Consequences of Text Reduction

**Bieke Luyckx, Tijs Delbeke,
Luuk Van Waes, Mariëlle Leijten & Aline Remael**

RESEARCH PAPER 2010-010
MAY 2010

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.213
phone: (32) 3 265 40 32
fax: (32) 3 265 47 99
e-mail: joeri.nys@ua.ac.be

The papers can be also found at our website:
www.ua.ac.be/tew (research > working papers) &
www.repec.org/ (Research papers in economics - REPEC)

D/2010/1169/010

LIVE SUBTITLING WITH SPEECH RECOGNITION

CAUSES AND CONSEQUENCES OF TEXT REDUCTION

Bieke Luyckx^{1,2}, Tijs Delbeke^{1,2}, Luuk Van Waes¹, Mariëlle Leijten¹, Aline Remael²

University of Antwerp¹

Artesis University College Antwerp²

ABSTRACT

Speech technology has made it possible to use speech recognition for simultaneous subtitling of live television broadcasts via the technique of respeaking. Despite the considerable prior research into the quality of live subtitling using speech recognition, little research has focused on the quantitative aspects of subtitles. Although live subtitles are nearly always a reduced form of the spoken comments, the exact causes of text reduction are still largely unidentified.

This study aims at a better understanding of the causes and consequences of text reduction in a live subtitling context. Three excerpts of an infotainment talk show were subtitled by twelve respeakers of the Flemish public television. They were instructed to do this in three different reduction conditions. Various subtitle features, such as reduction percentages and delay, as well as measures of the respeakers' working memory were collected. Both a quantitative and qualitative analysis were carried out. In the quantitative analysis we opted for a multilevel analysis to take into account the hierarchical nature of the data. In the qualitative analysis, we discussed the effects of commonly used reduction strategies.

The results show that reduction is not a random process. In contrast, it is largely determined by a number of external factors, viz. delay, amount of source text and the proportion of 'full' reductions. There is a large amount of evidence suggesting that respeakers prefer to omit certain comments rather than reducing them to a certain extent.

It also appears that the decision to fully omit a comment seems not to be primarily based on the amount of input, while the decision to partially reduce is. Differences in the capacity of the working memory do not seem to affect text reduction as such. Finally, the qualitative analysis demonstrated that respeakers use a wide variety of strategies to reduce the spoken comments in order to limit the loss of information as much as possible.

KEYWORDS

Real-time subtitling, live subtitling, respeaking, voice-writing, speech recognition, keystroke logging, reduction.

1. Introduction

1.1 TV Accessibility

Since its invention in the late 1920s and in spite of all the technological novelties created ever since, television does not seem to have faded in popularity. It stands tall as one of the most widespread sources of information and entertainment. However, as television itself is evolving into a highly versatile and interactive medium, it is partly doing so in response to the growing demand of 'media for all' (Diaz Cintas, Orero, & Remael, 2007). Indeed, in a time where the

media play a crucial role in people's perception and knowledge of the world, media accessibility is a human right. With the current prominence of audiovisual texts, audiovisual translation (AVT), which allows such texts to travel across linguistic borders, is also taking the centre stage in Translation Studies (TS) and translation practice. Both traditional forms of AVT, some of which have been around since the beginning of the 20th century, and more recent ones, make audiovisual texts 'accessible' to a wider audience. In this article, we focus on such audiovisual texts made for television and provided with live teletext subtitling.

1.2 Intralingual subtitling and its target audience

Subtitling is an AVT mode that has been around since the early 20th century and that has, traditionally served as a means to make television programmes and films in one language accessible for target audiences with a different mother tongue (see e.g. Gottlieb, 1992; Ivarsson, 1992). This type of subtitling is still common practice on television in smaller language communities like The Netherlands, Flanders, Portugal, Greece and the Scandinavian countries (Gottlieb, 1992). Unlike bigger countries like France, Italy and Spain, they have no dubbing tradition, for historical, language-political reasons, as well as economic ones; subtitling is simply much cheaper than dubbing (Remael, 2004). However, the application of subtitling was quickly seen as a valuable means to meet the demands of television accessibility made by aurally impaired viewers, broadening the idea of subtitling from a translation mode in a strict sense to a means of transferring speech into written language. This led to the distinction between *interlingual* subtitling, in which source and target language differ, and intralingual subtitling, in which they are the same (Gottlieb, 1992; Díaz Cintas, 2005; Díaz Cintas & Remael, 2007). As it was argued that not everybody would profit from intralingual subtitles at all times, a technical framework was needed that could provide subtitles that people would not have to watch if they did not want to. That is when the BBC came up with Ceefax in the 1970s, commonly known as teletext in many other European countries. It offers a means to display the subtitles when desired, creating so-called 'closed subtitling', a type of decodable and optional subtitling, as opposed to 'open subtitling', which the viewer cannot switch off (Díaz Cintas & Remael 2007).

In the United States and Canada, the term *captioning* (whether open or closed) is used to refer explicitly to the method used for transferring aural information into text, for people with a hearing impairment. Subtitling is, then, associated only with a 'real' translation process in which source and target language differ. Unlike subtitles, captions, "...contain not only dialog, but sound effects, onomatopoeias, and other visual cues as well ..." (Robson, 2004, p.4). However, Neves (2008, p. 130) sees no real semantic difference between the two terms and finds "the debate [...] rather sterile and quite worthless". In fact, she may be right in saying that the two terms are mere geographical varieties for the same concept, since in practice no intrinsic difference can be shown between them. Both terms simply refer to the conversion of the audio track of a television programme or another audiovisual production into text that is displayed on a screen (Neves, 2008).

The assumed non-existence of a difference between subtitling and captioning finds additional support in some thoughts on the reception of closed, teletext-based¹, subtitling. Whereas historically it is clear that closed subtitling was intended to serve first of all the needs of viewers

¹ The decoding device for closed captions used in the US and Canada is called Line 21 and intrinsically differs from the teletext technology. Since the text output is comparable, though, we will not bother to go into technical details of either of the systems. For convenience sake, we will continue to refer to the technique that is applied for decoding and visualizing closed subtitling as 'teletext'.

with an aural impairment, the D/deaf² and the hearing impaired (a diverse audience in itself) nowadays no longer constitute the largest interest group. As is shown by the research project on television access services held by the UK Office of Communications (Ofcom), of the 7.5 million people who actively used the intralingual subtitles, “6 million did not have a hearing impairment” (Ofcom, 2006). A similar situation seems to be the case in most countries. Polls from the Flemish public television VRT, for example, indicate that over 70% of the people using closed captions hear normally (Doens, 2000). Non-native speakers, elderly people for whom the speech goes too fast, customers of noisy bars or parents who are aurally distracted by their boisterous children appear to switch on teletext subtitles quite regularly. In The Netherlands, 41% of the foreign viewers use teletext captions (SOAP, 2007). In India, intralingual subtitles have proven to be a valuable means to improve literacy in children (Kothari, 2008; Kothari, Pandi, & Chudgar, 2004).

Besides stressing the already heterogeneous character of the group of D/deaf and Hard-of-Hearing people (see also De Linde & Kay, 1999 and Neves, 2005 for a detailed discussion), Neves (2008) too acknowledges that the diversity of the public goes beyond just hearing impaired viewers. Still, she decides to use the term ‘*Subtitling for the Deaf and the Hearing impaired*’ (SDH) systematically to refer to ‘intralingual captioning’. Though defensible historically, this may not be the best way of addressing the issue in our current day situation in which we are aware of the obvious diversity of same language subtitling users.

1.3 Real-time or live subtitling

A few other major distinctions that are traditionally mentioned in the subtitling literature deserve our attention. First, a rather evident but highly important divide needs to be made between *real-time* or *live* subtitling and *pre-preparable* subtitling (Remael, 2007). Evident, since in fact it only concerns a difference in programme type (pre-recorded versus live or as-live³). Important, because it has obvious consequences for subtitling practice. From here on, we will refine our focus and deal exclusively with the subtitling of live and as-live programmes, as we wish to learn more about the specificity of this specific type of AVT.

A dichotomy that is virtually only relevant for live subtitling, is that between *block* and *scrolling/snake* mode (Remael, 2007). These terms describe two different ways in which the subtitles can be displayed on the screen. When (live) subtitles are presented on the screen word for word, we speak of scrolling mode or snake. Subtitles that appear as a whole (one to three lines at a time) are said to be presented in block mode.

Often, if not always, snake is attended with (near-)*verbatim* subtitling, when the subtitler tries to stick to the original dialogue as closely as possible, whereas *edited* subtitling, when dialogues are paraphrased rather than transcribed, usually uses block mode. The distinction between verbatim and edited subtitling will prove to be essential to our discussion, as it forms one of the main foci of our experiment.

² According to Neves (2005, p. 84) “*deaf* simply refers to someone who cannot hear well enough to process aural information conveniently. Considering somebody *Deaf* means accepting the fact that that person belongs to the Deaf community that, even if a minority, has rules and codes of conduct that differentiate it from all others”.

³ *As-live programs* are really recorded beforehand, but are subtitled as if they were live. This means that no preparation was possible and that the subtitler has to perform at the speed of the program. From now on, we will only use the terms ‘live’ or ‘real-time’ subtitling, thus addressing the real-time character of the subtitling act rather than that of the program itself.

1.4 Towards respeaking

Live subtitling has come a long way and has experimented with various techniques since its inception in the early 1980s (Lambourne, Hewitt, Lyon, & Warren, 2004). Still, none of these allowed subtitlers to reach the high standards that had to be met. Traditional keyboards obviously did not allow the subtitlers to keep up with the pace of the programme dialogues and the use of velotype, a custom-designed keyboard, requires lengthy training. Moreover, if the subtitlers pushed their speed performance to the limit, a considerable loss of quality was unavoidable. Significant improvements were made when stenography was used as a subtitling tool, both in terms of speed and accuracy (Robson, 2004). Since the BBC introduced speech recognition as a technique for live subtitling in 2001, however, this has become the standard real-time or live subtitling method in many countries. Speed and accuracy easily outbid that of keyboard and velotype writing. The big advantage over both stenography and velotyping, moreover, is that the training period is fairly limited. There seems to be a great lack of trained stenographers and velotypists simply because the technique is so hard to master entirely (Lambourne, Hewitt, Lyon, & Warren, 2004; Van der Veer, 2007).

A typical misconception about real-time subtitling and the use of speech recognition is that spoken television dialogues are input directly into a speech recognizer, which automatically transfers this speech into a subtitling program. This kind of *speaker-independent* speech recognition system, in which the software has not been trained to recognize the voice of a specific individual, is not yet accurate enough for subtitling purposes (Robson, 2004). Moreover, natural speech, redundant in nature, is not very suitable for direct use in subtitles without at least some degree of adaptation. Stop words, hesitations, false starts and the like, are not desirable in subtitling. Instead, a *speaker-dependent* system is used. Here, the voice writer⁴ has actively trained the speech recognition software by creating a profile, talking to the program repeatedly and building up a personal lexicon. The more the software is used and trained, the better the results are.

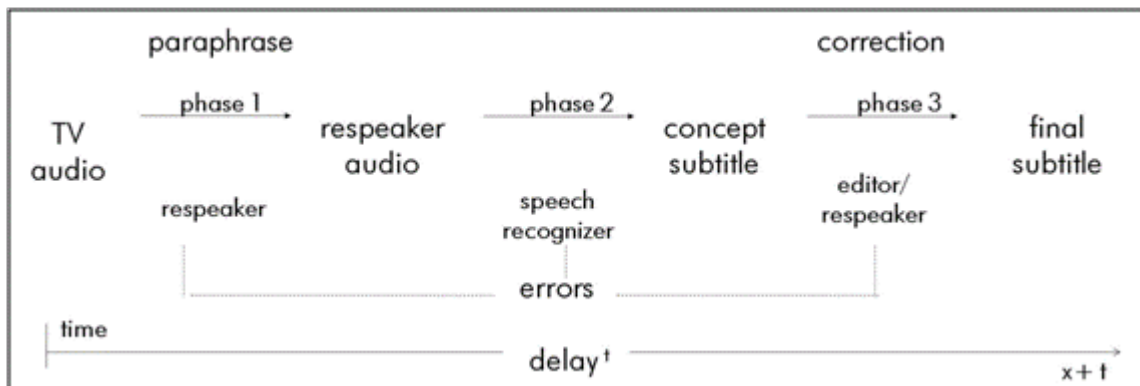


Figure 1. Temporal representation of the live subtitling process.

In an average live subtitling session, one person watches and listens to the television programme as it is broadcast live. Wearing a headset, he or she simultaneously repeats or paraphrases what is being said, an act that is known as *respeaking*. Figure 1 visualizes the basic

⁴ Our use of the term *voice writer* here merely refers to ‘somebody who uses speech recognition as a writing tool’. In Northern America, however, this term can also be used as a synonym for what is usually called *respeaker* in Europe, both indicating ‘a live subtitler using speech recognition’.

process of speech recognition-based live subtitling and shows where three major topics, referred to as error, delay, and reduction, manifest themselves.

The respeaker speaks directly to a speech recognizer, which results into a concept subtitle. Sometimes, errors that are made by the respeaker or by the speech recognizer itself are corrected before they are put on the air. This can be done either by the respeaker proper or by an additional editor who will quickly correct the output of the speech recognition program before the subtitles are broadcast.

Despite the advantages mentioned above (speed, accuracy and limited training period), the use of speech recognition for subtitling is by no means a flawless technique. Subtitles may contain errors, lag behind the original television dialogues, causing a 'delay', and viewers often complain that the subtitles do not sufficiently cover the aural information provided by the programme.

The occurrence of delay and errors is largely due to the speech recognition process itself. The rendering of information, however, is for the greater part controlled by the respeaker. Although the written version of the spoken comment is nearly always a reduced form, this does not inevitably mean that the quality of the text is affected. Uncompleted thoughts, interjections, and repetitions are just a few examples of sentence parts that carry no essential information, and that can therefore be deleted or summarized. A different word choice or emphasis, however, can quickly result in a change of meaning. In other words, text reduction can be problematic as well.

According to Diaz Cintas & Remael (2007), there are two types of text reduction: partial and total reduction. Partial reduction is achieved through condensation, a more concise rendering of the source text. Total reduction is achieved through deletion or omission of lexical terms. Very often both processes are combined, leading to the rewriting that is so typical of subtitling. No clear rules can be given as to when to condense and reformulate, or when to omit. Diaz Cintas & Remael (2007) suggest to: (a) eliminate what is not relevant for the comprehension of the message and/or (b) reformulate what is relevant in as concise a form as is possible or required. The reduction rate largely depends on the speech rate of the source text. Although there are no strict fixed rules for subtitling, a number of guidelines is assumed to be known by subtitlers (Diaz Cintas & Remael, 2007): (1) pay attention to the transition from oral to written mode, (2) formulate short, well-formed sentences, (3) pay attention to segmentation and lay-out, (4) match the subtitles with the visual information, (5) render the essence of the message within the given reading speed, leave out repetitions and features of orality and leave out or rewrite information of secondary importance, and (6) keep in mind what your target audience is expected to know. A detailed description of various common subtitling and reduction strategies can be found in *Audiovisual Translation: Subtitling* (Diaz Cintas & Remael, 2007).

1.5 Two live subtitling traditions

When verbatim subtitling is aimed at, as is done by the BBC for example, a corrector is hardly ever used. The presence or absence of an additional corrector is closely related to both the display and the reduction modes mentioned earlier. In the case of literal subtitling, the subtitles immediately build up word by word on the screen (scrolling mode) leaving no time for correction, either by a corrector or by the respeaker. On the other hand, when decently paraphrased subtitles are the goal rather than a verbatim rendering, one- to three-line subtitles will appear on the screen as a whole (block mode). The respeaker can monitor the subtitles as they are dictated, and has the possibility to quickly correct and edit them before they are put on the air. Sometimes, this monitoring (and fast editing) is done by a corrector. In Flanders, a

corrector is often used for fast live programmes (cycling reports, news programmes etc.), whereas real-time subtitling of programmes with a slower pace (like a tennis match) is usually done solo.

There seems to be considerable disagreement between countries (as well as researchers and users) concerning the preferred live subtitling technique. Robson (2004, p.20) states that “editing [...], as opposed to presenting verbatim captions” might be “the most-argued facet of captioning”. More specific studies comparing scrolling to block mode and verbatim to edited subtitling are needed, but whichever style turns out to be ‘better’, a decisive factor in this issue is certain to be the influence of *habit*.

Two major problems are raised by ‘advocates’ of block subtitling. First of all, they doubt the quality of the scrolling subtitles, since no correction or editing can be executed prior to broadcasting. Second, they worry about the reading speed, which they claim to be too elevated, even when the reading act is combined with normal or residual hearing. Jensema, Danthurti, & Burch (2000) found that deaf viewers spent over 80% of their time reading the subtitles. Even though the average reading speed of a normally reading adult is set at about 250 words per minute (Gibson & Levin, 1975), it is clear that fully verbatim subtitling would leave anyone with very little time for watching the programme, since a normal conversation can easily speed up to rates beyond 200 wpm. The introduction of the so-called 6-second rule (Gielen & d'Ydewalle, 1989), long used by subtitlers, set the maximum rate of subtitling at about 125 wpm, even though wpm-rates vary nationally today and can reach 160 wpm. Though this seems to be enough for normally hearing viewers, Schilperoord, de Groot, & van Son (2005) observe that this might not be the case for deaf viewers who are slower readers. On the other hand, comparing the words that were spoken in three Dutch television programmes to the subtitles that were created for them, Schilperoord, de Groot, & van Son (2005) found that a considerable part of the coherence relations was lost with subtitling, thus leading to a loss of explicitness and to changes in meaning. This might be, in fact, an argument in favor of verbatim subtitling. An experiment conducted by Ward, Wang, Paul, & Loeterman (2007) compared the comprehension rate of near-verbatim captions with edited captions, but no significant differences were found. They admitted, however, that their sample of 15 deaf and hard-of-hearing children might have been too small to draw far-reaching conclusions. Moreover, they suggested that for this kind of reception studies reading levels should be controlled.

More research needs to be done based on data of subtitling from both traditions if we wish to take firm position in the debate (see, for instance, Romero Fresco, 2009).

1.6 Live subtitling in Flanders

Although the USA and the UK undoubtedly are to be considered the pioneers of both closed and live subtitling, many countries in Europe were quick enough to follow in their footsteps. Flanders launched teletext on 8 May 1980 and started experimenting with real-time subtitling as early as 1981, then using a combination of a person revoicing what was said in the television programme and a typist entering the revoiced text into the computer as quickly as possible. In a way, the respoken technique finds its origin here and was refined and adapted with the advent of speech recognition, some 20 years later. The Flemish public television (VRT), as did the BBC, launched speech recognition as their standard method in 2001 (VRT, 2009). The success of the respoken technique enabled VRT to provide more and more programmes (live as well as pre-recorded) with intralingual subtitles. At present, VRT subtitles around 67% of its programmes intralingually, only giving precedence to Britain - BBC1 tops the list with 99% (Orero & Looms,

2008) - and The Netherlands, reaching 75% (SOAP, 2007)⁵. The 2006 VRT management contract set high benchmarks for the intralingual subtitling division, requiring that no less than 95% of Dutch programmes be subtitled by the end of 2010. Although agreeing with the necessity to strive for an increase in the number of programmes that carry subtitles, Remael (2007) expresses a deep concern about quality, especially when it comes to live subtitling: "The danger in the developments in live subtitling [...] is that quality might be sacrificed for quantity" (Remael, 2007).

Meanwhile, VRT goes on expanding its repertoire of intralingually subtitled programmes on teletext. In 2008, a gap of 30% still had to be filled by the end of 2010. At the end of 2009, however, this percentage had already decreased to 20% and mainly concerns live programmes. Most sports events are subtitled, as well as news items and some live entertainment shows. The biggest challenge (and probably the most urgent one) is the subtitling of real-time talk shows, topicality programmes, debates and the like. As we will see, this is part of the reason why we have used talk show footage as input material in this experiment.

1.7 Live subtitling versus interpreting

Interpreters undertake various cognitively demanding tasks such as retrieving, retaining, producing and monitoring information almost concurrently. Respeaking in a live subtitling context is sometimes linked to simultaneous interpreting (Eugeni, 2008), since live subtitling also involves the four main activities put forward by Jones (2002) for interpreting: listening, understanding, analyzing, and re-expressing. Indeed, live subtitling too poses a serious challenge for an individual's cognitive resources. In order to understand and manage the very specific reduction process involved in live subtitling, it is important to identify the role of the working memory during this task.

Working memory is considered to be a limited-capacity mechanism that is involved in both the processing and storing of currently active information while tasks are being carried out (Baddeley & Hitch, 1974). Simultaneous interpreting, sometimes referred to as 'listening and speaking concurrently' or 'holding the spoken message while simultaneously formulating and articulating the translated message' (Mizuno, 2005), makes use of this memory system (Osaka, 2002). Mizuno (2005) proposed an enlarged embedded processes model in which the working memory system and language comprehension and production systems constitute indispensable parts (see Figure 2). During live subtitling, respeakers simultaneously perform two major tasks, namely comprehending the audio comments and producing subtitles. The long-term memory, which includes the lexicon of both the source and target language and automated translation procedures, is here of less importance. Therefore, this model can partially be applied to the live subtitling process as well.

⁵ For an overview of European stipulations concerning subtitling, sign language and audio description, we refer to Remael (2007: 24-29).

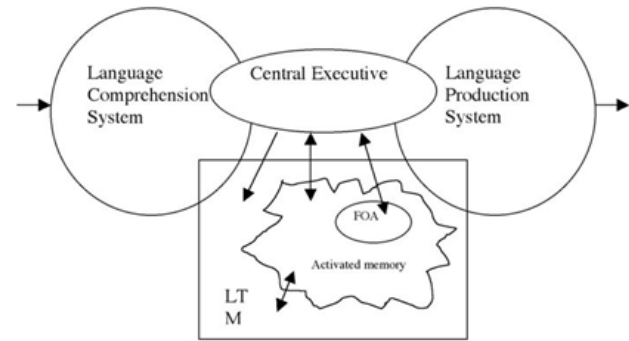


Figure 2. Process model for simultaneous interpreting, embedded in the working memory system and language comprehension/production systems (Mizuno, 2005). FOA: focus of attention, LTM: long-term memory.

As can be seen in Figure 2, the central executive and the long-term memory consist of a part of the language comprehension system and the language production system. The central executive is involved in the control of the focus of attention and coordination of the working memory system (Cowan, 1995) and does not itself have storage capacity (Baddeley, 1993). If attention switching or coordination of tasks takes long to complete, respeakers have to remember memory items longer, risking the loss of information altogether (Towse & Houston-Price, 2001). Similarly, if parsing of incoming speech in the language comprehension system or speech planning in the language production system takes longer, it will switch away from other activities that should be completed in a timely manner, risking the breakdown of the overall tasks (Mizuno, 2005). Therefore, the working memory provides a buffer for language comprehension and production. The buffer might be used as a means of maintaining subsequent words while the analysis of an earlier portion is going on (Mizuno, 2005). Composing a subtitle thus requires shifting between – and presumably concurrently activating – various memory processes, which puts a great strain on the working memory.

2. Research project

2.1 Objective

The overall ambition of this study is to improve speech recognition-driven live subtitling. To reach this goal, we have in a previous stage collected a corpus of live subtitled sports programmes at VRT. Some exploratory analyses of these data gave us an impression of the state of the art in Flanders and enabled us to design an experimental setting in which the phenomenon *text reduction* was explored, in relation to *delay* and *cognitive factors*. At a later stage, our findings will have to be implemented into training programmes for respeakers. Courses in respeaking are already taught (van der Veer, 2007; Arumi Ribas & Romero Fresco, 2008) in some countries, often as a part of an interpreting or translation curriculum. Both from schools and from television companies, however, we hear calls for a better theoretical foundation of the problems that respeakers have to deal with in practice. It is our belief that if some obnoxious errors and flaws in live subtitling can be detected and objectified, training programmes can be improved and the overall quality of live subtitling will increase.

2.2 Research questions

Based on the literature described in the introduction, we have selected a number of variables which we assume have an effect on text reduction. Matching our ambitions described in the paragraph above, the ultimate objectives of this paper can then be described as an attempt to find out the following:

1. To what extent is text reduction in live subtitling affected by the production delay, the amount of audio comment input, the proportion of full reductions and the cognitive span of the respeaker?
2. Which strategies do respeakers use to achieve text reduction?

Specific attention will be paid to (a) the amount of spoken information from live television programmes that is actually rendered in real-time subtitling, (b) the influence of reduction on delay and vice versa, (c) the influence of cognitive span on reduction and, (d) the influence of reduction strategies on text content. To tackle these issues, we have conducted both a quantitative and qualitative analysis.

3. Method

3.1 Participants

In all, 12 subtitlers (8 male, 4 female) of the Flemish public television (VRT) participated in the experiment. All participants were native speakers of Dutch, with a mean experience time with Dragon Naturally Speaking (DNS) of 3.33 years (standard deviation 2.18 years). The mean reading span and backward digit span of the respeakers were respectively 73.36 (standard deviation 6.80) and 9.73 (standard deviation 1.85). The standard for this reading span test is 66, indicating that the scores for the respeaker group in general are above average.

Participation in the experiment was entirely voluntary, unpaid and took place during office hours.

3.2 Data collection

The television excerpts we asked the participants to subtitle were taken from three episodes of a Flemish infotainment talk show named 'Phara'. The participants were familiar with this programme but they had not previously subtitled it, neither live nor in a pre-prepared setting.

The decision to choose a talk show, a genre that was not being subtitled live at the time, resulted from the consideration that strong objectives have been determined for VRT: -95% of all Dutch language programmes subtitled by end 2010. This means that talk shows too will have to be subtitled shortly. On the one hand, the results could therefore count as an indication of the respeakers' actual performance, while on the other they could uncover specific problems that must be dealt with if the goal of subtitling 95% of all programmes is to be reached. More concretely, each of the excerpts contained Dutch dialogues lasting 15 minutes. They had a

comparable format (two interviewers and three interviewees) and a speaking rate of around 200 words per minute (resp. 184, 200 and 208 wpm).

Subtitling data

For the live subtitling tasks, the participants used speech recognition software Nuance Dragon Naturally Speaking 9.5 (DNS) in combination with the subtitling software Softel Swift.

Dragon Naturally speaking is a speech recognition software package developed by Dragon Systems. The program recognizes natural and continuous speech and converts it into text on the screen. NaturallySpeaking uses a minimal visual interface. Dictated words appear in a floating tooltip as they are spoken, and when the speaker pauses, the program transcribes the words into the active window at the location of the cursor. Thus, by using voice commands, texts are written, adjustments are made, menus are activated and programs are managed (Dragon Naturally Speaking, help).

Softel's Swift is a professional subtitling workstation which provides subtitle preparation, processing and transmission (www.softel.co.uk).

The subtitling tasks were monitored by means of the logging tool *Inputlog*, a software program developed at University of Antwerp (Leijten & Van Waes, 2006). Inputlog is designed to log and analyze writing data, but can also integrate the input of dictation devices using the speech recognition software Dragon Naturally Speaking. Thus, using this logging tool, all the respeakers' keystrokes, mouse movements and speech input data were meticulously registered throughout the entire experiment.

WritingMode	Output	StartClock	StartTime	EndClock	EndTime	ActionTime	PauseTime	X	Y
2	Movement	0:00:00	0	0:00:16	16532	16532	0	252	55
2	Left Button	0:00:16	16672	0:00:16	16797	125	16672	252	55
2	Movement	0:00:16	16691	0:00:17	17691	1000	219	219	260
2	Left Button	0:00:17	17907	0:00:17	17969	62	1016	219	260
3 - dict	context in	0:04:39	279051	0:04:40	280577	1526	410		
1	ENTER	0:04:42	282250	0:04:42	282328	78	3199		
1	ENTER	0:04:42	282578	0:04:42	282641	63	328		
3 - dict	de algemene verplichtingen van de werkgever met betrekking tot de bescherming van gezondheid en veiligheid van werknemers op de werkplaats	0:04:43	283502	0:04:51	291922	8420	924		
3 - dict	\punt	0:04:55	295051	0:04:55	295928	877	11549		
1	BS	0:05:01	301594	0:05:01	301688	94	6543		
1	SPACE	0:05:02	302297	0:05:02	302375	78	703		
3 - dict	kaderen in de ontwikkeling ervan een preventiebeleid \punt	0:05:02	302687	0:05:06	306248	3561	390		

Figure 3: Merged logging data of Inputlog and Dragon Naturally Speaking (Leijten, 2007).

As shown in Figure 3, both the subtitling and logging file can be combined and integrated into one general logging file, which is based on the convergence of timestamps (Leijten, 2007). The writing mode, the output, and the start and end time of each action are given (see 3.5 analysis).

Cognitive data

The memory span of all respeakers was examined using backward digit span (Wechsler, 2000) and reading span tests (Van den Noort, Bosch, Haverkort, & Hugdahl, 2008). The backward digit span is a traditional measure of working memory (May, Hasher, & Kane, 1999) which assesses the maintenance of information in short-term memory. This widely used measure is based on a simple verbal working-memory task (Van den Noort, Bosch, Haverkort, & Hugdahl, 2008).

The reading span test (RST), first devised by Daneman & Carpenter (1980), assesses people's ability to perform active processing of a stimulus while simultaneously buffering other information in the working memory. The use of the RST as a measure of central executive functioning is based on the idea that test performance reflects the total amount of attentional resources that can be allocated to information storage and manipulation (Just & Carpenter, 1992). People who score low on the RST, often referred to as low spans, are disproportionately affected by the demands of maintaining a memory load while concurrently performing a language comprehension task (Just & Carpenter, 1992). Over the past decade, several alternatives of the RST have been developed in order to further improve the test or to develop a version in a different language (e.g., Desmette, Hupet, Schelstraete, & Van der Linden, 1995; Kondo & Osaka, 2004). Van den Noort, Bosch, Haverkort, & Hugdahl (2008) developed a new standard computerized version of the RST in four languages (Dutch, English, German, and Norwegian), which meets strict methodological criteria that are the same for all four languages. Unlike many other reading span tests, this test is controlled for the length of the sentences used, the frequency of sentence-final words, and the presentation of the material. The main advantage of this version is that it creates a possibility for direct comparison between different research groups and across languages (Van den Noort, Bosch, & Hugdahl, 2006). In the present research, we used this test to determine the participants' reading span.

3.3 Design

Three respeaking tasks were organized for the experiment, each of which intended a different reduction condition. Every respeaker had to subtitle all three fragments, each in a different reduction condition. There were three possible reduction conditions: (a) verbatim, (b) summarized, and (c) highly reduced (see procedure 3.4). These three different reduction intensities aimed at clarifying various reduction strategies. Since each fragment was subtitled in the same condition by four respeakers, we obtained 12 subtitled excerpts of the same fragment, which can be seen as one of the strengths of this study. Consequently, we were able to compare subtitles of the same fragment subtitled by different respeakers in three reduction conditions which led to a better understanding of the reduction strategies. Which episode had to be subtitled in which reduction condition was controlled for order effects.

3.4 Procedure

The experiment was conducted at the VRT, in the room where most live subtitling sessions were usually done at the time. Each participant was asked to read an introductory page which explained that the experiment did not entail an evaluation by any means, but rather that it would be an opportunity to have the overall strengths and weaknesses of speech recognition-based subtitling identified. They were also asked to pre-read the experiment's planning and instructions. The estimated duration of the experiment was 90 minutes. The participants had the right to quit at any time.

First, the subtitlers had to execute the subtitling tasks one by one. At a later stage, the participants were asked to conduct two memory tasks. The procedure was as follows:

1. In order to create a realistic setting, the subtitlers were given the opportunity to train a couple of difficult names that had been previously selected and listed by the researcher. In reality, live subtitlers will also take some time to train tricky terms and names that they would expect to appear in the target program. In our case, three short lists of about 5 words each were provided and all the subtitlers actually took their time to train them.
2. After the training, the participants had to conduct the respeaking tasks. The order in which the participants had to subtitle the different television excerpts was consistent in that they all started off with the verbatim condition and ended with the highly reduced variety, but differed as to which excerpt was subtitled when. When subtitling in the verbatim condition, only the first 9 minutes of the fragment had to be covered, because we hypothesized that a concentration limit would be reached around that stage. For the two summarizing conditions, the entire 15 minute excerpt was used.
 - **Verbatim:** The participants had to focus on quantity and stay as closely to the spoken dialogue as possible. Avoiding errors and producing nicely readable subtitles was of less importance than aiming for 100% subtitling.
 - **Summarized:** The respeakers had to focus equally on quantity and on quality. The target rendering percentage was 50%. Added to the task description was that this summarized mode would probably be the way in which they usually subtitled live programmes.
 - **Highly reduced:** In this task, 25% subtitling was set as a guideline. The main goal here was to establish a product in which the subtitles very concisely render the core elements of the television programme dialogue. In addition to that, much attention had to be paid to the quality of the subtitles: errors had to be avoided and smooth editing was preferred.
3. At a later stage, the participants were submitted to two memory span tests. In the first one, the backward digit span (Wechsler, 2000), different series of digits were orally presented by the experiment leader and had to be repeated backwards by the participant immediately afterwards. The series got longer each time and ranged from 2 digits up to 8 digits at the end. Each string was presented twice before proceeding with a new one. When errors in two consecutive trials of the same length were made, the test was ended. The correct answers were scored and added up for each participant. A second test, the reading span test (RST), examined the respeakers' verbal working memory (van den Noort et al., 2008). The RST started with an instruction on the computer screen and two exercise trials, after which the experimental material was presented. Subsequently, the participants received 100 sentences in different set sizes (2,3,4,5 or 6 sentences), which they all had to read aloud. When a respeaker had completed all the sentences of a set, the word "recall" was presented. At that point, the participants had to recall the last word of each sentence in the set. The order of recall was free. The total number of remembered words was determined.

3.5 Analyses

To register the subtitling process, we opted for a combination of Inputlog and Dragon Naturally Speaking. Inputlog makes it possible to integrate the input of the dictation processes using speech recognition software. Thus, a detailed overview of the process data was compiled by

merging the keystrokes and mouse movements, registered by Inputlog, with the speech data that were logged by DNS. Once these two sources were integrated into one data file, a general analysis was generated in Inputlog, which enabled us to visualize the entire subtitling process step by step (Leijten, 2007). As can be seen in Figure 2, the eventual output contained indications of:

- **writing mode:** keyboard (1), mouse (2), and speech (3-dict)
- **output:** indicates which key was pressed, which mouse movement was made or what was spoken
- **start and end time:** the start and end time of that specific action, in dual coding: in hours:minutes:seconds and in milliseconds
- **action time:** the total time of that specific action, in milliseconds
- **pause time:** the length of the pause during that specific action, in milliseconds

The in- and out-times of the set of subtitles were calculated directly from the time stamps that were registered automatically by the Swift software.

Based on the original audio fragments, verbatim transcriptions⁶ were made of the dialogues of the three talk show excerpts and they were parsed into transcription units (TU's), matching the corresponding subtitles (see Figure 4). As the subtitles of all 12 respeakers were fairly diverse and not everything that was said in the television programme was actually subtitled, the parsing process was rather difficult. We chose to create a new TU whenever one of the subtitlers (for whatever reason) started a new subtitle. There are a few examples of TU's for which none of the respeakers created a corresponding subtitle. We chose to isolate these as TU's anyway, because semantically nor syntactically could they be merged into the previous or the succeeding TU. Time stamps for all TU's had to be created manually. This was done by means of a program called LiveComment, developed at University of Antwerp.

Based on the verbatim transcriptions of the source text, the transcribed subtitles and, associated time stamps, various variables were calculated (see Figure 4). All calculations were based on the level of the single TU. However, some variables were based on aggregated data per minute as well. This was done by merging all TU measurements of one variable in the same minute and calculating the mean value of these TU's⁷. As a result, each variable has one specific value for each minute. Merged data were only calculated on word level. Seven variables are especially important for our analyses:

- **Reduction condition:** The respeakers were instructed to subtitle in three different reduction conditions: verbatim, summarized, and highly reduced (see procedure 3.4). In the quantitative analyses, the verbatim condition was used as reference condition.
- **Number of characters in the spoken comment:** The number of characters and words of each of the TU's of the original audio comment was counted, both on TU and minute level.
- **Reduction:** Reduction was measured by executing character and word counts for each of the TU's as well as for the subtitles, both on TU and minute level.

⁶ We chose to create real word for word transcriptions, including repetitions and hesitations, in order not to omit any information accessible to normally hearing viewers. Exceptions are stop words like 'eh' or 'he' and the hesitation particle [ə:]. These were not included, because they were thought to have no relevant denotational value whatsoever.

⁷ Since the data is distributed rather normally, the mean is a good measure for centrality.

- **Reduction percentage:** Reduction percentages for each TU could be calculated, both on character and word level, by dividing subtitle counts by TU counts. The reduction percentages for words were calculated on minute level as well.
- **Percentage 100% reduction:** This measure was calculated by splitting up the total reduction percentage into 100% reduction percentages on the one hand, and the remaining reduction percentages on the other. This variable represents the proportion of all TU's that were 100% reduced, in comparison with the number of remaining reduction rates. At the TU level, a value for the percentage 100% reduction was allocated as well. Each TU was linked to the value of the minute in which it was nested. Consequently, all TU's within the same minute have the same value.
- **Rest reduction percentage:** This measure was calculated by splitting up the total reduction percentage into 100% reduction percentages and the remaining reduction percentages. The rest reduction percentage is the mean of the remaining reduction percentages.
- **Mean delay:** The degree to which a subtitle lags behind the corresponding TU – *delay* or *décalage* - was measured on two levels. First the “in-delay” was determined, showing the time gap between the exact moment at which a TU starts and the moment at which the corresponding subtitle is put on the air. The other level is the “out-delay”. Here, the instant at which the subtitle disappears from the screen is subtracted from the moment at which the TU ends. For both delay-levels, the result is a number of seconds, indicating the time that was needed to process the source text from the aural input to the broadcast subtitle. The mean delay is then the mean of the in-delay and out-delay, expressed in seconds.

Delay IN	Delay OUT	Audio comment	# char	# words	# words 60 sec	Delay IN	Delay OUT	Subtitles	# char	# words
0:01:53	0:01:54	en daar geen middelen voor hadden.	35	6					0	0
0:01:54	0:01:56	Mijn, mijn ouders zijn met niets begonnen	42	7		0:02:02	0:02:05	Mijn ouders zijn met niets begonnen	35	6
0:01:56	0:01:57	en die hadden heel veel arbeidsethos	37	6					0	0
0:01:57	0:02:00	en die hebben dus iets ongelooflijks gecreëerd daar.	53	8	187				0	0
0:02:01	0:02:02	Dus je mag dat ook niet vergeten, hé.	38	8					0	0
0:02:03	0:02:04	Het is echt een, voor mij is dat een verschuiving	50	10		0:02:09	0:02:14	Het is echt een verschuiving van het ene succesverhaal naar het andere	70	12
0:02:05	0:02:06	van het ene succesverhaal naar het andere.	43	7					0	0

Figure 4: Example of spreadsheet with the basic data of the audio comments and subtitles.

These data were transferred to SPSS, specifying which respeaker had broadcast which fragment in which condition. Furthermore, we added an additional variable, one minute time interval (1' time interval). This variable indicates in which minute, which condition and which fragment every subtitle is nested. Consequently, the individual characteristics of each subtitle were reflected and data were analyzed in function of time. To facilitate the interpretation of the parameters, we standardized the raw scores of the continuous variables. A standard score or z-score indicates how many standard deviations an observation differs from the mean. This dimensionless quantity has a mean of zero and a standard deviation of one. The raw scores can be found in the Appendix.

In order to analyze all these data, we used mixed models, a type of multilevel analysis. Multilevel analysis is a statistical methodology for analyzing data that have a hierarchical structure. This method allows variance in outcome variables to be analyzed at multiple hierarchical levels (Quené & van den Bergh, 2008). To conduct this analysis, we used R, an open-source package for statistical analysis (R Development Core Team, 2008).

4. Results

In the first section below, a summarized description of the data is given to make these more transparent. The second section consists of a quantitative analysis. Using multilevel modeling, we created a model that allows us to explore which factors affect the reduction process. Indeed, after having ascertained that some form of text reduction is required, respeakers have to decide what and how they are going to reduce the subtitle taking into consideration the essence of the information presented in the spoken comment. To illustrate these strategies, the third section contains a qualitative analysis based on some case studies.

4.1 Descriptives

The extensiveness of the data makes it difficult to get a clear overview of the variables. Therefore, this concise aggregated description aims at increasing the transparency of the data in order to improve the clarity of the quantitative analyses. Table 1 represents a number of descriptives that form the basis for further analyses.

Table 1: Mean values and SD per minute of the subtitles, for the three reduction conditions.

	Verbatim	Summarized	Highly reduced	Mean total
Number of subtitles	11.17 (1.67)	9.28 (2.07)	7.82 (1.55)	9.61 (2.37)
Reduction percentage for words	54.61 (9.18)	64.78 (10.46)	71.06 (9.67)	64.85 (11.67)
Percentage 100% reduction	43.09 (12.40)	52.34 (14.63)	60.02 (12.05)	53.16 (14.66)
Mean delay (sec)	10.89 (2.00)	11.73 (2.62)	12.89 (3.41)	12.02 (2.96)
Rest reduction percentage	19.27 (13.57)	26.51 (13.36)	25.84 (18.24)	24.58 (15.70)

To obtain a better understanding of the different reduction strategies, the respeakers were instructed to subtitle in three different reduction conditions. As shown in Table 1, a difference in

reduction percentage between the conditions could indeed be perceived. As expected, the reduction percentage increased as the instructed reduction percentage increased. A remarkable result concerns the reduction percentage of the verbatim condition. Although the respeakers were instructed to subtitle as much as possible, more than 50% reduction occurred. Therefore, it appears that our respeakers at least found it impossible to literally subtitle a live talk show.

Another important finding relates to the fluctuation of text reduction in function of time. As shown in Figure 5, the fluctuations in the number of words in the spoken comment (top line) are mirrored to a certain extent in the number of words in the final subtitles. This holds for all three reduction conditions. For instance, an increase in the number of spoken words also leads to a relative increase in the number of words that are produced when subtitling. This finding suggests that text reduction is – at least partially - based on the audio input, independently of the reduction condition.

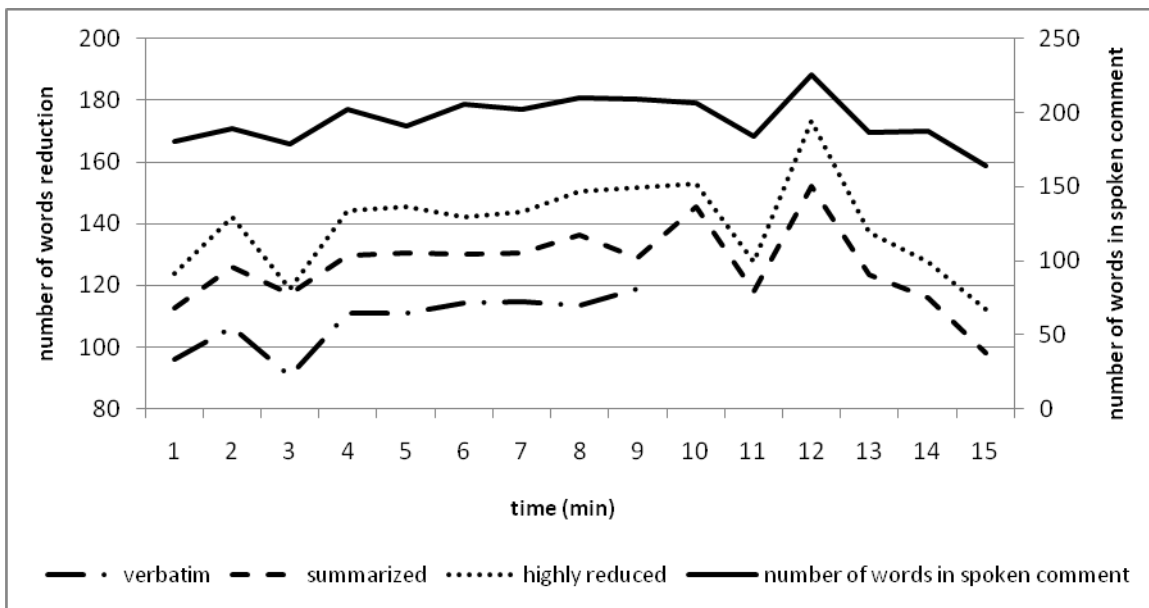


Figure 5. Number of words in spoken comment and associated number of words reduction, per minute, for the verbatim, summarized and highly reduced condition.

A different way to explore the reduction strategies is based on the proportion of utterances that were fully reduced (= 100% reduction). As shown in Figure 6, this percentage varies between the three conditions and increases in relation to the level of reduction that was targeted (cf. *infra* for statistical evidence). Therefore, we assume that respeakers make use of omissions as a strategy to further achieve reduction rather than shortening the subtitles. This supposition is reinforced by the number of subtitles produced, since this number also increases from the verbatim over the summarized to the highly reduced condition. The rest reduction percentage on the other hand, remains relatively stable across both summarized and highly reduced reduction conditions, which indicates that respeakers mainly use complete omissions to adapt the reduction percentage in an advanced reduction stage.

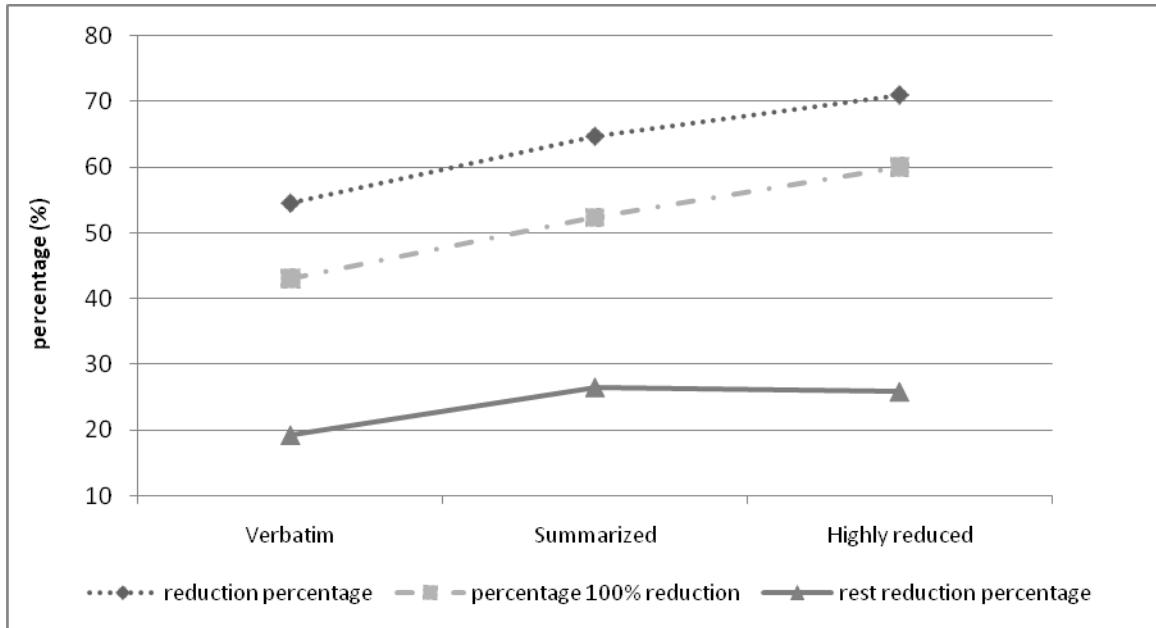


Figure 6. Reduction percentages per minute, for the verbatim, summarized and highly reduced condition.

Besides text reduction, the delay is an important characteristic of the live subtitling process. A mean delay of 12 seconds can be observed. Consequently, the reader receives the subtitles on average 12 seconds after the original audio comment, possibly disconnecting the information of the subtitle from the images in question.

4.2 Quantitative analysis

In this section we describe the quantitative analysis that we performed. The main goal of this analysis is to provide a model that explains the causes of text reduction.

Multilevel modelling

A typical assumption of multilevel models is that data are strictly hierarchical in structure. That is, each lower level unit is nested in a single unit or group at the next higher level. However, some analyses involve more than one type of grouping within a higher level, i.e. when units from the immediate lower level are nested in a single unit of each higher level. This also applies to our data. In our 2-level model, there is not just one random effect at level 2, but two crossed random effects. As can be seen in Figure 7, each observation at level 1 is nested in the combination of these two random effects. In other words, each subtitle is nested in a single respeaker and in a single minute interval⁸, whereas respeakers and minute intervals are crossed random factors at level 2 of the model. Because of these hierarchical characteristics of the data, we opted for a multilevel mixed-effects modelling of the data (see Quené & Van den Bergh, 2008 for further details).

⁸ Each minute interval represents one consecutive minute of the original audio fragment.

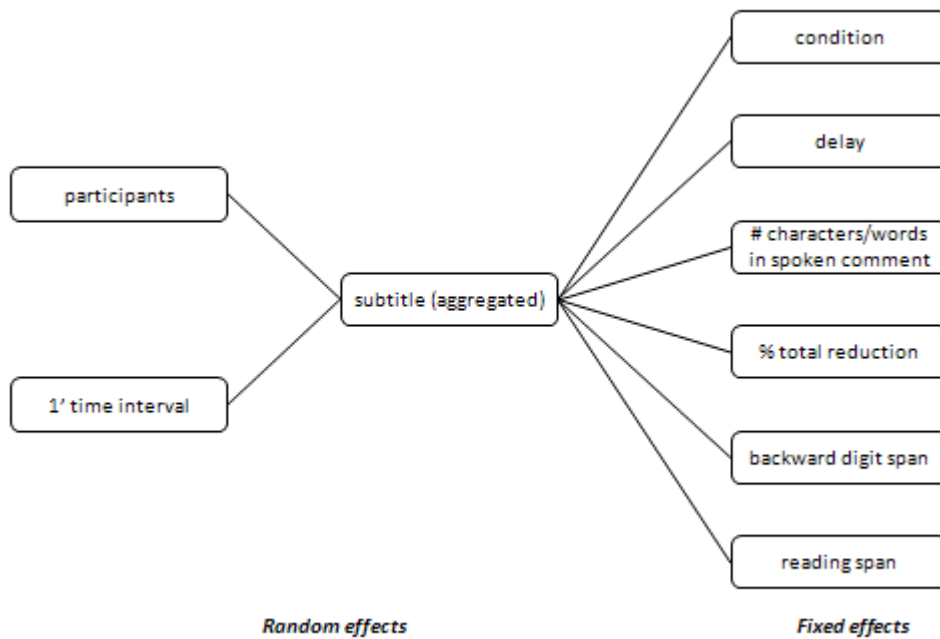


Figure 7: Mixed model describing the relation between participants (Level 2), 1' time interval (Level 2) and subtitle (Level 1) used in multilevel analysis.

The empty mixed-effects model for crossed random effects is given as:

$$Y_{i(jk)} = \gamma_{0(00)} + (u_{0(j0)} + v_{0(0k)} + e_{i(jk)}) \tag{1}$$

As shown in (1), the multilevel model consists of two parts, the fixed part and the random part. The fixed part, $\gamma_{0(00)}$, models the effect of the overall mean that underlies all observations. Afterwards, multiple fixed effects are added to the model. In the random part, between brackets, the total variance of Y is decomposed into three components: (a) the unique component of each participant $u_{0(j0)}$, (b) the unique component of each minute interval $v_{0(0k)}$, and (c) the residual component $e_{i(jk)}$, which represents the interaction between the participant and the minute interval. Each random effect is paired with an estimate of the standard deviation that characterizes the spread of the random effects for the intercepts⁹.

4.2.2 Models

The quantitative analysis consists of two models. In a first stage, we fitted a model based on aggregated data per minute in order to estimate the reduction percentage for words and in order to be able to analyze the data in function of time. In a second stage, we fitted a model with the subtitle (TU) as lowest level and the reduction percentage for characters as dependent variable. This analysis aimed at examining the rest reduction percentages. This second model is based on a larger set of observations as each individual subtitle is taken into account¹⁰.

⁹ The intercept is the overall mean that underlies all observations, viz. subtitles, participants and 1' time interval. It is the value when all explanatory variables equal zero, that is, have a mean value.

¹⁰ We have to be careful in interpreting this analysis since the results exclusively reflect the reduction process of the comments that are completely or partially subtitled. The variable *delay* only has a value if

Both models were estimated in two steps. In a first step, we estimated the 'zero model'. The zero model, shown in (1), is an 'empty' model, with a constant as its only explanatory variable. It gives us an estimate of the overall average (the intercept $\nu_{0(00)}$) and the variances at the different levels. As noted in the method section, the verbatim condition was used as reference condition. In a next step we extended the zero model by including the variables that characterize the subtitles. Obviously, the two following questions immediately arise: (a) Are these fixed effects parameters justified? and (b) Is the extended model significantly better than the zero model? In addressing the first question, we used Markov chain Monte Carlo (MCMC) sampling to assess the significance at the 5% significance level. With this method, we can estimate p-values based on the t -statistics (Baayen, 2006, p.270). The answer to the second question is based on the likelihood ratio test. The likelihood ratio test takes the squared log likelihood for the smaller model and compares it with the log likelihood for the larger model. A significant decrease indicates an improvement over the empty model (Baayen, 2006).

Model 1

In a first step we estimated the multilevel model based on aggregated data per minute to explore the variance at 1' time interval, participant, and subtitle level. The estimated parameters from the zero aggregated and net zero aggregated model are given in Table 2. A way to determine whether multi level analysis is necessary is based on the intra class correlation (ICC). The ICC is the proportion of the variance in the outcome that is attributed to the higher level units. A low ICC implies that the participants or minute intervals of the different groups do not systematically differ, whereas a high ICC indicates that the response is highly clustered within groups. In the zero model, the ICC on the 1' interval and participant level amounts to respectively 55% and 18%. In other words, 55% of the variance between the subtitles can be explained by the specified minute in which the subtitle is nested, whereas 18% can be attributed to the participants.

In a second step we added the explanatory variables condition, mean delay, number of words in the spoken comment, the percentage 100% reduction, backward digit span, and reading span to the model. Both conditions ($t = 5.35, p \leq 0.001$; $t = 6.99, p \leq 0.001$), the average delay ($t = 3.59, p \leq 0.001$), the number of words in the spoken comment ($t = 7.68, p \leq 0.001$) and the percentage 100% reduction ($t = 16.43, p \leq 0.001$) all appeared to be significant. The backward digit span and the reading span, on the other hand, did not have a significant effect on the reduction percentage ($t = -0.57, p = 0.570$; $t = 0.39, p = 0.695$).

Although all explanatory variables yield a significant effect, it was still unclear whether the net zero model is an improvement over the zero model at this point. Therefore, a likelihood ratio test was used to compare the fit of the two models. The -2 log likelihood of each model is given in Table 2. The considerable reduction in -2 log likelihood indicates that the net zero model is indeed an improvement over the empty model ($\chi^2(5) = 799.97, p \leq 0.001$).

the utterance is subtitled. When the respeaker decides not to subtitle the utterance, this value is missing. Consequently, the whole observation will be excluded from the analysis.

Table 2. Estimated parameters (with standard error of estimate and 95% confidence interval) for the reduction percentage of the Zero model (N=468) and the Net zero model (N=405).

	Reduction percentage	
	Zero model	Net zero model
	Est. (SE)	Est. (SE)
Fixed Part		
Intercept	64.852 (1.675)	59.661 (1.079)
Condition 2	-	6.337 (1.185)
Condition 3	-	8.580 (1.228)
Mean delay	-	1.094 (0.307)
Number of words in spoken comment	-	3.381 (0.440)
Percentage 100% reduction	-	6.444 (0.392)
	Est. [HPD95]	Est. [HPD95]
Random Part		
1' interval level variance	77.521 [4.372,5.802]	16.213 [1.737,2.871]
Participant level variance	24.730 [2.739,6.294]	2.944 [0.755,2.844]
Residual variance	38.451 [7.295,8.565]	19.909 [4.781,5.656]
-2 log(lh)	2769561.64	1598201.64
p-value		≤ 0.001

Note. Est. = parameter estimate, SE = standard error, HPD95 = 95% confidence interval, S = Summarized, HR = Highly Reduced

Considering the condition, delay, input, and the percentage complete reductions, the ICC at 1' interval and participant level decreases respectively from 55% and 18% to about 42% and 8%. A part of the variance has been explained by the fixed variables.

The estimated mean reduction percentage after taking into account the explanatory variables is almost 60%. In other words, when subtitling in the verbatim condition, with an average delay, average number of words in the spoken comment, and an average percentage of 100% reduction, a reduction percentage of 60% can be expected. By changing the reference condition, the mean reduction percentages for both other conditions were calculated. For the summarized and highly reduced conditions, a mean reduction percentage of respectively 65% and 68% was found. All three conditions differed significantly (Condition 1-2: $t = 5.35$, $p \leq 0.001$; Condition 1-3: $t = 6.99$, $p \leq 0.001$; Condition 2-3: $t = 2.22$, $p = 0.029$).

A rather contradictory result that needs explaining is the positive coefficient for the average delay. This result indicates that a large delay is attended with a high reduction percentage. At first sight, however, this is not what we expected. A high reduction percentage is logically associated with a low delay since one gains time by subtitling less. This analysis, however, is based on aggregated data, which causes the positive correlation. Each subtitled comment has a particular reduction percentage and delay. If the utterance has not been subtitled, a reduction percentage of 100% without any delay is reported. When the delay increases too much, the respeaker often chooses not to subtitle a number of subsequent utterances. Therefore, the following reduction percentages will be 100% without any reported delay. Consequently, by

merging the data, a large delay will be linked to the corresponding reduction percentage, but also to the subsequent reduction percentages if they amount 100%. As a result, a large delay seems to be attended with a large reduction percentage, which is expressed in a positive coefficient. An example can be found in Appendix B.

The large positive coefficient of the variable *percentage 100% reduction* reveals an interesting finding concerning the reduction process. As explained above, a positive coefficient indicates a positive relationship. In other words, an increase in reduction percentage is attended with an increase in the percentage that is 100% reduced. More specifically, if the proportion of complete reductions increases by 10%, the reduction percentage increases by 4.4% (see Appendix A). Hence, it appears that respeakers prefer omitting a comment rather than shortening it when a higher degree of reduction is targeted (or needed).

As mentioned above, this analysis revealed no significant effect of the backward digit span and the reading span.

Model 2

In the next step, we proceeded from an analysis on the aggregated level to an analysis with the subtitle as lowest level.

First, we conducted an analysis without predictors to assess whether multilevel was needed. Table 3 gives an overview of the estimated parameters. The intra class correlations at the 1' interval and participant level are respectively 3% and 1%. Both values showed to be significant, making multilevel modeling necessary.

In a next step, we added the various input, output and participant characteristics to the zero model in order to construct the net zero model. In this model, both conditions ($t = 3.6, p \leq 0.001$; $t = 4.15, p \leq 0.001$), the average delay ($t = -15.99, p \leq 0.001$) and the number of characters in the spoken comment ($t = 38.34, p \leq 0.001$) all had a significant effect on the reduction percentage. As already observed in the previous model, both the backward digit span and the reading span proved not to be significant ($t = 0.96, p = 0.339$; $t = -1.61, p = 0.108$). The percentage 100% reduction was not significant either ($t = -0.83, p = 0.409$).

By adding significant explanatory variables, we expect the extended model to be an improvement over the empty model. This is confirmed by a significant decrease in deviance ($\chi^2(4) = 62053, p \leq 0.001$).

Table 3. Estimated parameters (with standard error of estimate and confidence interval) for the reduction percentage for the reduction percentage of the Zero model (N=9386) and the Net zero model (N=3701).

	Reduction percentage	
	Zero model	Net zero model
	Est. (SE)	Est. (SE)
Fixed Part		
Intercept	65.759 (1.657)	16.638 (1.887)
Condition 2	-	6.907 (1.917)
Condition 3	-	7.962 (1.916)
Mean delay	-	-7.881 (0.493)
Number of characters in spoken comment	-	18.516 (0.483)

	Est. [HPD95]	Est. [HPD95]
Random Part		
Minute id level variance	58.165 [6.145,8.819]	38.665 [4,615,7.125]
Participant level variance	24.667 [3.097,8.211]	14.940 [2.193,7.353]
Residual variance	1808.027 [41.909,43.148]	754.662 [26.916,28.165]
-2 log(lh)	2362251609	308775184
p-value		≤ 0.001

Note. Est. = parameter estimate, SE = standard error, HPD95 = 95% confidence interval, S = Summarized, HR = Highly Reduced

In the verbatim condition, with mean values for all variables, a mean reduction percentage of 16.6% was observed. This percentage reflects the reduction percentage of those utterances that are partially or fully subtitled. In the summarized and highly reduced condition, reduction percentages of respectively 23.5% and 24.6% were found. As reported earlier, both reduction conditions differed significantly from the verbatim condition (Condition 1-2: $t = 3.60$, $p \leq 0.001$; Condition 1-3: $t = 4.15$, $p \leq 0.001$). The reduction conditions 2 and 3, however, did not significantly differ ($t = 0.61$, $p = 0.543$).

In contrast to the first model, a significant negative effect of the delay was observed. The raw scores (Appendix C) showed that a delay increase of one second is attended with a decrease of the reduction percentage of 1.75 percent.

The lack of significance of the variable ‘percentage 100% reduction’ is entirely consistent with what we expected since all complete reductions were excluded by including the delay as an explanatory variable. As outlined in the method section, all subtitles within the same minute are linked to the mean percentage 100% reduction for that specific minute. If, in that particular minute, a large number of complete reductions occur, both the reduction percentage and the mean percentage 100% reduction for that minute will both be large in a relative proportion. In this model, however, only the comments that are not or partially reduced were included in this analysis. The complete reductions have been left out. As reported in the descriptives, a mean rest reduction percentage of 24.58%, this is the mean reduction percentage without complete reductions, can be observed. This relative low percentage is still linked to the percentage 100% reduction of the minute in which the subtitle is nested. Consequently, a distortion of the original relation occurs since the low rest reduction percentages are still attended with either low or high percentages 100% reduction.

In order to verify this assumption, we removed the delay from the model whereby the complete reductions were added to the analysis. We assumed that excluding the delay would result in a significant effect of the percentage 100% reduction. This variable indeed appeared to be significant ($t = 16.73$, $p \leq 0.001$). In this alternative model, the number of characters in the spoken comment, however, did not reach the significance level of .05 ($t = -0.31$, $p = 0.753$). The absence of significance is probably due to the nature of the choice for complete reduction, which seems to be based on other factors rather than on the number of characters in the spoken comment. The correlation between the number of characters in the spoken comment and the reduction percentage supports this assumption. More specifically, although a correlation between the number of characters in the spoken comment and the total reduction percentage is absent ($r = -0.020$), a strong positive correlation can be found between the number of characters in the spoken comment and the reduction percentage of the comments

that are subtitled ($r = 0.503$). In other words, if a respeaker chooses to subtitle a comment, a large input is attended with a high reduction percentage. However, if the respeaker decides not to subtitle the comment, no correlation can be found (-0.052). In view of these findings, the decision to reduce 100% of the comment seems not to be based on the amount of input, while the decision to partially reduce is. The delay might be an alternative explanation. More research is needed to further clarify this issue.

Taking into account the condition, delay, and input, an ICC of 5% and 2% can be observed respectively at the minute id and participant level. These low values, however, are significant. Consequently, 5% of the differences between subtitles can be explained by the minute in which they are nested. Furthermore, 2% of the differences are attributed to the participants.

These analyses show that the delay as well as the number of spoken words in audio comment and the percentage 100% reduction influence the reduction degree. At the same time, it is important to examine which types of reduction strategies are used.

4.3 Qualitative analysis

The degree of reduction is not always a flawless measure to evaluate the amount of meaning a subtitle actually retains. How much of the source text message is retained, is especially important since most deaf viewers demand verbatim subtitles in order to have full access to information available to the hearing population (Ofcom, 2005). As mentioned in the introduction, respeakers use some commonly known guidelines for subtitling in a more general sense, but no explicit or fixed rules regarding subtitling or reduction exist. In general terms, one could say that the respeaker must act on the principle of relevance. But then again, some subtitling choices may be more appropriate than others. The examples below illustrate some of the strategies respeakers often use. Generally, different subtitling strategies are used in combination. These strategies have to be seen as subtitling solutions that are resorted to whenever respeakers are confronted with a challenging information load rather than as general subtitling rules.

In addressing this issue, we will illustrate some of the reduction strategies respeakers often use (Diaz Cintas & Remael, 2007). In particular, this analysis aims at illustrating different effects of text reduction on the meaning of the source text. Each example contains a combination of strategies but we will mainly pay attention to one or two strategies. The extent of the corpus makes it impossible to look at all data with respect to content. Therefore, we restrict ourselves to a number of fragments that exemplify various characteristic reduction strategies. The design of the experiment gives us the unique opportunity to compare different subtitles of the same fragment as each fragment was subtitled by 12 respeakers.

Using a shorter near-synonym or equivalent expression

Using shorter synonyms is a relative simple strategy to reduce subtitle length. As can be seen in Table 4, comment (1) of the source text is repeated and confirmed in comment (2). Both respeakers made successful use of a synonym by subtitling the confirmative repetition ‘sommige mannen doen dat inderdaad ook al [some men indeed already do that]’ in a condensed form. The confirmative words ‘ja [yes]’ and ‘dat klopt [that is right]’ cover all information contained in the repeated sentence. Thus, this reduction strategy seems to work well.

In addition to this confirmation, a contrastive comparison introduced by the relational marker ‘maar [but]’ has been added. This superlative forms the essence of the entire utterance.

The contrastive comparison, however, has only been subtitled by respeaker 1. Although this omission does not affect the comprehensibility of the dialogue, it certainly has an effect on the meaning of the utterance. Moreover, by omitting the comparative relation, the main idea of the utterance is lost. Respeaker 2 has changed the original meaning from ‘starting an argument’ into ‘confirming an empirical fact’.

Table 4. Examples of text reduction processes

	Audio comment	Respeaker 1	Respeaker 2	Word red 1	Word red 2
(1)	Ja, maar doen sommige mannen dat toch ook al met hun lichaam? [Yes, but do some men do that already with their body]	Maar sommige mannen doen dat ook met hun lichaam [But some men do that with their body as well]	Mannen doen dat ook al [Men already do that]	2	7
(2)	Sommige mannen doen dat inderdaad ook al, maar bij vrouwen is dat veel extremer. [Some men indeed already do that, but in women it is more extreme.]	Ja, maar bij vrouwen is het extremer [Yes, but in women it is more extreme]	Dat klopt [That is right]	7	12

Omissions and changing questions into affirmative sentences

Reduction is a comprehensive concept. As pointed out in the introduction, it includes both omissions and paraphrases. A combination of both is often observed.

In this example, the respeaker profits by changing the question into an affirmative sentence since two sentences were merged into one (Table 5). In utterance (5), a question that refers back to (1) is asked. Rather than subtitled this question literally, the respeaker rephrases the utterance and subtitled it indirectly in the subsequent one. ‘Ik ben opgegroeid [I grew up with the Laatste Nieuws]’ refers back to (1) which means the question in (5) becomes redundant. This is a typical subtitled trick.

Next to condensing, the respeakers also use omissions to reduce text length. Relevance is the key to decide if utterances can be omitted. As can be seen in Table 5, utterances (3), (4) and (5) have been omitted altogether. The unfinished thought and the backchanneling words could be experienced as disturbing since they potentially interfere with the main line of thought. Utterance (4) does contain concrete information without incomplete thoughts, repetitions or stop words, but this information is not vital to follow the logical structure of the reasoning and can be seen as a side remark. The question is: do these omissions have a significant impact on the comprehensibility of the entire passage? Utterance (3) contains two types of information: (a) ‘Ah ja [Ah yes]’ and ‘Ja [Yes]’ have a confirmative function and are used for backchanneling, and (b) ‘Dus dat is een, dat was Het Volk [So, that is a, that was Het Volk]’ is an expansion of (2), in which the name of the paper asked at in (1) is given. The extension in (3) refers to the former name of the paper. Neither of the information units in (3) are essential in order to understand the logical sequence of the utterances, as they contain no new information. In conclusion, omission in this context makes the main line of thought more transparent and straightforward by eliminating secondary comments, which is especially important for a target audience with slow readers.

Table 5. Examples of text reduction processes

	Audio comment	Respeaker	Word reduction
(1)	Met welke krant bent u opgegroeid? [With which paper did you grow up?]	Met welke krant bent u opgegroeid? [With which paper did you grow up?]	0
(2)	<i>Wel, met met De Nieuwe Gids.</i> [Well, with De Nieuwe Gids]	Met de nieuwe gids [With de nieuwe gids]	2
(3)	Ah ja. <i>Dus dat is een, dat was Het Volk.</i> Ja. [Ah yes. So, that is a, that was Het Volk. Yes]	-	11
(4)	<i>Maar mijn vader schreef artikels over economie in De Nieuwe Gids.</i> [But my father wrote articles about economy in De Nieuwe Gids.]	-	11
(5)	En u? En. [And you? And.]	-	3
(6)	<i>Het Laatste Nieuws natuurlijk, he.</i> Met Het Laatste Nieuws. [Het Laatste Nieuws of course, isn't it. With Het Laatste Nieuws.]	Ik ben opgegroeid met het Laatste Nieuws [I grew up with the Laatste Nieuws]	2

Note. Italics in the main text indicate a second speaker.

In other cases, omission can be problematic. As shown in Table 6, both respeakers here chose to omit utterance (3). However, only respeaker 1 compensates this loss by conveying the information in a previous subtitle. The utterances in 1 and 2 are quite literally reproduced by the respeakers. At first sight, the difference between the two subtitles may appear trivial. Both subtitles seem to exhibit just a little lexical nuance in the verb (ligt [lies] → bestaat [exists]). Nevertheless, taking into account the different target audiences, a considerable difference in meaning occurs. In the source text, utterance (3) gives additional information about utterance (2), by which the meaning of comment (2) is clarified. Respeaker 1 covers this clarification by replacing the verb 'ligt [lies]' by the verb 'bestaat [exists]', whereas respeaker 2 simply does not clarify comment (2) and ignores (3). Both allochtones and older people may have difficulties in correctly interpreting such nuances, especially with fast subtitles. The probability that some target audiences, especially a target audience that cannot combine reading and listening, will interpret the meaning of 'ligt' literally, which complicates the comprehensibility, is thus real. Therefore, just a small nuance in the choice of words can be problematic for overall comprehensibility in some cases, whereas in others deleting a larger chunk of text hardly has any effect.

Table 6. Examples of text reduction processes

	Audio comment	Respeaker 1	Respeaker 2	Word red 1	Word red 2
(1)	we zeggen: [we say:]	En dan zeggen we bij het terugkomen vaak: [And then we often say while returning:]	Wij zeggen dan: we gaan naar huis, eens kijken of ons land er nog ligt [Then, we say: we are going home, see if our country still lies there]	-6	-13
(2)	we gaan naar huis, eens kijken of ons land er nog ligt. [we are going home, let's see if our country still lies there]	"Eens kijken of ons land nog bestaat." [Let's see if our country still exists.]		5	12
(3)	En dan, en dan bedoelen we eigenlijk: kijken of ondertussen Vlaanderen misschien al bij die Hollanders hoort, en, en, en het Franstalige gedeelte bij Sarkozy hoort. [And then, and then we actually mean: see if meanwhile Flanders already belongs to the Dutchmen, and, and, and the French part belongs to Sarkozy]	-	-	26	26

Simplifying indicators of modality

Omitting or simplifying indicators of modality can save space and time. As shown in Table 7, both respeakers here decided not to subtitle the first phrase in (1), an indicator of modality. It concerns a repetition of oral features and can therefore be omitted. In fact, this omission does not only shorten the subtitle, it also makes it more readable.

Other reductions occur as well. Both respeakers have left the additional information about Didier Reynders out as it is supposedly known information. Since the full name of the Minister is still here, the deletion has little influence on the comprehensibility of the utterance. An important remark in this context concerns people's background knowledge. World knowledge is crucially important in comprehending the meaning of texts. For example, foreigners are not aware of the political situation in our country and may have interpretation problems due to this omission. Thus, one always has to keep in mind what one's target audience is expected to know. However, in this example, the audio fragment of Didier Reynders itself was accompanied with the subtitle 'Didier Reynders, MR chairman', which makes subtitling his function redundant.

Another observation worth mentioning is restrictive substitution. Respeaker 1 replaces 'Franstalig België [French-speaking Belgium]' by the shorter equivalent expression 'Wallonië [Wallonia]'. Respeaker 2 even takes it a step further. He opts for just one subtitle by merging the utterances in (1) en (2) into one. Thus, with a reduction percentage of 58% for subtitle 1 and 75% for subtitle 2, we can conclude that a significant reduction took place. Nevertheless, no significant loss of meaning occurred (if one knows who Didier Reynders is).

Table 7. Examples of text reduction processes

	Audio comment	Respeaker 1	Respeaker 2	Word red 1	Word red 2
(1)	Maar laat ons misschien. <i>Het is zoiets ja</i> , eens luisteren naar een van de cruciale spelers in Franstalig België, [But let us maybe. <i>It is like yes</i> , let's listen to one of the crucial players in French-speaking Belgium,]	We luisteren naar een cruciale speler in Wallonië [We listen to a crucial player in Wallonia]	We luisteren even naar Didier Reynders [We listen to Didier Reynders]	11	13
(2)	vicepremier en MR-voorzitter Didier Reynders [vice prime minister and MR chairman Didier Reynders]	Didier Reynders [Didier Reynders]	-	3	5

Note. Italic text indicates a second speaker.

In sum, the examples analyzed in this section illustrate that respeakers use a wide variety of strategies to reduce the subtitles in such a way that the loss of information is limited as much as possible. The analyses shed more light on the effect that quantitative reduction may have on comprehension, and show the delicate relation between the (reduced) subtitles and the original content.

5. Discussion

Within the constraints of time, place, and cognition, respeakers have to listen and speak concurrently, which makes live subtitling a particularly challenging task. Consequently, the written version of speech is nearly always a reduced form of the spoken comments. The experiment presented in this article aimed at examining the causes and consequences of text reduction. With this issue in mind, live subtitling data were collected and both a quantitative and qualitative analysis was conducted. Preliminary to these analyses, a concise description of the data was presented in order to increase the transparency of the dataset.

The descriptives show that it is almost impossible to subtitle literally when live subtitling. Although the respeakers were instructed to literally subtitle in the verbatim condition, no one succeeded in doing so. Moreover, on average, less than half of the source text was subtitled. Thus, we can conclude that, while subtitling a live talk show, a substantial reduction is inevitable. These results are consistent with the findings of Romero-Fresco (2009). He reported that even when respeakers were encouraged to produce verbatim subtitles, they did not seem able to do this. This finding has important implications for a number of target audiences. Many deaf people demand verbatim subtitles in order to have full access to all original information available to the hearing population. However, considering these results, this demand seems unrealistic, most certainly, if using block titles.

The quantitative analysis, which allowed us to explore the causes of text reduction, showed that reduction is not a random operation. Text reduction is largely determined by a number of external factors.

In the aggregated model per minute, text reduction was affected by delay, number of words in the spoken comment and percentage 100% reduction. In fact, an increase in one of the mentioned variables results in an increase in the reduction percentage. With regard to the reduction strategies, the results suggest that respeakers prefer to omit a comment rather than reducing it. This idea is partially supported by the rest reduction percentage. The reduction percentage of the utterances that are partially reduced is relatively constant across both reduction conditions, while the proportion of full omissions differs. This is unlike the verbatim condition, where a lower percentage was found. The reason for this deviation is not entirely clear. A possible explanation is that respeakers use both complete omissions and partial reductions to reduce up to a certain reduction percentage. If they, subsequently, decide to further reduce, mainly complete omissions are used. It might also be the case that the deviation is due to the data itself. In the verbatim condition, two extreme outliers can be observed, which minimize the rest reduction percentage. Overall, the present experiment clearly supports the idea that respeakers use complete reductions to further realize reduction.

In the second model, we aimed at analyzing the individual subtitles. Both the delay and number of characters in the source text considerably influenced the reduction percentage. A decrease in delay is attended with an increase of text reduction. This finding seems to support the idea that one can manipulate text reduction by means of the delay. However, it is still unclear whether text reduction is influenced by delay or vice versa. By contrast, an increase in the number of characters in the source text is attended with an increase in the reduction percentage. Based on an additional analysis, it appears that the decision to fully omit a comment seems not to be based on the amount of input, while the decision to partially reduce is. It is quite possible that respeakers first assess the delay to decide if complete reduction is necessary. If not, partial reduction occurs, which will be based on the amount of input. Thus, the delay may be an important decisional factor for the choice between complete and partial reductions. This result implicates that, when using an antenna delay, reduction strategies might change.

The present experiment fails to confirm the direct influence of working memory on text reduction. No significant effect of working memory was found, neither in model 1 nor model 2. A possible explanation might be found in the fact that all respeakers, except for 2, had a considerably higher reading span than the average score for adults. A ceiling effect might have occurred here, obscuring our ability to observe differences in performance. In a comparable context, viz. an interlingual interpreting task, Osaka (1994) argued that the interpreting task has a significant relation to the listening span task of the interpreter and that interpreting processes are influenced by the working memory. By contrast, Liu, Schallert, & Carroll (2004) also assessed the general working memory capacity of professional interpreters, beginning, and advanced students by using the listening span test. No significant difference among the groups was found. An alternative explanation was reported by Just, Carpenter, and colleagues (Daneman, 1991; Daneman & Carpenter, 1983). They suggest that working memory varies as a function of an individual's efficiency with the specific task processes.

The final section explores the subtitles from a more qualitative perspective to explore the impact of text reduction with respect to content. The examples presented show that, although commonly known subtitle guidelines exist, various respeakers subtitle the same utterance in a different way. Additionally, an inevitable loss of information could be observed during text reduction. However, this loss mostly concerns inferior or superfluous information. Moreover, overall comprehension of the source text is generally maintained, as supported by Romero-Fresco (2009). His data suggests that the editing process, as performed by respeakers, loses a minimal amount of important information, given the respeakers' ability to edit words that do not contribute to the idea units conveyed in the original text.

In interpreting the results, two remarks have to be taken into consideration. First, the speech rate of the source text strongly varies within and between each fragment. Second, the respeakers were not used to subtitling literally, especially not in dialogical context. Therefore, we may expect to see a lower reduction percentage in the verbatim condition when working with respeakers who are trained to subtitle literally. Moreover, due to the specific characteristics of the programme subtitled in this study, we do not claim that our findings can be generalized across different types of live programmes.

To conclude, we would like to formulate some suggestions for further research. First, it goes without saying that further research is needed to confirm and/or refine the results reported here. Taking into consideration that live programmes vary greatly with respect to amount of spoken information, density of content, degree of self-explanation, redundancy etc., it is certainly advisable to reiterate this type of research with different materials. Moreover, by focusing more explicitly on the interplay of complete and partial reductions, for example, it should also be possible to gain a better insight into the respeakers' (unintentionally) pursue of an optimal reduction percentage.

Secondly, since in the near future a limited antenna delay may become a standard in broadcasting live programmes, it might be worthwhile to further investigate reduction in relation with antenna delay. If time constraints become less important in the context of (quasi) live subtitling, text reduction may be affected by some other variables than those identified in the current experiment. This complementary perspective may provide crucial information concerning the relation between delay and reduction.

Thirdly, it deserves recommendation for follow-up research to make both a quantitative and qualitative inventory of the errors that occur in the process of subtitling. As shown in Figure 1, errors – and especially the correction of errors – might cause considerable delay, leading indirectly to (extra) reduction. Moreover, it would be beneficial to determine also the exact *cause* of a given subtitling errors. Their origin may be technical (e.g. software / hardware failure), it may be a human error (e.g. slurred or hesitating speech) or a combination of both. This identification process, obviously, would have to include both paraphrasing and subtitling errors as well as errors that were made throughout the process but that were corrected before they were broadcasted. The logging tool we used in this experiment, viz. Inputlog, provides us with information to further explore the subtitling process data from this perspective. It does not only provide an online log of the dictated and corrected text, it is also possible to connect the linear representation of text production and revision to a segmented audio recording of the dictated text.

Finally, to improve our understanding of the role of working memory, more attention needs to be paid to the variance in working memory load during live subtitling tasks and its influence on the choice for specific text reduction strategies. Of course, also inter- and intra-personal aspects should be taken into account when approaching these aspects.

Acknowledgements

We especially would like to thank Sven de Mayer for his patience in explaining us the principles of multilevel analyses. In addition, we would like to thank Astrid Geudens helping us to find a Dutch version of the reading span test. We would also like to thank Maurits Van den Noort for providing his computerized version of the Reading Span Test. Bart Van de Velde did a great job in developing the time registration program LiveComment.

REFERENCES

- Arumi Ribas, M., & Romero Fresco, P. (2008). A Practical Proposal for the Training of Respeakers. *The Journal of Specialised Translation* (10), 106-127.
- Baayen, R. (2006). Mixed Models. In R. Baayen, *Practical Data Analysis for the Language Sciences with R* (pp. 263-328). Nijmegen.
- Baddeley, A. D. (1993). Working Memory or Working Attention. In A. D. Baddeley, & L. Weiskrantz, *Attention: Selection, Awareness, and Control: A Tribute to Donald Broadbent*. Oxford: Oxford University Press.
- Baddeley, A. D., & Hitch, G. (1974). Working Memory. In *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.
- Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. New York: Oxford University Press.
- Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, 20, 445-464.
- Daneman, M., & Carpenter, P. A. (1980). Individual Differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Daneman, M., & Carpenter, P. A. (1983). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- De Linde, Z., & Kay, N. (1999). *The Semiotics of Subtitling*. Manchester: St. Jerome Publishing.
- Desmette, D., Hupet, M., Schelstraete, M. A., & Van der Linden, M. (1995). Adaptation en langue française du "Reading Span Test" de Daneman et Carpenter [A French adaptation of Daneman and Carpenter's "Reading Span Test"]. *L'Année Psychologique*, 95, 459-482.
- Diaz Cintas, J. (2005). Back to the Future in Subtitling. *MuTra 2005 - Challenges of Multidimensional Translation*. London.
- Diaz Cintas, J., & Remael, A. (2007). *Audiovisual Translation: Subtitling*. Manchester: St. Jerome Publishing.
- Diaz Cintas, J., Orero, P., & Remael, A. (2007). *Subtitling for the Deaf, Audio Description, and Sign Language*. Amsterdam: Rodopi.
- Doens, E. (2000). *Kwalitatief Onderzoek van Ondertiteling voor Doven en Slechthorenden*. VRT Studiedienst.
- Eugeni, C. (2008). A Sociolinguistic Approach to Real-time Subtitling: Respeaking vs. Shadowing and Simultaneous Interpreting. *English in International Deaf Communication*, 72, 357-382.
- Gibson, E., & Levin, H. (1975). *Psychology of Reading*. Cambridge: MIT Press.
- Gielen, I., & d'Ydewalle, G. (1989). Hoe worden ondertitelde televisieprogramma's bekeken? *De Psycholoog*, 24, 425-431.

- Gottlieb, H. (1992). Subtitling. A New University Discipline. In C. Dollerup, & A. Loddegaard, *Teaching Translation and Interpreting. Training, Talent and Experience*. (pp. 161-170). Amsterdam: John Benjamins Publishing Company.
- Ivarsson, J. (1992). *Subtitling for the Media: A Handbook of an Art*. Stockholm: Transedit.
- Jensema, C. J., Danthurti, R. S., & Burch, R. (2000). Time Spent Viewing Captions on Television Programs. *American Annals of the Deaf*, 145 (5), 464-468.
- Jones, R. (2002). *Conference Interpreting Explained*. Manchester: St. Jerome Publishing.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 98, 122-149.
- Kondo, H., & Osaka, N. (2004). Effects on concreteness of target word on verbal working memory: An evaluation using Japanese version of reading span test. *Japanes Journal of Psychology*, 71, 51-56.
- Kothari, B. (2008). Let a Billion Readers Bloom: Same Language Subtitling (SLS) on Television for Mass Literacy. *International Review of Education*, 54, 773-780.
- Kothari, B., Pandi, A., & Chudgar, A. R. (2004). Reading Out of the 'Idiot Box': Same-Language Subtitling on TV in India. *Information Technologies and International Development*, 2 (1), 23-44.
- Lambourne, A., Hewitt, J., Lyon, C., & Warren, S. (2004). Speech-based Real Time Subtitling Services. *International Journal of Speech Technology*, 7, 269-279.
- Leijten, M. (2007). *Writing and Speech Recognition: Observing error correction strategies of professional writers*. Utrecht: Lot publications.
- Leijten, M., & Van Waes, L. (2006). Inputlog: New perspectives on the logging of on-line writing processes in a Windows environment. In K. P. Sullivan, & E. Lindgren, *Computer Key-Stroke Logging and Writing: Methods and Applications* (Vol. 18, pp. 73-94). Oxford: Elsevier.
- Liu, M., Schallert, D. L., & Carroll, P. J. (2004). Working memory and expertise in simultaneous interpreting. *Interpreting*, 6 (1), 19-42.
- May, P. C., Hasher, L., & Kane, J. M. (1999). The Role of Interference in Memory Span. *Memory & Cognition*, 27 (5), 759-767.
- Mizuno, A. (2005). Process Model for Simultaneous Interpreting and Working Memory. *Meta: Translators' Journal*, 50 (2), 739-752.
- Neves, J. (2008). 10 fallacies about Subtitling for the d/Deaf and the hard of hearing. *The Journal of Specialised Translation* (10), 128-143.
- Neves, J. (2005). *Audiovisual Translation: Subtitling for the Deaf an Hard-of-Hearing*. Roehampton University. London: University of Surrey-Roehampton.
- Ofcom. (2005). Subtitling - An issue of speed? London.
- Ofcom. (2006). *Television Acces Services. Review of the Code and Guidance*. Opgehaald van Ofcom: <http://www.ofcom.org.uk/consult/condocs/accessserv/access.pdf>
- Orero, P., & Looms, P. O. (2008). *Pilot of Mature Digital Television Access Services*. Opgehaald van Brunel University West London: <http://dea.brunel.ac.uk/dtv4all/ICT-PSP-224994-D23.pdf>
- Osaka, M. (2002). *Nou no memo-cho: wakingu memori (Working memory: The sketchpad in the brain)*. Tokyo: Shinyosha.
- Osaka, M. (1994). Riidingu-supan no kenkyu 5: doujityuyaku tesuto to wakingu memori (Study of reading span 5: Test of simultaneous interpreting and the working memory). *Nihon-shinrigakkai dai-58-kai taikai ronbunhu (Proceedings of the 58th conference of Japan Psychological Society)*, (p. 710).
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59, 413-425.

- Rashbash, J., Steele, F., Browne, W., & Prosser, B. (2004). *A user's guide to MLwiN Version 2.0*. Opgeroepen op January 12, 2010
- Remael, A. (2004). Irak in het Vizier: Hedendaagse Internationale Politiek in Ondertitels. *Filter* , 11 (3), 3-11.
- Remael, A. (2007). Sampling Subtitling for the Deaf and the Hard-of-Hearing in Europe. In J. Diaz Cintas, A. Remael, & P. Orero, *Media for All* (pp. 23-52). Amsterdam: RODOPI.
- Robson, G. D. (2004). *The Closed Captioning Handbook*. Oxford: Elsevier.
- Romero-Fresco. (2009). More haste less speed: Edited versus verbatim respoken subtitles. *Vigo International Journal of Applied Linguistics VIAL* , 6, 109-133.
- Schilperoord, J., de Groot, V., & van Son, N. (2005). Nonverbatim Captioning in Dutch Television Programs: A Text Linguistic Approach. *The Journal of Deaf Studies and Deaf Education* , 10 (4), 402-416.
- SOAP. (2007). Opgehaald van Samenwerkingsverband Ondertitel Alle Programma's: http://www.nvvs.nl/soap/_index.php
- Team, R. D. (2008). *R: A language environment for statistical computing*. Opgehaald van R Foundation for Statistical Computing: <http://www.r-project.org>.
- Towse, J. N., & Houston-Price, C. M. (2001). Reflections on the Concept of the Central Executive. In J. Andrade, *Working Memory in Perspective*. Hove: Psychology Press.
- Van den Noort, M. W., Bosch, M. P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist* , 11, 289-269.
- Van den Noort, M. W., Bosch, M. P., Haverkort, M., & Hugdahl, K. (2008). A Standard Computerized Version of the Reading Span Test in Different Languages. *European Journal of Psychological Assessment* , 24 (1), 35-42.
- Van der Veer, B. (2007). De Tolk als Respaeker: een Kwestie van Training. *Linguistica Antverpiensia* , 315-328.
- VRT. (2009). *Terugblik. Een historische schets van Teletekst Ondertiteling*. Opgehaald van VRT: www.vrt.be/vrt_master/888/PDF/888_terugblik2010_voorjaar.pfd
- Ward, P., Wang, Y., Paul, P., & Loeterman, M. (2007). Near Verbatim Captioning versus Edited Captioning for Students Who Are Deaf or Hard of Hearing: A Preliminary Investigation of Effects on Comprehension. *American Annals of the Deaf* , 125 (1), 20-28.
- Wechler, D. (2000). *WAIS-II Nederlandse bewerking: Afname- en scoringshandleiding [WAIS-III Dutch version: User manual]*. Lisse, The Netherlands: Swets and Zeitlinger.

APPENDICES

Appendix A – Raw scores

Table A. Estimated parameters (with standard error of estimate and confidence interval) for the reduction percentage of the Zero model (N=468) and the Net zero model (N=405).

	Reduction percentage	
	Zero model	Net zero model
	Est. (SE)	Est. (SE)
Fixed Part		
Intercept	64.852 (1.675)	59.661 (1.079)
Condition 2	-	6.337 (1.185)
Condition 3	-	8.579 (1.229)
Mean delay	-	0.373 (0.105)
Number of words in spoken comment	-	0.140 (0.018)
Percentage 100% reduction	-	0.440 (0.027)
	Est. [HPD95]	Est. [HPD95]
Random Part		
1' interval level variance	77.521 [4.372,5.802]	16.213 [1.754,2.881]
Participant level variance	24.730 [2.739,6.294]	2.944 [0.780,2.824]
Residual variance	38.451 [7.295,8.565]	19.909 [4.777,5.668]
-2 log(lh)	2769561.64	1598201.64
p-value		≤ 0.001

Note. Est. = parameter estimate, SE = standard error, HPD95 = 95% confidence interval

Appendix B – Example of merged utterances.

Table B. Example of merged utterances.

	Spoken comment (TU)	Subtitle	Reduction percentage	Delay
(1)	Ik, ik denk dat, dat bijna Bobbejaanland een noodzakelijke voorwaarde was om dit te kunnen maken, [I, I think that, that Bobbejaanland almost was a necessary condition to realize this]	Denk zelfs dat Bobbejaanland een grote voorwaarde was om dit te kunnen maken [Even think that Bobbejaanland was a large condition to realize this]	18,75	0:00:18
(2)	dus ik ben heel blij dat we, dat we dit nu hebben kunnen doen samen.	-	100	

[So, I am very glad that we, that we could have done this together now]

(3)	En we zijn alleen nog dichterbij mekaar gekomen, dat kan ik in ieder geval zeggen. [And we only have come closer together, that can I say, in any case.]	-	100
Total			72,92 0:00:18

Appendix C – Raw scores

Table C. Estimated parameters (with standard error of estimate and confidence interval) for the reduction percentage for the reduction percentage of the Zero model (N=9386) and the Net zero model (N=3701).

	Reduction percentage	
	Zero model	Net zero model
	Est. (SE)	Est. (SE)
Fixed Part		
Intercept	65.759 (1.657)	16.638 (1.887)
Condition 2	-	6.907 (1.917)
Condition 3	-	7.962 (1.916)
Mean delay	-	-1.745 (0.109)
Number of characters in spoken comment	-	0.643 (0.017)
	Est. [HPD95]	Est. [HPD95]
Random Part		
Minute id level variance	58.165 [6.145,8.819]	38.665 [4,615,7.125]
Participant level variance	24.667 [3.097,8.211]	14.940 [2.193,7.353]
Residual variance	1808.027 [41.909,43.148]	754.662 [26.916,28.165]
-2 log(lh)	2362251609	308775184
p-value		≤ 0.001

Note. Est. = parameter estimate, SE = standard error, HPD95 = 95% confidence interval

CONTACT

Luuk Van Waes
University of Antwerp
Department of Management
Prinsstraat 13
BE 2000 Antwerp, Belgium
luuk.vanwaes@ua.ac.be