Subtitling for the Deaf and Hard of Hearing:
A corpus-based methodology for the analysis of subtitles with a focus on segmentation and deletion

A thesis submitted to The University of Manchester for the Degree of
Doctor of Philosophy
in the Faculty of Humanities

2008

Dimitra Kalantzi

School of Languages, Linguistics and Cultures
The University of Manchester

Table of Contents

Final word count: 57,649 words (main text including footnotes and endnotes)

List of Tables

List of Figures

List of Abbreviations

AD – Audio description

AVT – Audiovisual Translation

BCI - Broadcasting Commission of Ireland

BSL - British Sign Language

DCMP - Described and Captioned Media Program

ECR - Extralinguistic culture-specific reference

ESIST - European Association for Studies in Screen Translation

HoH – Hard-of-Hearing

ILSP - Institute for Language and Speech Processing

IPTV – Internet Protocol Television

ITC - Independent Television Commission

MT – Machine Translation

MUSA - MUltilingual Subtitling of multimediA content

Ofcom – Office of Communications

POS - Part-of-speech

RAD - Royal Association for Deaf people

RNID - Royal National Institute for Deaf People

SDH – Subtitling for the D/deaf and HoH

SL – Sign Language

wpm – words per minute

See Appendices 4 and 7 for a list of the linguistic abbreviations used in this thesis.

Abstract

Subtitling for the D/deaf and HoH (SDH) is an invaluable service to millions of D/deaf and HoH viewers providing them with access to audiovisual media broadcast on television and increasingly in cinemas, DVDs and on the Internet. It is also a clear example of an AVT activity where the requirement to know and understand the intended addressee, a requirement present in all types of translation, is all-important. Or is it? Subtitles today still purport to be suitable for all, when by all accounts not only are the D/deaf and HoH two different groups, but what is more important within each group members differ not only in outlook, but also in language and communication needs. The time now seems opportune to answer crucial questions and identify the addressees and their needs/preferences. In order to provide an adequate reply to such questions, a tripartite large-scale research plan is deemed essential. This would combine i) corpus-based research describing current subtitle output with ii) empirical research into the needs and preferences of D/deaf and HoH people and iii) compare both with existing guidelines and update them as necessary.

This thesis has focused on the first and to a lesser extent third steps of the tripartite research plan sketched above. Accordingly, we will suggest characteristics to measure and metrics to use in order to describe subtitles. These will then be used to analyse a small sample corpus of BBC documentaries and current affairs programmes. The research is by no means large-scale, yet the aim is to provide a blueprint for large-scale research of the type advocated above. Some preliminary findings will also be reported in the belief that however tentative, they may provide significant initial data for developing relevant hypotheses. In addition, a new typology of deletions which is believed to be more intuitive than the one by Kovačič (1992, 1994, 1998) will be suggested. This will also be applied to the corpus and we will also investigate whether the deletion of oral features is as prominent in intralingual SDH and the genres covered by the sample corpus as it has been reported to be by previous interlingual subtitling studies focusing on films and TV series. Moreover, we will propose a comprehensive and updated list of linguistic recommendations regarding segmentation based both on the present corpus and the recommendations currently scattered in existing guidelines. These will also be presented in the form of linguistic rules in the belief that they could be further elaborated and used to either update existing automatic segmentation tools or to design a new tool that would alert subtitlers working off-line by flagging potential segmentation problems in subtitles.

University: The University of Manchester

Name: Dimitra Kalantzi

Degree Title: Doctor of Philosophy

Thesis Title: Subtitling for the Deaf and Hard of Hearing: A corpus-based methodology for the analysis of subtitles with a focus on segmentation and deletion

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

i) The author of this thesis (including any appendices and/or schedules to this thesis) own any copyright in it (the "Copyright") and s/he has given The University of Manchester the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

ii) Copies of this thesis, either in full or in extracts, may be made **only** in accordance with the regulations of the John Rylands University Library of Manchester. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

iii) The ownership of any patents, designs, trade marks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions") , which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

iv) Further information on the conditions under which disclosure, publication and exploitation of this thesis, the Copyright and any Intellectual Property Rights and/or Reproductions described in it may take place is available from the Head of the School of Languages, Linguistics and Cultures (or the Vice-President) and the Dean of the Faculty of Life Sciences, for Faculty of Life Sciences' candidates.

Acknowledgments

This thesis would not have been completed if it were not for the constant support and encouragement of my family, friends, professors and colleagues. I would like to express my sincere thanks and gratitude to:

- My supervisor Harold Somers for his help, patience and guidance while I was working on this thesis.

- All the people who have offered their unconditional help whenever I needed it, in particular Fotios Karamitroglou, Joselia Neves and Carlo Eugeni.

- The Greek partners of the MUSA project, in particular Mr Stelios Piperidis from ILSP and Manos Constantinou from Lumiere Cosmos Communications for providing me with the research material.

- My friends in Greece but also in the UK for not forgetting me and for being there for me through the good times and the (many) bad. I wouldn't have made it without their love and support. In particular, I would like to say a big thank you from the bottom of my heart to Eva, Emmanouela, Eftihia, Maria, Xaroula, Zoi and above all, Leda. I wouldn't have made it without you, *myfriend*.

- The Bakalas Foundation in Greece for its financial support during the first two years of this research.

- My baby niece for making me smile through the last difficult months of completing the thesis.

- Most of all, I would like to thank my parents, Panagiotis and Vassiliki, and my sister, Nicole for being there for me. Their support, patience and love have been invaluable.

1. Introduction

The importance of audiovisual media in our lives cannot be underestimated. Audiovisual media, in particular television, are powerful tools that inform, educate and entertain us. Their role in our everyday personal, social, political and cultural lives has always been vital and will no doubt continue to be so in the current era of global mass communication witnessing the proliferation of television networks and channels often transmitting to several countries, the advent of digital television, the implementation of new technologies, such as Internet Television, IPTV (Internet Protocol Television) and Mobile Television, and the millions of videos available on Internet sites, such as the BBC iPlayer, Hulu.com (in the US) or on a more informal level Youtube.com.

Within this context, the importance of audiovisual language transfer in providing everyone, including D/deaf[1] and Hard-of Hearing (HoH), as well as blind/visually impaired audiences, with access to television comes as no surprise. Indeed, accessibility is today one of the keywords in screen translation (Gambier, 2003, p. 179) and in Europe, it has now "finally become an explicit issue of debate in the European Union context" with legislation setting out requirements and benchmarks (Neves, 2007b, p. 260). Although many countries are only now beginning to catch up with the need to invest in media accessibility, the visibility the issue has gained both socially and research-wise can only make us optimistic about the future.

In this thesis, we will be concerned with D/deaf and HoH viewers and of the possible solutions currently available to cater for their needs, we will focus on subtitling rather than sign language interpretation. At first, subtitling for the D/deaf and HoH (SDH) was only intralingual, i.e. only involved language transfer from oral to written within the same language, but more recently, interlingual SDH is becoming

---

[1] Following convention, *deaf* is used to denote people with severe to profound hearing loss; *Deaf* (with capital D), on the other hand, is used to denote people who belong to the Deaf community. The term *D/deaf* is used to encompass both.

increasingly common particularly in the DVD market (see section 2.3.1). [2] Here, we will focus on intralingual SDH, although many of the notions we will refer to are equally applicable to both interlingual SDH and subtitling for hearing viewers.

So far, SDH practice seems to have placed greater emphasis on the issue of quantity setting targets for increasingly higher percentages of subtitling. But recently there have been increasing calls for greater attention to be paid to the issue of quality, as well. This is particularly true in countries such as the UK, where SDH has had a long tradition since the early 1980s and 100% subtitling has now been attained by the BBC, while interim targets for 2009 have been set to 89% and 80% for ITV/Channel 4 and Channel 5 respectively[3]. Among the essential ingredients in the quest for higher-quality SDH, two appear particularly prominent: i) the need to identify who the addressees are and ii) the need to identify their needs and preferences. Indeed, SDH is a clear example of an audiovisual translation (AVT) activity, where the requirement to know and understand the addressee, a requirement present in all types of translation, becomes all important. Both aspects, nevertheless, seem to have been largely neglected so far. Regarding the need to define the addressees, although the previous distinction between congenitally deaf people and people with acquired hearing loss has been superseded by the tripartite distinction between Deaf, deaf and HoH people, subtitling practice still lags behind with current subtitles still purporting to be suitable for all D/deaf and HoH viewers. Regarding the identification of viewers' needs and preferences, on the other hand, relevant research so far has been too fragmentary and restricted.

In order to reply to the fundamental questions set out above (*Who are the addressees?* and *What do they need/want?*), it seems that we would first need to:

1) carry out corpus-based research in order to provide a detailed description of the characteristics of current subtitle output, and

---

[2] In the USA, the term *captioning* is used instead of *subtitling* to refer to intralingual SDH. Here, like Neves, we will use the term SDH to refer to any type of captioning/subtitling solution that has been designed to cater for viewers with some degree of hearing loss (2007b, p. 255) and will distinguish between intralingual and interlingual SDH.

[3] For the relevant guidance, see Ofcom (2008) and Ofcom (2008b).

2) carry out empirical research involving D/deaf and HoH people in order to provide an account of how particular subtitle characteristics identified in (1) are related to viewers' response and how this response varies among different groups.

After thus identifying the addressees and their needs/preferences, the next research step would be to:

3) compare research findings with established guidelines and amend guidelines as necessary.

In order for findings to be reliable, it is imperative that such research is conducted on a large scale covering a large number of genres, programmes, subtitles and characteristics, as well as a large number of D/deaf and HoH people, relevant associations, broadcasters, subtitlers, subtitling associations such as the European Association for Studies in Screen Translation (ESIST), researchers and universities. In addition, in order for research findings to inform subtitling practice, there is little doubt that more than amending guidelines would be required (e.g. defining who the intended audience is, informing student and professional training, updating existing subtitling software and devising new tools). Nevertheless, this tripartite research plan is deemed an essential foundation for the pursuit of high-quality subtitling.

1.1 Research aims

Beginning with the tripartite research plan sketched above, the present research has focused on the first and to a lesser extent third steps. More specifically, regarding the first step, i.e. the description of current subtitle output, this thesis has **set out to suggest characteristics to measure and metrics to use in order to describe current subtitles.** This will be used to analyse a small corpus of BBC documentaries and current affairs programmes provided by the MUSA project.[4] The research is by

---

[4] The MUSA (MUltilingual Subtitling of multimediA content) project was a joint EU-funded project by the BBC (London), Systran (Paris), Universiteit Antwerpen, Katholieke Universiteit Leuven, the

no means large-scale, as it is restricted to two closely-related genres and a limited number of subtitles and characteristics. Yet, the main aim is to provide a blueprint for research of the type advocated above. In addition, some preliminary findings will be reported in the belief that however tentative they might be, they may provide significant initial data for developing relevant hypotheses.

Of the many characteristics discussed, particular emphasis has been paid to deletion and segmentation within subtitles. The choice has not been accidental. Regarding deletion, there are two reasons:

i) The question of whether subtitles should be verbatim, i.e. an exact and complete transcription of speech, or whether they should be edited (and deletion seems to be the most common editing strategy, see section 5.7.1) is a hotly debated and highly charged issue in SDH (see section 2.4.6). Nevertheless, very little research has been dedicated to describing this particular aspect of SDH. Are current subtitles verbatim? Are they edited? If so, how much editing is done? The only research that has been undertaken so far in Europe seems to be the one by De Linde (1997) and De Linde and Kay (1999), where findings are based on the analysis of 262 subtitles from a total of twelve programmes belonging to six different genres (see sections 3.3.1 and 3.3.5).

ii) Although deletion has been regarded as one of the defining features of subtitling and appears to be one of the phenomena most often mentioned in the subtitling literature (see section 2.4.6), it would seem that there is still no satisfactory way to classify deletions. The dominant classification suggested by Kovačič (1992, 1998) where deletions are categorised according to the linguistic function of their language sign (see also sections 2.4.6 and 4.3.1.10 of the present thesis), however useful it

---

Institute for Language and Speech Processing (ILSP) (Athens) and Lumiere Cosmos Communications (an Athens-based subtitling company). The main objective of the project, which run from 2000 until 2004, was the development of a system which combined together advanced text analysis, speech recognition, machine translation and other techniques to help in the preparation of subtitles. Only a prototype was made with demos available online at http://sifnos.ilsp.gr/musa/demos (last accessed on October 28th 2008).

might be, involves terms that are not intuitively clear (*ideational*, *interpersonal* and *textual*) and which might not be readily accessible to subtitlers.

For these reasons, regarding deletion, the **aim** of the present thesis is not only to provide relevant metrics, but also to suggest **a new typology of deletions**. These will then be used to analyse the research corpus and report initial findings. In addition, given that previous interlingual studies focusing on films and TV series have reported that the usual candidates for deletion are oral features, we will also investigate whether this finding is applicable to intralingual SDH and the genres covered by the present research corpus.

Regarding segmentation within a subtitle, the decision to focus on this aspect was dictated by the material available. That is, it soon became evident that the subtitles analysed contained a high proportion of line-break errors suggesting that poor (in terms of linguistics) segmentation is an issue when subtitling. Accordingly, this thesis has set out to **provide a list of the types of errors encountered** in the research corpus, as well as a **tentative account of the factors** that seem to influence the subtitlers' output when producing a subtitle with improper line breaks.

More importantly, it appears that i) there are no guidelines that encompass all recommendations currently available and ii) the general recommendation currently available to segment at the highest syntactic node possible, though powerful, does not always give the best result (see section 6.1). For these reasons, this thesis has also set out to compare findings from the present research corpus with existing guidelines thus entering the third step of the tripartite research plan outlined above. More specifically, the **aim** is to **propose a comprehensive and updated list of linguistic recommendations based both on the present corpus and the recommendations currently scattered in existing guidelines**. Such a list could then become a point of reference regarding the linguistic aspects of segmentation for both students and subtitlers, as well as for further research. In addition, these recommendations will also take the form of linguistic rules that could be elaborated on and used to update existing segmentation tools or to design a new tool.

1.2 Thesis Structure

To conclude this chapter, a summary of the chapters in the remainder of the thesis is provided as follows:

➢ Chapter 2 presents an overview of AVT, the typology of subtitling adopted for the purposes of the present thesis, as well as some of the general characteristics of subtitling relevant to this research. The discussion then turns to SDH, which is the focus of the thesis. The theoretical view of D/deaf and HoH viewers from a subtitling perspective is described and some basic information on hearing impairment is presented on the basis of the literature on Deaf studies. This information reveals the heterogeneity of the D/deaf and HoH audience, which is largely ignored by subtitling practice, where subtitles seem to be produced with an indeterminate audience in mind. The need to determine the audience, in particular the intended SDH audience is stressed and the tripartite research plan involving description of subtitle characteristics, identification of users' needs/preferences and comparison to existing guidelines is put forward.

➢ Chapter 3 discusses existing guidelines and empirical subtitling research relevant to the subtitling characteristics covered in this thesis. These include the duration of subtitles, the time between consecutive subtitles, the maximum number of lines, segmentation within a subtitle, reduction and the overall strategies used by subtitlers. Guidelines presented come both from the area of SDH, as well as from that of interlingual subtitling for hearing viewers.

➢ Chapter 4 provides a brief overview of corpora in translation and AVT in particular, while arguing for the necessity to adjust the current corpus typology so that intralingual SDH corpora are accommodated. The corpus used in this research is then described including its limitations, and the analysis process and some basic characteristics to measure and categories/metrics to use when describing subtitles are suggested. This includes a new typology for discussing deletions.

➢ Chapter 5 presents and discusses some preliminary results obtained from the analysis of the present corpus using the categories and metrics suggested in Chapter 4. Findings are also compared with any previous research available and tentatively related to existing guidelines. Particular emphasis is placed on the issue of deletion.

➢ Chapter 6 focuses exclusively on segmentation within subtitles from a linguistic perspective. Relevant findings from the present corpus are reported and then compared to previous research and related to existing guidelines. Given that current guidelines are found lacking in certain respects, an updated set of recommendations is suggested based both on our research corpus and on the recommendations currently scattered across available guidelines. This updated set of guidelines is envisaged to be helpful both to professional and trainee subtitlers. In addition, guidelines are also expressed in the form of linguistic rules that could be elaborated on and used to update existing segmentation tools or to design a new tool.

➢ Finally, in Chapter 7, the major findings and suggestions of this research bearing in mind limitations are briefly recapitulated and suggestions for further research are presented.

2. Subtitling

Up until at least the early 1990s and despite its quantitative importance, very little research had been devoted to subtitling and audiovisual translation in general (Delabastita, 1989, p. 193; Díaz Cintas, 1999, p. 31; Díaz Cintas, 2004, p. 50). Nevertheless, for years subtitling had been one of, if not the most widely used type of translation reaching a high number of people on a daily basis. The main reason for this paradoxical state of affairs seems to have been the fact that translation studies largely ignored subtitling and some researchers even questioned its status as a type of translation (Shuttleworth and Cowie, 1997, p. 161). On a wider scale, the neglect for audiovisual translation was symptomatic of the low status of translation itself within linguistic and literary studies and a result of their emphasis, copied by translation, on high prestige topics and genres (Delabastita, 1989, p. 193). However, after years of efforts, subtitling has now come to be regarded as "the most important translational activity of our time" (Díaz Cintas, 2004, p. 50) and is widely researched, as evidenced by the high number of publications, conferences and courses available.

After providing a definition of subtitling and a typology of audiovisual translation, this chapter presents a classification of subtitling according to a variety of criteria, briefly discusses some of the major characteristics of subtitling and then focuses on subtitling for D/deaf and HoH viewers (SDH), which is the focus of this thesis and "undoubtedly one of the forms of audiovisual communication which is developing most at present" (Díaz Cintas and Remael, 2007, p. 14).

2.1 Definition

Subtitling may be defined as the process of providing synchronised, often condensed written translations/transcriptions of the original dialogue and soundtrack of films and television programmes, and more recently live opera and theatre productions, games and videos broadcast on the Internet. In the case of interlingual subtitling, it

also "endeavours to recount ... the discursive elements that appear in the image", such as letters, inserts, graffiti, inscriptions, placards, etc. (Díaz Cintas and Remael, 2007, p. 8). In the case of SDH, whether intralingual or interlingual, information is also provided regarding paralinguistic and extralinguistic features of the soundtrack, such as tone of voice, music, sounds and sound effects, as these too contribute, sometimes crucially, to the atmosphere and full meaning of the programme. In addition, information regarding speaker identification is often essential so that viewers can identify who is saying what. Along with dubbing, subtitling is one of the two main types of audiovisual translation that exist today.

2.2 Typology of Audiovisual Translation

Drawing on the typologies suggested by Gambier (1994, p. 276; 2003, p. 172-177), the following types of AVT can be identified as presented in Table 1:

Table 1. Typology of Audiovisual Translation

Types of Audiovisual Translation		
Written	Oral	Visual-gestural
Subtitling	Dubbing	
Surtitling	Voice-over	
	Narration	
	Commentary	
	Interpreting	
	Sight translation	
Subtitling for the Deaf and HoH	Audio description	Sign language interpretation

A short definition of each category, apart from sign language interpretation, which is rarely mentioned in this context, is provided by Gambier (ibid). Focusing on

subtitling, which is the topic of this thesis, it is important to point out that whereas in his first typology, Gambier only suggests one subtitling category (and that is presumably interlingual subtitling for hearing viewers), in his second typology, a distinction is drawn between interlingual subtitling for any audience, including D/deaf and HoH viewers, and intralingual SDH. For reasons analysed in sections 2.3.1 and 2.3.2, I have instead chosen to distinguish subtitling primarily in terms of audience rather than on linguistic grounds.

Moreover, following Díaz Cintas and Remael, I have opted for the term "audiovisual translation" (AVT), which is becoming the standard referent in recent years, to describe the field (2007, p. 12). Admittedly, this term is not readily applicable to non-prototypical audiovisual activities, such as intralingual SDH and audio description (AD) for the blind and partially sighted, unless a broader definition of translation is adopted. As Díaz Cintas and Remael themselves acknowledge, "these new practices [SDH and AD] have brought in further terminological disarray, especially because of the fact that none of them ... implie[s] the transfer from a source to a target language, one of the traditionally defining features of any translation activity" (ibid). Jakobson's (1959) famous taxonomy of translation might be of help in this matter, as it would allow us to categorise intralingual subtitling for D/deaf and HoH, as well as hearing viewers as a type of intralingual translation. However, its contribution seems to be of limited value in the case of AD. The main goal of AD is to "provide information of all the visual elements – action, costumes, settings, images – of theatre, television/film, museum exhibitions, and other events" (Audio Description Associates brochure, no date); in this context, verbal language in the form of text shown in a scene or shown as subtitles constitutes only a small proportion of the types of information relayed. Thus, in order to include AD in our taxonomy, we would first need to extend Jakobson's definition of intersemiotic translation ("an interpretation of verbal signs by means of signs of nonverbal sign systems") (Jakobson, 1959) so that it becomes bi-directional. We would then be able to categorise AD, which involves interpreting a mainly non-verbal non-linguistic source by means of verbal signs, as a type of intersemiotic translation.

2.3 Typology of Subtitling

Many different types of subtitling have been suggested in recent years. Drawing from previous studies, mainly Gottlieb (1998), Díaz Cintas (2001), Bartoll (2004) and Díaz Cintas and Remael (2007), the following typology has been adopted in this thesis.

2.3.1 Linguistic parameters

Linguistically, subtitles can be interlingual, intralingual or bilingual. Interlingual subtitling involves language transfer from oral to written between two different languages, whereas intralingual subtitling involves language transfer from oral to written within the same language. Bilingual subtitling like interlingual subtitling involves different languages, however, in this case language is transferred from one language into two different languages simultaneously. This type of subtitling is less frequent and is used in countries where two languages are used, such as Finland, Belgium and Israel, or in international film festivals (Díaz Cintas and Remael, 2007, p. 18).

Traditionally, interlingual subtitling has been associated with hearing viewers who want to watch a foreign language programme, while intralingual subtitling has been associated with D/deaf and HoH audiences, who want to watch a programme in the language of the country in which they live. Although this distinction might have been accurate in the past, it failed and still fails to acknowledge the following facts:

- Like all viewers, D/deaf and HoH people also wish to view foreign language programmes.
- Interlingual subtitles for hearing viewers are not appropriate for D/deaf and HoH people, as they are not tailored to their specific needs. In particular, not only are subtitle rates too high, but what is more important, only the linguistic features of the soundtrack are transferred (De Linde and Kay, 1999, p. 1). Moreover, no information is provided regarding the remaining elements of

the soundtrack (e.g. sound effects, music) or speaker identification, both beneficial to D/Deaf and HoH viewers.

More importantly, this broad classification is far from accurate today, as it fails to account for professional practice itself. Not only has interlingual subtitling for D/deaf and HoH viewers been a reality for several years, but it is now gaining greater visibility thanks to the DVD market (Díaz Cintas, 2004, p. 200). In addition, although intralingual subtitling is mainly used by D/deaf and HoH audiences, it is also employed to cater for a diverse set of hearing people and for various purposes. Thus, it can be used by hearing viewers, such as immigrants and students learning a foreign language, children learning their mother tongue or anyone watching a broadcast in noisy environments, such as airports and pubs.[5] In the case of karaoke, it can also be used to subtitle songs or musicals so that hearing people can sing along. Occasionally, it is also used to subtitle parts of programmes (and more rarely entire programmes) whenever people speak the language with a strong, local or foreign accent so that they may be understood by viewers speaking that language (Díaz Cintas and Remael, 2007, p. 17).

Given this variation, it seems essential to introduce a further classification in terms of audience to be used complementary to the linguistic classification.

2.3.2 Audience

This parameter distinguishes between subtitling for hearing viewers and subtitling for D/deaf and HoH people (SDH). It brings to the fore the different needs of the D/deaf and HoH audience (for more on this, see section 2.5), while allowing us to override the linguistic classification of subtitles. Under this categorisation, therefore, each type encompasses both interlingual and intralingual subtitling reflecting professional practice in a more accurate manner. The different uses of each type were mentioned in section 2.3.1. Intralingual SDH, which is the focus of this thesis, and to

---

[5] In fact, it would seem that in the UK, the vast majority of viewers have no hearing impairment at all (http://www.guardian.co.uk/media/2006/mar/27/broadcasting.uknews1?gusrc=rss&feed=global , last accessed on October 28th 2008); the same seems to be the case in the USA (Erard, 2001).

a lesser extent interlingual SDH will be further elaborated in section 2.5. Suffice it here to mention that both subtitling for hearing viewers and SDH can cater for a wide audience, which is far from homogeneous.

2.3.3 Time

Along this parameter, a distinction is drawn between pre-prepared (off-line), live (real-time) and semi-live subtitles. Pre-prepared subtitles are produced in advance and the subtitler is usually, though not always, provided ample time to carry out their work (Díaz Cintas and Remael, 2007, p. 19). On the other hand, live subtitles are prepared and cued live, while the programme is transmitted. They are used for live programmes, such as the news, live interviews and political statements, although they might also be added to pre-recorded programmes (Softel, 2001, p.1). Finally, semi-live subtitling refers to the practice of combining live with pre-prepared subtitles. Given that many programmes are not totally live, it is indeed possible to integrate pre-prepared subtitles for the pre-edited segments of the programme, although they too will have to be cued live during transmission.

Undoubtedly, live and to a lesser extent semi-live subtitling are very demanding processes. They are both primarily associated with intralingual SDH, although they are also occasionally produced interlingually for the benefit of hearing viewers (see for instance Den Boer, 2001). Intralingual live SDH has been traditionally produced by a stenographer or velotypist (Ivarsson and Carroll, 1998, p. 134). More recently, a new approach called respeaking, which involves the use of speech recognition software, has been gaining ground. In the words of Eugeni, "respeaking is the production of real time subtitles by means of speech-to-text technology transcribing the vocal input of the operator who repeats or reformulates the audiovisual source text" (forthcoming, no page). In this context, the subtitler, rather like a simultaneous interpreter, listens to the source text, processes it in his/her mind and utters (respeaks) the subtitle, which is then transcribed by the speech recognition software.

On a similar note, but on an experimental level, some researchers have also investigated the potential of automatic speech recognition in the form of a fully-automatic system that would transcribe the programme audio directly into the respective intralingual subtitles (see the VOICE project[6]; Piperidis et al., 2004; Daelemans et al., 2004). Still more efforts seem to have been dedicated to the potential of Machine Translation (MT) for the production of live interlingual subtitles. Several projects have been undertaken, such as MUSA (see Piperidis et al., 2004; Tjong Kim Sang et al., 2004), eTITLE[7] (see Melero et al., 2006), ALTo (see Toole et al., 1998; Popowich et al., 2001) and the project by Nyberg and Mitamura (1997). What is even more interesting and perhaps surprising in this case is the fact that not only has MT been researched, but it has also has been commercially used for subtitling purposes for some time now in the USA.[8] Nevertheless, as Díaz Cintas and Remael point out, as far as the particular system is concerned,

> this approach seems to be driven solely by economic forces and interests... Quality seems to be disregarded altogether and the results are still a long way from being satisfactory. ... examples shown on their website ... are not only lexically and syntactically incorrect in Spanish but also nonsensical... an attack on the Spanish language ... (2007, p. 21)[9]

---

[6] More information is available on the project's website http://voice.jrc.it/home_en.htm (last accessed on October 28th 2008)

[7] Like MUSA (mentioned in section 1.1), eTITLE was an EU-funded programme (from 2004 until 2006), which combined state-of-the art Machine Translation and Translation Memory, as well as other advanced technologies for automated speech-to-text, sentence compression and subtitle placement among others. Its goal was to produce tools to assist in the multilingual subtitling of audiovisual material through the web (Melero et al., 2006). Although there is a dedicated website advertising the final commercial version of the system (http://www.etitle.co.uk/index.php), this has still not become available (the website still states 2006 as the expected release date). In addition, the commercial version of the system is not envisaged to incorporate Machine Translation, as according to the relevant information "eTITLE does not automatically produce subtitles, but works alongside the subtitler helping them to work more effectively" http://www.etitle.co.uk/faq.php

[8] The system is called TranslateTV (TTV) and further information can be obtained online at www.translatetv.com (last accessed on October 28th 2008).

[9] Reference should also be made here to Sync-Now, an automated web-based service enabling transcript synchronization and/or captioning of audio/video content submitted by customers. Two captioning services, *Ready ToEncode Captions* and EZ *Captions*, are advertised on www.sync-now.com (last accessed on October 28th 2008). However, the latter service, advertised as an end to end solution for captions, is not yet available. In addition, there is no information regarding the type of language technology tools that might be used to produce the captions.

2.3.4 Technical parameters

In technical terms, subtitles can be either open or closed. Open subtitles are inseparable from the image, they are not optional and cannot be turned off. Closed subtitles, on the other hand, are hidden. In the case of television, they are broadcast separately and are only accessible to viewers who have a special decoding device, for instance a teletext decoder. In the case of DVD, the subtitles can be activated separately using the appropriate menu.

2.3.5 Medium

Subtitles can also be categorised according to the medium used for the distribution of the subtitled audiovisual product. At the moment, we can distinguish between cinema, television, video, DVD and Internet subtitles, as well as subtitles for computer games. Different media require different subtitling solutions. For instance, regarding the time a subtitle should remain on screen, television subtitles tend to be slower than subtitles for other media, "the reason being that the profile of the television viewer is in general more heterogeneous and the subtitles have to satisfy all viewers" (Díaz Cintas and Remael, 2007, p. 24). Other features influenced include the maximum number of characters per line, subtitle position on the screen (although there is now a trend for all media to centre subtitles), and line-break application (with cinema for instance preferring shorter subtitle lines).

It is important to note that the taxonomy presented in sections 2.3.1 to 2.3.5 is not complete. Other classifications exist distinguishing subtitles in terms of projection method, position on the screen (centred vs. non-centred), and mobility (mobile vs. fixed). Nevertheless, the taxonomy adopted here includes all major distinctions relevant to this thesis. In addition, we need to point out the somewhat transient nature of at least some of the classifications. For instance, as norms are changing, the distinction between centred vs. non-centred subtitles might disappear altogether. More importantly, the impact of technology has created new classifications or changed traditional classifications enriching them with new categories while making

others dated. Given the close relationship of subtitling with technology, this trend is more than likely to persist in the future.

2.4 General Subtitling Features

This section describes the main features of subtitling. Its purpose is not to present norms or guidelines or to refer to particular features of subtitling, such as font type or position on the screen, punctuation conventions, etc. Information on such features can be found in many publications, including Díaz Cintas and Remael (2007), Ivarsson and Carroll (1998) and Karamitroglou (1998). All features relevant to this research will be presented in detail in Chapter 3.

2.4.1 Accessibility

According to Gambier, "the keyword in screen translation is now accessibility" (2003, p. 179). In his discussion, the term is used to cover a variety of features, such as acceptability, readability and relevance, but along with Díaz Cintas and Remael (2007, p. 13), we will focus on its social dimension. As Neves points out, accessibility is the reason why subtitling and other forms of AVT actually exist (2005, p. 16). Indeed, the main purpose of AVT is to make an audiovisual programme accessible to viewers who would otherwise be deprived of the full message due to linguistic and/or sensorial problems.

The concept of accessibility undoubtedly entails a viewer-centred approach to AVT. In order to facilitate access, the needs of the audience must be taken into account so that different solutions are provided for different audiences. The requirement to know and understand the intended addressee, a requirement present in all types of translation, becomes all-important in AVT, particularly within the context of SDH, sign language interpretation and AD. Focusing on SDH, knowing the D/deaf and HoH and their needs better will enable both practitioners and researchers to better provide for such viewers analysing and adapting present practices and envisaging

new ways forward (Neves, 2005, p. 76). SDH and its audience will be discussed in detail in section 2.5.

2.4.2 Vulnerability

Subtitling is unusual in that it is an overt type of translation, presenting the end-product alongside the original (Gottlieb, 1994b, p. 102)[10]. This co-occurrence allows viewers to compare the subtitles with the original version and makes subtitling susceptible to criticism (ibid) or "vulnerable" to use the term suggested by Díaz Cintas (2001, p. 133).[11] This is particularly true in the case of interlingual subtitling for anyone with some knowledge of the original language, and in the case of intralingual SDH for viewers with some residual hearing or lip-reading skills. However, even viewers with no such skills can understand discrepancies between the subtitles and the original and become critical. This might happen, for instance, "cuando hay una clara desproporción entre la duración de los diálogos y los subtítulos o cuando la actuación enfadada del actor nos hace intuir una retahíla de exabruptos y los subtítulos ofrecen una versión muy eufemizada" (Díaz Cintas, 2001, p. 133).

2.4.3 Spatiotemporal constraints

This is one of the main features of subtitling. In the words of Marleau,

> Le traducteur ... est limité par le temps, car la durée de projection des sous-titres est toujours fonction de la durée du dialogue dit. La traduction est également soumise aux contraintes d'un espace limité. (1982, p. 275)

Luyken et al. further elaborate on these notions (1991, p. 42-43). On the one hand, the amount of space available for subtitles is severely limited by the physical

---

[10] Gottlieb has borrowed the terms *overt* and *covert* from House (1977).
[11] For a discussion of this issue, see also Gottlieb, 1994b; Karamitroglou, 2001; Díaz Cintas and Remael, 2007.

limitations of the size of the screen and the necessity of not obscuring the image. On the other hand, the amount of time available for displaying the subtitle depends on the pace of the soundtrack dialogue, the amount of text in the subtitle, the average reading speed of viewers and the need to keep a minimum interval between subtitles.

2.4.4 Polysemiotic nature

Audiovisual products are texts of polysemiotic nature, as they employ both the visual and acoustic channels to get the message across to their audience (Delabastita, 1989, p. 196). Given that both channels can be used to transmit both verbal and non-verbal signs, Delabastita proceeds to suggest four semiotic categories relevant to subtitling (ibid, p. 199); these are later taken up by Gottlieb:

1) the verbal audio channel: dialogue, background voices and lyrics

2) the non-verbal audio channel: music, natural sounds and sound effects

3) the verbal visual channel: captions and written signs in the image, such as letters, documents, road signs, place names, advertisements, etc.

4) the non-verbal visual channel: picture composition and flow

(1994, p. 265)

Subtitles constitute only a single part of the verbal visual channel and should thus "form an organic whole with the other components", which remain intact in order for viewers to fully understand and appreciate the audiovisual product (Luyken et al., 1991, p. 154). Although this dimension of subtitling is usually considered a constraint on the subtitlers' work, at times, positive feedback from the visual or audio channel can support them in finding an appropriate subtitling solution.

2.4.5 Conversion from speech into writing

Subtitling, whether interlingual or intralingual, differs from other types of translation in that it involves a shift in mode from speech into writing. Baker notes:

It is important to remember that there are significant differences between written and spoken languages. ... In TV subtitling we are trying to use a written medium to convey speech. Because of the differences in design features between written and spoken language, this is a novel and truly creative enterprise. (1981, p. 10)

Gambier takes this issue even further by stating that "avec l'AV, les frontières oral/écrit sont constamment brouillées" (2004, p. 170). For one thing, screen dialogue differs from everyday natural dialogue (Díaz Cintas, 2001b, p. 33; Remael, 2004, p. 105), as it is the result of continuous transformation, whereby "le scénariste conçoit les dialogues comme de la parole mais il les écrit; le metteur en scène lit le script ...; [et les acteurs] donnent une version orale" (Gambier, 2004, p. 171). However, despite its hybrid nature, screen dialogue is still considered to contain many, if not all of the characteristics of spoken language, characteristics that distance it from writing, such as repetitions, false starts, unfinished sentences, self-contradictions, interpersonal signals, etc. That is why many of these features are often omitted from subtitles, as they would be odd and distracting when transferred to the written mode (Díaz Cintas, 2001b, p. 33; Kovačič, 1998, p. 79; Gottlieb, 2005, p. 19; see also detailed discussion in section 2.4.6).

Furthermore, the nature of subtitles themselves as written language has also been questioned.

Subtitles are written, yet they are supposed to reflect spoken communication. ... This raises a theoretical question regarding the status of subtitles: are they an independent stratum of language, intermediate between speech and writing, or are they rather a sub-category of either of the two [?] (Kovačič, 1998, p. 79-80)

The blurring between speech and writing in subtitling is manifest in yet another way. According to one of the major situational parameters identified by Biber, writing is permanent and enables a prolonged visual inspection of the text so that the reader can

either take his/her time or skim an entire text in a few seconds; speaking on the other hand, is temporary, produced much faster and has to be understood by the listener at the time of production (1988, p. 42). Although subtitles are written and thus share the permanence of writing, at the same time, they also share the temporary nature of speech in that they only allow the viewer a few seconds to process each subtitle. The viewer, like a listener, and unlike a typical reader, must comprehend language as it appears in front of his/her eyes.

It would appear that no definite answers have been provided concerning both the extent to which screen dialogue resembles every day dialogue, and the issue of the written, spoken or independent nature of subtitles. More importantly, for both issues discussion seems to have unduly focused on scripted film dialogue. Díaz Cintas and Remael do mention that unscripted speech in interviews is different from both real-life conversation and film dialogue (2007, p. 64). Further research would be needed, however, to determine the extent to which these three modalities are different from and similar to each other. What would be even more welcome would be extending research to other forms of speech in audiovisual programmes, such as non-dialogic scripted speech as that found in documentaries and news, and non-dialogic unscripted speech as that which may be found in talk shows.

2.4.6 Reduction

Reduction has come to be regarded as one of the defining features of subtitling. It also appears to be one of the phenomena most often mentioned, even if only to be touched upon, in subtitling literature (see for instance Marleau, 1982; Delabastita, 1989; Luyken et al., 1992; Gottlieb, 1994; Brondeel, 1994; De Linde, 1995; Shuttleworth and Cowie, 1997; Ivarsson and Carroll, 1998; Gottlieb, 1998; Tomaszkiewicz, 1998; O'Connell, 1998; James, 1998; De Linde and Kay, 1999; Díaz Cintas, 2001; Georgakopoulou, 2003; Neves, 2005; Díaz Cintas and Remael, 2007). More importantly, reduction is one of the most discussed issues within the context of SDH; in particular, the question of whether subtitles should be verbatim, i.e. an exact and complete transcription of speech, or whether they should be

reduced/edited is a hotly debated and highly charged issue (for more on this, see subsection *Reduction and verbatim subtitling in SDH* and section 3.3.5). Given the prominence of reduction in the relevant literature, this section is relatively longer than others. Yet, the information presented here is no more than a brief overview of a limited number of issues which surround reduction and are relevant to this thesis.

What is reduction: Two types of reduction have been traditionally identified: total reductions (deletions/omissions) and partial reductions (condensations) (Kovačič, 1994, p. 247). According to Ivarsson and Carroll, deletion is usually better than paraphrase, as it is easier, less intrusive and thus less irritating to viewers with even a limited knowledge of the original language (1998, p. 86). More often than not, however, the two processes go hand in hand, as deleting part of the dialogue usually makes it necessary to paraphrase and condense the rest to a certain extent (Ivarsson and Carroll, 1998, p. 87; Diaz-Cintas, 2001, p. 124-126).

Occasionally, partial reductions have been identified with paraphrases rather than condensations (e.g. De Linde, 1995; Ivarsson and Carroll, 1998). Although this is true to a certain extent (many paraphrases are shorter than the expressions they replace), this is not always the case; for instance, quite often, lexical and syntactic structures are substituted by structures that may be simpler, yet are longer than the original. It might therefore be more accurate to view paraphrases as a separate category and only equate them with partial reductions when they actually involve condensation and reduction.

Why reduce:

Spatiotemporal constraints: The very nature of subtitling with its spatial and temporal constraints (see section 2.4.3) imposes a need for brevity and often makes complete transcription of the original dialogue impossible.

Average speech rate vs. average reading rate: The main reason why complete transcription of speech is often impossible is that the average speech rate is higher

than the average reading speed – we tend to talk much faster than we can read (Marleau, 1982, p. 276, 277). In the UK, "the average reading speed of adult viewers is estimated at 66% of the average speaking speed" and as a result, the dialogue has to be reduced by around one third (De Linde and Kay, 1999, p. 11). The need to condense subtitles becomes even greater with particularly fast speech; as Ivarsson and Carroll aptly put it,

> "Some people talk nineteen to the dozen with words tumbling out so fast that they manage to say in a few seconds three or four times as much as there is space for in the two subtitle lines below the picture… In such cases the dialogue obviously has to be condensed". (1998, p. 85)

Diasemiotic nature of subtitling: Subtitling necessarily involves the conversion of speech into writing, as the source-language (SL) speech becomes either SL writing (in the case of intralingual subtitling) or target-language (TL) writing (in the case of interlingual subtitling). In both cases, the transition from spoken to written language is more than an act of transcription. As was mentioned in section 2.4.5 and will be further explored later in this section, oral features are often either deleted or condensed when crossing over to writing. Quite often, such elements are considered redundant, while at times, transcribing them would impede smooth reading and comprehension.

Intrasemiotic and intersemiotic redundancy: Both intrasemiotic and intersemiotic redundancy further the tendency for reduction (Gottlieb, 1994, p. 273). Intrasemiotically, the inherent redundancy of speech (whether spontaneous or deliberate) often calls for reduction, whereas intersemiotically viewers may use information from the rest of the audiovisual channels (image, acting, sound and language) to supplement the content of the subtitle. This positive intersemiotic feedback helps in a way neutralise most of the effects of reduction so that viewers do not miss as much of the audiovisual programme as would be implied by a mere quantitative linguistic analysis (Gottlieb, 2005, p. 21).

<u>Need to appreciate the entire programme:</u> One of the main reasons why reduction is necessary has to do with the need to allow viewers adequate time to appreciate the whole audiovisual product. In polysemiotic texts (see section 2.4.4), successful reception depends on the viewers' ability to take in all communicative channels present. After all, watching television or a DVD or going to the cinema is about watching rather than reading.

<u>Nature of reading subtitles</u>: Reading subtitles on screen is different from conventional reading in the following important respects:

- Viewers' attention is divided between two different activities, as they have to both read the subtitles and watch the programme.
- Viewers have no control over the speed of reading, as subtitles are only available for a short time.
- Viewers cannot go back to any subtitle in the way they would be able to turn back to a sentence on the printed page.

Under these special circumstances, subtitles need to be constructed in a way that enables the audience to understand them without difficulty first time around and in the limited time they appear on screen. As a result and in conjunction with other factors mentioned in this section, it might often be necessary to reduce the original text.

<u>Need for extra space for paralinguistic and extralinguistic information:</u> In the case of SDH, subtitles have to provide access not only to the dialogue, but also to the remaining elements of the soundtrack (e.g. sound effects, music), as these too contribute, sometimes crucially, to the atmosphere and full meaning of the programme; they also have to provide additional information so that viewers may properly identify who is saying what. Although paralinguistic and extralinguistic information can often be provided on a third subtitle line, this might not always be possible. In addition, it may sometimes be best to keep the number of lines displayed to a minimum, hence the importance of deleting or condensing part of the dialogue.

Last but not least, we might point out that it seems peculiar to use verbatim subtitles for D/deaf and HoH viewers when studies of interlingual subtitling for hearing viewers mention figures of quantitative losses of up to 40% (see section 3.3.5). As Gambier notes, it appears "as if the hearing impaired ... [are] assumed to read faster than "ordinary" viewers" (2003, p. 176).

What to reduce: According to the relevant research available, the usual candidates for reduction/deletion are interpersonal elements and oral features.

Oral features: Due to the diasemiotic nature of subtitling, oral features of the dialogue tend to be thought of as redundant and are either deleted or condensed when crossing over to writing. Such oral features include the following:

- pauses, hesitations, false starts, repetitions, self-corrections and interruptions
- discourse markers and pragmatic connectors (e.g. *you know*, *well*)
- unfinished sentences and 'grammatically unacceptable' constructions
- slips of the tongue, self-contradictions, ambiguities and nonsense
- overlapping speech

(adapted from Gottlieb, 2005, p. 18)

Since these features are all characteristic of spoken language, they are usually the first to be omitted when subtitling (De Linde, 1995; Karamitroglou, 1998; Ivarsson and Carroll, 1998; De Linde and Kay, 1999; Kovačič, 1998b). However, as De Linde (1995) and Kovačič (1998b) point out, although these elements of speech appear redundant and thus perfectly omittable, they might be integral to a character's style of discourse and/or the portrayal of their character. In addition, trimming such features off the discourse has the effect of "normalizing the text, by presenting the … audience with a version less non-standard than the original" (Gottlieb, 2005, p. 21). Moreover, when a spoken utterance only consists of these so-called redundant elements (e.g. repetition or false starts), viewers might "feel irritated if they hear something being said and yet no subtitle appears on the screen" (Smith, 1998, p. 146).

Interpersonal elements: According to De Linde and Kay, a common conclusion of the few studies that "have dealt with the underlying principles of editing in interlingual subtitling (de Linde 1995, Gottlieb 1992, Kovačič 1994)" is that elements serving an interpersonal function are most likely to be omitted (1999, p. 48). The term "interpersonal" was first used for subtitling purposes by Kovačič (1992) and was borrowed from Halliday's theory of language (1985, 1994), which distinguishes three types of language functions as shown in Table 2:

Table 2. Halliday's Language Functions

Functions of language		Description	
1. Ideational function	a. Experiential function	Language is used to organize, understand and express our perceptions of the world and of our own consciousness.	The experiential is largely concerned with content or ideas.
	b. Logical function		The logical is concenred with the relationship between ideas.
2. Interpersonal function		Language is used to enable us to participate in communicative acts with other people, to take on roles and to express and understand feelings, attitude and judgements.	
3. Textual function		Language is used to relate what is said (or written) to the rest of the text and to other linguistic events. This involves the use of language to organize the text itself.	

(adapted from Bloor and Bloor, 2004, p. 9)

Although an important tool in accounting for deletions, it seems that the linguistic terms involved (*ideational*, *textual*, *intepersonal*) are not intuitively clear and might not be readily accessible to subtitlers, students and researchers who are not familiar or are little familiar with Halliday's theory. More importantly, no definitions or examples of relevant categories are provided whenever these terms are used in the subtitling literature (for more, see section 3.3.5, p. 68)[12]. The only relevant information comes from Díaz Cintas and Remael, who define intepersonal elements as:

---

[12] Unfortunately, Kovačič's research was mainly reported in her PhD thesis which is not readily available, as it was written in Slovene (1992).

elements that may signal power relations between interlocutors and thereby establish character... [They] contribute little to propositional content in the strict sense, formally ... occupy a somewhat isolated position at the beginning of the sentence, for instance, or between commas. (2007, p. 165)

Examples provided include greetings, interjections, vocatives, formulas of courtesy, etc. (ibid).

It is also important to note that despite its usefulness, Halliday's theory is not sufficient in accounting for deletions, as similar interpersonal features may be treated differently in different contexts. In order to account for this variation, Kovačič complemented Halliday's functional model by drawing on Relevance theory as put forward by Sperber and Wilson (1986) and later adapted for translation purposes by Gutt (2000). According to the main principle of this theory, translators/subtitlers omit elements of the text that are considered less relevant to the reader/viewer for understanding the message and the overall story. Thus, deleting such elements relieves viewers of some processing effort, without making the processing of the rest of the text more difficult (Kovačič , 1994, p. 250). This way, the same element may be treated differently when appearing in different contexts, as it may be highly relevant in one context and dispensable in another. In each case, it is up to the subtitler to evaluate the relevance of each element to the understanding of the message and accordingly to decide whether to retain or omit it. The subtitler's evaluation is based on his/her intuitions or beliefs about what is relevant to the particular audience:

The translator does not have direct access to the cognitive environment of his audience, he does not actually *know* what it is like – all he can have is some assumptions or beliefs about it. (Gutt, 2000, p.118)

Naturally, these assumptions might be inaccurate or plain wrong. In addition, "the decision as to what is necessary and what is superfluous in translation is very individual and will differ from subtitler to subtitler" (Kovačič, 1998, p. 79).

Cohesive devices: Cohesion is one of the seven standards of textuality as defined by de Beaugrande and Dressler (1981). Its role is to make relationships between entities in a text explicit thereby reflecting underlying coherence in the surface structure of the text. It is created by using four main types of cohesive elements (Halliday and Hasan, 1976):

1) Reference - anaphoric and cataphoric, exophoric and endophoric
2) Substitution and ellipsis
3) Conjunction
4) Lexical cohesion - re-iteration (synonymy, repetition, etc.) and collocation

Although cohesive devices are important in creating cohesive harmony and thus holding the text together as a whole, they are often omitted when subtitling (De Linde and Kay, 1999; Neves, 2007). As a result, subtitled texts tend to be somewhat cohesively weaker than their corresponding source texts (Titford, 1982, p. 115). Weakening cohesion may, however, make a text more difficult to process and result in a loss of meaning. For this reason, great care should be taken when omitting cohesive devices (De Linde and Kay, 1999, p. 30).

Reduction and verbatim subtitling in SDH: Although reduction is considered one of the main features of subtitling for hearing viewers, this is not true regarding SDH. Particularly in the case of intralingual SDH, reduction and editing in general are often challenged by the current trend for verbatim subtitling. This trend is evident both in subtitling practice, where there have been reports of subtitle rates of up to 250 words per minute (Neves, 2005, p. 142-143), as well as in users' views as expressed in relevant surveys (Kyle, 1992; IFF Research and Sancho-Aldridge, 1997; RNID, 1999). More importantly, verbatim subtitling has become a banner for Deaf

Associations that consider any kind of editing as a form of censorship and are thus completely opposed to it (Neves, 2005, p. 142).

On the other hand, the same surveys mentioned above have also recorded viewers' comments and complaints about not being able to follow fast, verbatim subtitles. Still other studies have indicated that verbatim subtitling does not on the whole fulfil the needs of D/deaf and HoH people and have suggested the need for editing and/or reduction (Neves, 2005; Araujo, 2004; Franco and Araujo, 2003). As Neves puts it:

> Not having enough time to read subtitles, not having useful time to process information, not understanding the meaning of certain words ... cannot be understood as being given equal opportunities. (2005, p. 150)

It would seem that the problem with these opposing views is that they are rather extreme. It might be more useful and accurate to view verbatim transcription and editing/reduction as complementary rather than opposing strategies in the subtitler's arsenal. For some subtitles, verbatim transcription may indeed be possible and sufficient. For others, it might be and it is often necessary to reduce and/or adapt the original utterance. In all cases, what is important is to determine when and why to choose each strategy. In order to do that, it would be necessary to determine among other things the base-line subtitle rate for the intended audience and more importantly the audience itself. Both issues have not, however, been determined satisfactorily yet (see sections 3.3.1 and 2.5.4 respectively).

2.5 Subtitling for D/deaf and HoH viewers (SDH)

As mentioned in section 2.3.2, we can distinguish two types of subtitling among others: subtitling for hearing viewers and subtitling for D/deaf and HoH people (SDH). In the past, these two types were respectively identified with interlingual and intralingual subtitling, however, as explained in section 2.3.1, this association is now dated. The distinction between subtitling for hearing viewers and SDH is very

important, as it constitutes the first step towards recognising and catering for the different needs of D/deaf and HoH people.

2.5.1 Importance of SDH

The importance of subtitling for D/deaf and HoH people cannot be overstated (De Linde and Kay, 1999, p. 10). In the UK, according to the latest statistics available by the Royal National Institute for Deaf People (RNID), the Royal Association for Deaf people (RAD) and Deafness Research UK[13], approximately nine million people, that is one in seven, are deaf or HoH. Of these, the majority are people who have developed a hearing loss as they get older (age-related hearing loss, also known as presbyacusis) and their numbers are expected to rise, as the number of people over 60 increases. Deaf people, i.e. people with severe or profound deafness (see section 2.5.3), only account for approximately 7% of the total population with hearing loss (688,000 people), while British Sign Language (BSL) is the first or preferred language of around 50,000 to 70,000 people. The situation is somewhat similar in the rest of Europe. In 2005, there were 81.5 million deaf and HoH people across Europe and by 2015 this figure is expected to reach 90 million (Bobeldijk, 2004).

2.5.2 SDH audience from a subtitling perspective

Traditionally, two groups have been identified within the D/deaf and HoH audience from a subtitling perspective: congenitally deaf people and people with acquired hearing loss usually associated with old age.

Congenitally deaf people: According to Ivarsson and Carroll, only one in a thousand people are born deaf, so this first group is rather small (1998, p. 131). Members use sign language as their real mother tongue and primary means of

---

[13] Information available only at
http://www.rnid.org.uk/information_resources/aboutdeafness/statistics/
http://www.royaldeaf.org.uk/index.php?option=com_content&view=article&id=55&Itemid=68
http://www.deafnessresearch.org.uk/factsheets/deafness-the-facts.pdf (all links last accessed on October 28th 2008)

communication and thus, their ability to use and understand the language of the society in which they live (which is their second language) is often limited and accompanied by low reading levels (ibid). Consequently, using subtitles may be extremely difficult for this group and this difficulty is further aggravated by deaf people's limited breadth of knowledge which is due to limited access to information throughout their education and general life (De Linde and Kay, 1999, p. 12). To overcome subtitling problems for this audience, it is "important to clarify the vocabulary and simplify the syntax used … to the greatest possible extent and to keep the subtitle speed low" (Ivarsson and Carroll, 1998, p. 132).[14] Moreover, all acoustic information crucial to the plot needs to be represented in the subtitles and special care has to be given in order to enable viewers to identify who is speaking and what is happening off-screen.

People with acquired hearing impairments: This group, which is much larger than the first one, consists of people who have acquired deafness or other hearing impairments later in life. The vast majority of this group become deaf or HoH in their 50s and 60s (De Linde and Kay, 1999, p. 11) and given the UK and Europe's increasing ageing population, their numbers are expected to rise even more. Unlike congenitally deaf people, people with acquired hearing loss share the mother tongue of their social environment and are more likely to have had an education within the hearing community and to exhibit average reading skills. In addition, there is no evidence to suggest that their reading ability is adversely affected by their hearing loss (De Linde and Kay, 1999, p. 11). Such viewers are thus more likely to find it easier to read subtitles at normal presentation rates and to understand complex syntax and vocabulary (Ivarsson and Carroll, 1998, p. 132).

However, there appear to be a few problems with this distinction. For one thing, it perpetuates the erroneous belief that the D/deaf and HoH are one and the same group (see Neves below). In addition, it unduly restricts the D/deaf and HoH audience; apart from the congenitally deaf and people who have acquired hearing loss much later in life, there are also congenitally HoH people, as well as children who have

---

[14] Such a suggestion would more than likely be seen as patronising and condescending by many D/deaf and HoH viewers and associations.

acquired hearing impairments soon after birth or at a very young age. Moreover, even if we take this distinction for granted, there seems to be another problem: as mentioned above, people with acquired hearing loss are less likely to encounter problems with reading and subtitling. However, as was also mentioned, the vast majority of this group become deaf or HoH in their 50s and 60s and older people seem to be more likely to have problems with reading subtitles, particularly fast ones.

Some of these problems have been addressed by Neves, who has called for greater emphasis to be paid to the elemental distinction between Deaf and HoH people arguing that their needs are not similar in nature and thus as currently practiced, SDH usually fails to cater for the needs of either group (2005, p. 22). More recently, she has also suggested that it might actually be more useful to think in terms of four (sub)groups each characterised by different needs:

- the profoundly deaf who use sign language as their first language and read written text as their second language
- the oral deaf who relate to a spoken language as their first language and see themselves as part of the hearing community even though they themselves have very little or no residual hearing
- the late deafened who have acquired deafness to greater or lesser degree at some stage but who have a memory of hearing, and
- the hard of hearing, who have some degree of hearing capacity

(Neves, 2007b, p. 262)

Although this distinction is more accurate than the traditional distinction between congenitally deaf people and people with acquired hearing loss, they are both bound to make generalisations. Generalisations cannot be avoided when it comes to such a massive audience; however, it is important to stress that as is true for hearing people, D/deaf and HoH people are a highly heterogeneous group consisting of individuals whose reading ability and thus ability to use subtitles is determined by the complex interplay of a variety of factors. Such factors and other aspects of hearing loss that

influence the acquisition of oral language literacy skills will be discussed in the following section.

2.5.3 Some basics about hearing impairment

Definition: "Hearing impairment is a generic term referring to all types, causes and degrees of hearing loss", thus encompassing both deaf and HoH people (Paul and Quigley, 1990, p. 40). Occasionally, the term is used in a more restrictive way and a distinction is made between *hearing loss* and *deafness* (Moores, 1987, p. 9). However, it is the general and more common sense that is used in this thesis.

It is important to note here that there is not only a clinical perspective on hearing impairment, but also a cultural one. The former approach defines hearing impairment in terms of audiological parameters and generally views hearing impairment in a negative light as a pathological condition. The cultural view, on the other hand, does not treat hearing impairment as a pathological condition, but focuses instead on the language, experience, customs and values of the Deaf community/culture. This distinction is partly reflected in the terms *deaf* and *Deaf*, which following convention are used respectively to denote people with severe to profound hearing loss and people who belong to the Deaf community.

Degree of hearing loss: An individual's degree of hearing loss can be measured using an audiometer, that is by presenting single sounds via headphones and noting the level at which the person can detect the sounds (Watson et al., 1999, p.3). Typically, hearing is considered normal up to a hearing threshold of about 26dB in the better ear (see Table 3). Over that limit, people with hearing loss of up to 70dB are considered HoH, whereas people with hearing loss over 91dB are considered deaf. People with hearing loss between 71 and 90dB can be considered either deaf or HoH depending on their use of residual hearing; it has also been observed that within this range "some individuals are socially deaf ... but more of them are merely hard-of-hearing" (Davis, 1978, p. 88 quoted in Paul and Quigley, 1990, p. 41).

Table 3. Degree of Hearing Loss

Degree of Impairment	Description	Label
Up to 26 decibels	Normal	Normal hearing
27 to 40 decibels	Slight	Hard-of-hearing
41 to 55 decibels	Mild	Hard-of-hearing
56 to 70 decibels	Moderate	Hard-of-hearing
71 to 90 decibels	Severe	Hard-of-hearing, or Deaf
91+ decibels	Profound (Extreme)	Deaf

(Paul and Quigley, 1990, p. 41)

Age of onset: Two types of hearing loss can be distinguished along this parameter, prelingual and postlingual hearing loss. Prelingual hearing loss refers to the condition of people whose hearing impairment is present at birth or occurs within the first two years of life prior to the development of speech and language. Postlingual hearing loss, on the other hand, refers to the condition of people whose hearing impairment occurs following the acquisition of speech and language (Moores, 1987, p. 9). As Paul and Quigley (1990, p. 41-42) point out:

> When degree of impairment is considered in conjunction with age of onset, one sees significant effects on the development of primary language and secondary language skills... A child who becomes deaf after age 2 has language, communication, and educational needs that are very different from those of a child who becomes deaf before the age of 2.

Likewise, "the later the onset of hearing impairment, the greater linguistic competence will have been gained and therefore the easier the interaction with the hearing community" (Neves, 2005, p. 78). In addition, there appears to be an age

after which the reading ability of people who lose their hearing later in life, particularly in the case of age-related hearing loss, is not adversely affected (De Linde and Kay, 1999, p. 12).

First language: This variable refers to the language with which people have been brought up and is relevant in the case of prelingually deaf and HoH people. Thus, depending on their parents' decisions, children may have been brought up using:

i) the oral language, e.g. English,

ii) sign language (SL), e.g. BSL,

iii) any other form of oral-based manual code, e.g. fingerspelling, the Rochester Method, signed English (Paul and Quigley, 1994, p. 24-33),

iv) both the oral language and SL or another form of oral-based manual language.

Although it is still debatable whether oral language, oral-based manual communication or SL is better at developing oral language skills, acquiring a first language (whatever that might be) at as early an age as possible seems to have critical implications for later success (Paul and Quigley, 1994, p. 144). Overall, it appears that deaf children who are better readers have had early access to language (usually SL) and have also been exposed to English (Marschark, 1997, p. 140).

Education: Education is inextricably linked to the issue of language. In particular, the issue concerning which mode or form of communication should be used by and with hearing-impaired students is still highly controversial and continues to attract attention even 200 years after it was originally raised (Paul and Quigley, 1994, p. 19). Three major modes or communication philosophies may be distinguished: (a) oralism, which places major emphasis on the use of the language in its oral form, i.e. speech, (b) SL and (c) Total Communication, which advocates the right to use or choose from all forms of communication, oral and manual/signed alike.

Another important issue within the context of education is that of integration, that is the attendance of regular teaching classes as opposed to special schools for the

D/deaf and HoH. For instance, in the case of HoH students, there is evidence suggesting that other factors being equal, including the degree of hearing impairment, the more fully mainstreamed the average HoH child is, the greater his/her academic achievements are (Ross et al., 1976, p. 48).

Degree of hearing loss, age at onset, first language and education are all important aspects of hearing loss and among the most important factors influencing the acquisition of oral language skills. Other factors considered important in that respect include early identification and intervention, adequate exposure and acquisition of language and communication skills early in life, the encouragement of parents and teachers, parent and child motivation, and the quality of teaching. The complex interplay of all these factors accounts for the heterogeneity of D/deaf and HoH people in terms of oral language skills and makes catering for their needs even more demanding.

2.5.4 SDH for whom?

Even on the basis of this limited information on hearing impairment, there is no doubt that not only are the D/deaf and HoH two different groups, but what is more important within each group members differ not only in outlook, but also in language and communication needs to such an extent that we might in fact be dealing with three (deaf, Deaf, HoH) or even four distinct groups (as suggested by Neves, 2007b). In one of the earliest large-scale surveys to be conducted in the UK, Kyle reports:

> the main finding of the study which one returns to again and again is that there are several populations embedded in the classification "deaf" and they are typically untypical. (1992, p. 153)

On a similar note, the RNID 1999 research points out:

> There is a need to acknowledge that deaf and hard of hearing people have the same range of intelligence and interests as everyone else; that, like

everyone else, they read and assimilate information at different speeds...
There is therefore a diversity among deaf and hard of hearing people
which raises the question to what extent developing technologies can
promote/assist choice for this audience. (p. 34)

This heterogeneity raises an important question as to the intended audience of SDH:
is it the Deaf, the deaf or the HoH? Is it prelingually or postlingually hearing-
impaired people? And within each group, what is the profile of the average viewer
constructed by broadcasters, subtitling companies and subtitlers? And how is this
profile related to specific subtitle characteristics? Even if we adhere to the traditional
subtitling distinction between congenitally deaf people and people with acquired
hearing impairment, it is not clear whether the intended audience is the congenitally
deaf, people with acquired hearing loss or both. Subtitles today seem to be produced
with an indeterminate audience in mind. More importantly, who *should* the intended
SDH audience be? And how can this be determined? For instance, on a purely
numerical basis, SDH should be addressed to HoH people who have lost their
hearing later in life; on the other hand, viewing SDH as a social service would more
than likely place emphasis on the D/deaf as the intended audience. Could there
possibly be a golden mean that would help us accommodate as wide a range of
viewers as possible?

It appears that these questions have not been answered in full yet, as research has still
to provide a detailed description of the characteristics of current subtitle output, as
well as an adequate account of how they are related to D/deaf and HoH viewers'
response and how this response varies among the D/deaf and HoH. Surprisingly, this
is true even in countries such as the UK, which is one of the few countries worldwide
with a long tradition and a high volume of intralingual SDH and where many
relevant studies and surveys have been undertaken (see for instance Kyle, 1992; De
Linde, 1997; IFF Research Ltd and Jane Sancho-Aldridge, 1997; RNID, 1999; NSM
Research, 2001; Ofcom 2005).[15] Although relevant research has a lot of useful

---

[15] This is particularly true for television broadcasting, and without a doubt, the BBC is at the forefront
in this area subtitling 100% of its programmes since May 2008 (see

information to offer, most of it has so far focused on issues, such as determining priorities for subtitling, reasons for using or not using subtitles and describing attitudes towards SDH and the use of subtitling. There has been very little research focusing on quality issues (such as subtitle speed, verbatim vs. editing subtitling, preference for block vs. scrolling subtitles, use of colour and add-on lines) and even less research has sought to provide an answer to the issues posed in the previous paragraph: what is the profile of current subtitle output and how are particular subtitle characteristics related to D/deaf and HoH viewers' response? The only exceptions seem to be Kyle (1992) and to a much greater extent the research by De Linde which has been reported in both De Linde (1997) and De Linde and Kay (1999).

The importance of De Linde's research lies among other things in that it set out to measure subtitle characteristics and relate them to viewers' response in order "to identify ways in which to improve the effectiveness of television subtitles for deaf and hard-of-hearing people" (1997, p. 3). However, despite its undoubted significance, it only constitutes the first step towards answering these questions due to its restricted nature; indeed, although many different genres were covered, only two programmes per genre and very few subtitles per programme were studied (approximately 25 subtitles per programme) and only a limited number of subjects were involved (twenty in all, of whom ten were deaf and ten were hearing) including no HoH people or people over 53 years old. Thus, the findings reported by De Linde (1997) and De Linde and Kay (1999) can only be seen to be indicative and need to be corroborated and extended by further research. Such research would need to cover a large number of genres, programmes, subtitles and characteristics, as well as a large number of D/deaf and HoH people covering as wide a range of the audience as possible. It seems that such large-scale research can only be carried out effectively by broadcasters in close association with D/deaf and HoH people and associations, subtitling associations such as ESIST and possibly universities.

---

http://www.bbc.co.uk/pressoffice/pressreleases/stories/2008/05_may/07/subtitling.shtml , last accessed on October 28th 2008).

In addition, as mentioned in Chapter 1, it is essential not only to match subtitle characteristics with audience needs and preferences, but also to compare both with established guidelines and amend as necessary. Existing guidelines and empirical subtitling research relevant to this thesis will be presented in detail in Chapter 3.

3. Existing guidelines and empirical research

This chapter focuses on existing guidelines and empirical research relevant to the subtitle characteristics covered in this thesis. After a short introduction, a brief description of each characteristic is provided along with a presentation of relevant guidelines both from the area of SDH and interlingual subtitling for hearing viewers. This is then followed by an overview of the findings of any relevant empirical research available.

3.1 Guidelines

Guidelines or codes of good practice are now "becoming common currency in the [subtitling] profession" (Díaz Cintas and Remael, 2007, p. 79). Most of these documents are in-house manuals for internal use "written out by more experienced professionals, who compile their knowledge so as to help [less experienced,] less qualified or less proficient professionals to do a better job" (Neves, 2005, p. 43). Apart from professional guidelines, there are also guidelines that have been compiled by organisations such as the Office of Communications (Ofcom) in the UK and guidelines that have been compiled by researchers and/or professionals and have appeared in academic journals and books. In all cases, however, the goal is the same: to promote best practices and to ensure a minimum level of subtitling quality. It is also important to note that even though they might be used in a prescriptive manner, guidelines are not to be considered as absolute rules, but rather as recommendations and a listing of suggestions for further questioning and reflection (ibid, pp. 73-74).

Given the focus of this thesis on intralingual SDH, the majority of guidelines that have been consulted belong to this field of subtitling. Many of them come from the UK; among these, the BBC guidelines are the most important, as not only are they the only in-house guidelines available for consideration in this thesis, but they are also the most relevant given that the programmes in the research corpus all come from the BBC. SDH guidelines from other countries have also been considered for

comparison purposes. For the same reason, it was also decided to include guidelines relevant to interlingual subtitling for hearing viewers. Guidelines relevant to the characteristics covered in this thesis will be presented in detail in the respective sections in the remainder of the chapter.

3.2 Descriptive research

Despite the wealth of subtitling research that has become available in the last decade or so, relatively little research has been dedicated to measuring and describing current subtitle output. For instance, regarding reduction, although verbatim vs. edited subtitling is a much discussed issue especially in SDH, we still do not know to what extent current subtitles are verbatim and to what extent they are edited. Limited research is available and this has tended to concentrate on interlingual subtitling for hearing viewers. Even then, focus has been on films and related genres, such as sitcoms and soap operas, and the programmes analysed have been limited in number ranging from just one to twelve programmes (Gottlieb, 1992; Lomheim 1995; Koljonen, 1995; Lomheim, 1999; Georgakopoulou, 2003; Pedersen, 2003).[16] In the absence of any large-scale research, it is thus surprising, to say the least, to read generalisations such as the one by Gottlieb, who asserts that "in television subtitling the text volume is typically reduced by one third" (1998, p. 247).

The situation is somewhat similar in the context of SDH regarding both the issue of reduction and to an even greater extent a host of other issues. Very little research is available and in fact, the only empirical studies of this type seem to be by De Linde (1997) and De Linde and Kay (1999) in the UK and Jensema and McCann (1995) and Jensema et al. (1996) in the USA. Since the focus of this thesis is the status of intralingual SDH in the UK, the results reported by Jensema and his associates will be provided for comparison purposes only. As pointed out in section 2.5.4, despite its undoubted merits and importance, De Linde's research is rather restricted as

---

[16] Georgakopoulou and Pedersen's (2003) research projects, which involve the analysis of 12 programmes, seem to be the exception in that respect.

although it covers six different genres, it only covers two programmes per genre and only approximately 25 subtitles per programme. Without a doubt, much more empirical work is required.

3.3 Specific guidelines and empirical research

This section presents guidelines and any empirical findings relevant to the subtitle characteristics covered in this thesis. The following characteristics are discussed: subtitle duration and delay time between two consecutive subtitles (both aspects of the temporal dimension of subtitling), maximum number of lines, line segmentation, reduction and overall strategies used.

3.3.1 Duration of subtitles

As mentioned in section 2.4.3, the amount of time available for displaying a subtitle depends on the pace of the soundtrack dialogue, the amount of text in the subtitle, the average reading speed of viewers and the need to keep a minimum interval between subtitles (Luyken et al., 1991, p. 42-43). Reading speed in turn is affected by a host of factors, such as the rate of presentation, the quantity (in terms of characters) or complexity of the information presented in the subtitle, the type and pace of visual information on screen, the genre of the programme, the distribution channel, the level of literacy of the audience as well as their degree of familiarity with the source language and subtitling (see Luyken et al., 1991, p. 44; Ivarsson and Carroll, 1998, p. 65; De Linde and Kay, 1999, p. 6; Díaz Cintas and Remael, 2007, p. 96).

Before presenting the guidelines available, it is important to note that there appears to be a dearth of relevant empirical research involving D/deaf and HoH people. Many guidelines are still more or less based on the 6 second rule, whereby three seconds are available per full subtitle line. Further research is now required to determine whether that rule is still valid or whether it needs to be updated. Such research would need to involve a large number of D/deaf and HoH people, a large

number of programmes and subtitles per programme, as well as a variety of different subtitle speeds per programme.

Research in that direction has already been conducted but not to a sufficient extent. In the UK, starting from the premise that "in practice, the average subtitle speed for pre-recorded programmes is commonly around 160 words per minute", Ofcom has already conducted a study "to evaluate how acceptable increased subtitling speeds are for different types of deaf and hard-of-hearing people, and what factors are important" (2005, p. 6, 8). Findings confirm among other things that the issue of speed is extremely complex influenced by the degree of deafness, age, reliance on subtitles and literacy level of the viewers, and the main recommendation issued is that subtitling speed should not normally exceed 180 words per minute (wpm) (ibid, p. 5). However, the research is of limited value, as only 64 participants were involved and no viewers with mild or slight hearing loss were included. More importantly, there is no information about the number of subtitles shown per programme and the actual subtitle rates experimented with, and no figures are provided. A larger-scale research in a similar direction is that of Jensema in the USA (1998). On the basis of a sample of 578 participants, he reports that the rate at which subtitle speed was most comfortable for viewers was about 141 wpm. Nevertheless, despite the large sample involved and the useful data provided, findings and suggestions can only be considered as a basis for further research: materials used were not authentic (consisting of videos taken of posters), each video lasted for only half a minute, the mean age of participants was a mere 31.7 years (too young considering many people lose their hearing later in life) and a high percentage of participants were hearing rather than deaf or HoH (262 out of a total of 578).

SDH guidelines

The first SDH guidelines available in the UK recommended a maximum subtitle presentation rate of 120 wpm, that is approximately three seconds per full line (Baker, 1981, p. 10; Baker et al., 1982, p. 13). More recent SDH guidelines, on the

other hand, assume a faster reading speed and have raised that maximum to 140 wpm, which corresponds to approximately two seconds and 15 frames per line (ITC, 1997, p. 9; ITC, 1999, p. 11). In addition, Ofcom, the company that replaced the ITC (Independent Television Commission), seem keen on revising this maximum and setting it even higher (Ofcom, 2005); however, no new guidelines have become available as yet. The BBC Subtitling Guide. on the other hand, provides the following standard timings:

- A short and familiar word or phrase – 1.12 to 2 seconds
- Up to half a line – 2 to 2.12 seconds
- One line – 2.12 to 3 seconds
- One line and a little bit – 3.12 seconds
- Up to one and a half lines – 4 to 4.12 seconds
- Two lines – 5 to 6 seconds
- Two lines and a little bit – 6.12 seconds
- Two and a half lines – 7 seconds
- Three lines – 7.12 to 8 seconds[17]

(BBC, 1998, p. 41)

As pointed out in the Guide, these timings are only intended as general guidelines and should thus be applied unless the context demands otherwise (ibid, p. 39). That is, in some cases, extra time should be provided in order to allow viewers to read the subtitle, e.g. when the subtitle includes an unfamiliar word/phrase or a long figure, when there are several speakers or when there is a lot happening on screen. Similarly, extra time should be allowed for synchronisation purposes when the original speech is slow. On the other hand, there are also cases, where it is often necessary to provide less time than specified by the recommended standard timings. Such cases include the need to respect shot changes, as well as the need to avoid

---

[17] Here, we need to point out that the use of seconds per line rather than wpm used in other guidelines renders the BBC standard timings rather vague. In all cases, it would seem that both information about seconds per line and wpm would need to be complemented by information about characters per second, as "The correspondence between x number of subtitle characters and y words per minute is of course extremely approximate and varies from one language to another, depending on the language's average word length and speaking speed" (Ivarsson and Carroll, 1998, p. 67).

editing words that can be lip-read, humorous words and critical information in news items or factual programmes.

Regarding SDH guidelines from other countries, in Ireland, on the basis of the guidelines by ITC, the Broadcasting Commission of Ireland (BCI) has published its own set of guidelines, where it is recommended that the presentation rate of subtitles should not exceed 160 wpm, i.e. two seconds per full line (2005, no page). In Australia, rates are even higher with recommendations of 180 wpm (Deafness Forum of Australia, 2004, no page). In the USA, the Described and Captioned Media Program (DCMP), with its emphasis on promoting and providing equal access through described and subtitled educational media, recommend different rates for different types of content; thus, lower- to middle-level educational media should not exceed presentation rates of 120 to 130 wpm, upper-level educational media may be captioned at slightly higher speeds, adult special-interest media should be subtitled at a rate of 150-160 wpm and even higher if required, and adult movies [sic] should be captioned at a near verbatim rate, but not faster than 235 wpm (2007, p. 12). Unfortunately, however elaborate, this classification is rather obscure, as it is not clear what is meant by each programme category.

The most detailed SDH guidelines regarding subtitle display times have been suggested in Portugal by Neves (2004). She recommends a baseline presentation rate of 120-130 wpm and suggests specific presentation rates depending on the number of lines, syntax, vocabulary complexity, etc. (ibid, p. 29-30). She also recommends that no subtitle should be left on screen for less than one second and ten frames and that no subtitle should be kept for more than six seconds (ibid, p. 29).

Overall, recommendations on subtitle display times seem to be more tentative than guidelines regarding other characteristics. For instance, according to the ITC guidelines, "presentation rates will depend upon the programme content" (1999, p. 11), while according to Neves, subtitle presentation rates have to be determined on the basis of the rhythm, style and language of each film/programme (2004, p. 29).

Interlingual guidelines

Guidelines for interlingual subtitling for hearing viewers are more in line with the SDH recommendations by Baker (1981), Baker et al. (1982) and Neves (2004), as illustrated below. It is also interesting to note that unlike most SDH guidelines, guidelines for interlingual subtitling are more often expressed in seconds per line rather than wpm.

Duration of a full two-line subtitle: Most of the guidelines available agree that a full two-line subtitle should be available on screen for a maximum time of five to six seconds (Ivarsson and Carroll, 1998, p. 63; Karamitroglou, 1998, no page; Díaz Cintas, 2001, p. 113; Díaz Cintas and Remael, 2007, p. 89). No subtitle should be kept for more than six seconds, as "this would cause automatic re-reading of the subtitle, especially by fast readers" (Karamitroglou, 1998, no page). The only exception is the subtitling of songs, "in which case the subtitle should be left hanging on screen beyond the six seconds if the rhythm requires it" (Díaz Cintas and Remael, 2007, p. 97; see also Ivarsson and Carroll, 1998, p. 65).

Duration of a full single-line subtitle: Again, the majority of guidelines recommend that full single-line subtitles are displayed for about three to three and a half seconds (Ivarsson and Carroll, 1998, p. 63; Karamitroglou, 1998, no page, Díaz Cintas and Remael, 2007, p. 97)

Duration of a single-word subtitle: According to older guidelines, the minimum duration even for the briefest subtitle is at least one second and a half (Ivarsson and Carroll, 1998, p. 63; Karamitroglou, 1998, no page; Díaz Cintas, 2001, p. 113). This seems now to have been replaced by a recommendation to retain the subtitle for one second (Díaz Cintas and Remael, 2007, p. 90).

Duration of a subtitle consisting of one and a half lines: Subtitles consisting of one and a half lines are not commonly mentioned in existing guidelines, as it is seems that their duration is to be deduced on the basis of the guidelines available for single-

line and two-line subtitles. When mentioned, it is recommended that they remain on screen for approximately four seconds (Ivarsson and Carroll, 1998, p. 63).

Duration of a three-line subtitle: Given the relevant recommendation to limit the number of lines in any subtitle to two (see section 3.3.3), there are no guidelines on three-line subtitles and their duration from the perspective of interlingual subtitling for hearing viewers.

Empirical research

As mentioned in section 3.2, there has been very little empirical research measuring the characteristics of subtitle output. Regarding the duration of current subtitles and subtitle presentation rates, there are two relevant studies within the context of SDH and one from the area of interlingual subtitling for hearing viewers.

The first SDH study and the only one available from the UK is the one by De Linde (1997), followed by De Linde and Kay (1999). On the basis of their research and using a sample of 262 subtitles from twelve programmes belonging to six different genres, they report the subtitle presentation rates shown in Table 4:

Table 4. Presentation rates by De Linde and Kay (1999, p. 45-46)

Programme type	Presentation Rate (wpm)
Comedy	125
Documentary	126
Drama Serial	132
Film	67
Game show/Chat show	133
Magazine	124

As can be seen, more interactive programmes, such as game shows, chat shows and drama serials were found to exhibit higher subtitle rates than narrative-based programmes, such as documentaries. The researchers acknowledge that the figure for

films is unusually low; this was due to the nature of the two samples used, as they both "contained a large proportion of non-speech sound such as animal noises, background music, machine bleeps and so on" (ibid, p. 45). This limitation provides further evidence in support of the need for large-scale research analysing a larger sample consisting of more programmes and subtitles per programme.

The second research within the context of SDH comes from the USA and is an example of the type of research needed in the UK. Jensema and his associates undertook an extensive project measuring among other things the average presentation rates of TV subtitles (1995, 1996). The research involved 183 programmes from 13 different genres, namely children's animation programmes, children's educational programmes, children's action programmes, prime-time drama, situation comedies, films, news, documentaries, talk shows, soap operas, music specials, sports and live performances. Detailed statistics are provided for all programmes and for each genre separately, including wpm, characters per minute, characters per word, lines per minute, words per line and characters per line (1996, p. 287). Table 5 presents wpm results for all 183 programmes and separately for documentaries, which are the focus of the present thesis.

Table 5. Presentation rates, adapted from Jensema et al. (1996, p. 287)

Words/minute	All programmes (N = 183)	Documentary (N = 17)
Mean	141	139
Standard Deviation	21	12
Maximum	231	161
Minimum	74	113
Range	157	48

Documentaries along with films, prime-time dramas and sitcoms were found "to cluster around the mean captioning speed of 141 WPM" (ibid, p. 288). Soap operas, news programmes and talk shows were the programmes exhibiting the highest presentation rates, while sports programmes and music specials were the programmes with the slowest subtitle speeds.

The last study comes from the field of interlingual subtitling for hearing viewers. Analysing an episode from three TV series subtitled into Norwegian, *Maigret* (a French detective series, with emphasis on the depiction of milieu and characters, and with a literary and contemplative tone), *Golden Years* (an American science fiction series consisting of plenty of action, dramatic developments and short, sharp exchanges), and *Allo! Allo!* (a British comedy series set during the Second World War and known for its burlesque humour largely based on a constant stream of verbal jokes and situation comedy), Lomheim (1999) measured the display time of each subtitle and reported the results shown in Table 6:

Table 6. Number of subtitles according to exposure times (highest percentage for each programme shaded) by Lomheim (1999, p. 196)

	Total subtitles	<2 secs	2-3 secs	3-4 secs	4-5 secs	5-6 secs	>6 secs
Maigret	586	0 (0.0%)	27 (4.6%)	126 (21.5%)	209 (35.7%)	130 (22.2%)	94 (16.0%)
Golden Years	338	3 (0.9%)	66 (19.5%)	99 (29.3%)	69 (20.4%)	57 (16.9%)	44 (13.0%)
Allo! Allo!	286	25 (8.7%)	77 (26.9%)	74 (25.9%)	64 (22.4%)	34 (11.9%)	12 (4.2%)

On the basis of these statistics, Lomheim concludes that the three programmes display clear differences and he emphasises the importance of the pace of the dialogue apart from the number of characters in determining exposure time. Although it is true that the length of a subtitle is not a sufficient factor in deciding its duration, it would have been interesting to know for how many of the subtitles analysed by Lomheim exposure time could have actually been determined on the basis of subtitle length alone.

3.3.2 Time between consecutive subtitles

A time break needs to exist between all subtitles, otherwise the viewer might not perceive the change of subtitle text. This break "signal[s] to the brain the disappearance of one subtitle ... and the appearance of another" and is essential,

particularly when the two subtitles are similar in terms of length and layout (Karamitroglou, 1998, no page).

SDH guidelines

Although much information is provided about timing, there is only one recommendation relevant to the minimum number of frames between consecutive subtitles. Thus, according to the BBC Subtitling Guide, if subtitles follow each other immediately, the subtitler should advance the in-cue by one frame rather than make the in-cue the same as the preceding out-cue, as this makes the subtitles transmit very abruptly (1998, p. 40).[18] The remainder of SDH guidelines offer no relevant recommendations. Emphasis is instead placed on other issues of the temporal dimension, such as presentation rates (see section 3.3.1), synchronisation, i.e. when a subtitle should appear and disappear in relation to the corresponding utterance, and the need to respect shot changes/cuts that signify a thematic change in the programme (see Baker et al., 1982; ITC, 1997; ITC, 1999).

Interlingual guidelines

Older guidelines recommend a minimum of four to five frames between subtitles, even when the dialogue is ongoing (Ivarsson and Carroll, 1998, p. 64-65; Karamitroglou, 1998, no page; Díaz Cintas, 2001, p. 113). This minimum seems to have now been lowered to two to three frames (Díaz Cintas and Remael, 2007, p. 92).

Empirical research

The only empirical research measuring subtitle intervals is by Lomheim (1999). Measuring the intervals between subtitles for each of the three programmes analysed, he provides the statistics shown in Table 7:

---

[18] We need to point, however, that the particular recommendation is rather vague, as it is not clear whether the term"subtitles that follow each other immediately" refers to unfinished utterances of one speaker carried over multiple subtitles, fast dialogue, interrupted discourse or all of the above.

Table 7. Time intervals between subtitles, adapted from Lomheim (1999, p. 199)

	Maigret	*Golden Years*	*Allo! Allo!*
<2 frames	0	0	0
2 frames	324	0	0
3 frames	6	24	48
4-12 frames	33	88	96
13-24 frames	36	44	30
1-2 secs	53	57	37
2-4 secs	40	48	39
> 4 secs	93	76	35

Lomheim uses subtitle intervals as an indication of the rhythm or pace of subtitles; longer pauses between subtitles equal a less frantic rhythm. On the basis of his results, he notes that

> a rapid pace in the subtitles (short intervals) does not necessarily go hand in hand with a fast pace and dramatic action in the programme. On the contrary, the material in this study reveals that the programme with the slowest pace of action has the fastest pace of subtitling (1999, p. 199).

3.3.3 Maximum number of lines

As mentioned in section 2.4.3, the space available for presenting subtitles is severely limited due to the physical limitations of the size of the screen, the necessity of not obscuring the image and the need to display subtitles long enough so as to allow viewers adequate reading time (Luyken et al., 1992, p. 42). This limitation has direct influence on the maximum number of lines to be displayed at any time, as well as on the preference for one-line, two-line, three-line or even four-line subtitles.

SDH guidelines

The initial guidelines by Baker recommended a maximum subtitle length of two lines (1981, p. 3). This recommendation was later qualified to allow room for three-line

subtitles as long as no important picture information is obscured (Baker et al., 1982, p. 9) and has since been endorsed by both the ITC (1999, p. 7) and the BCI (2005, no page). The BBC Subtitling Guide also allows for a maximum of three lines (1998, p. 34); when possible, however, it is preferable to fit subtitle text on one or two lines rather than two or three lines respectively unless the result is a very bad line break (ibid, p. 28). Similarly, the Deafness Forum of Australia also allow room for a maximum of three lines per subtitle, although they too state their preference for one- and two-line subtitles (2004, no page). A similar guideline is provided by the DCMP, who recommend using only a maximum of two lines per subtitle, but also allow three and even four lines of text if "a one- or two-line caption would interfere with pre-existing graphics or be confusing in speaker identification" (2007, p. 7). Neves too recommends a maximum number of two lines per subtitle occasionally allowing for three lines to provide information, such as speaker identification and information about sound and sound effects (2004, p. 16). Overall, it would appear that the preferred maximum length is two lines per subtitle and that longer subtitles are only to be used under special circumstances.

Interlingual guidelines

Interlingual guidelines differ from SDH guidelines in that they all agree that the maximum number of lines per subtitle should be limited to two (Ivarsson and Carroll, 1998b, no page; Karamitroglou, 1998, no page; Díaz Cintas, 2001, p. 112; Díaz Cintas and Remael, 2007, p. 82). The only exceptions are SDH and bilingual subtitling where three and even four lines can be used. Additional recommendations are available regarding when to use a single line and when to opt for a two-liner with most guidelines suggesting that as long as subtitle text can be presented in a single line, a one-liner is usually preferable (Ivarsson and Carroll, 1998, p. 76; Díaz Cintas and Remael, 2007, p. 86). Still more recommendations refer to the ideal length of each line in the case of a two-line subtitle and to the issue of line breaks, which is covered in detail in section 3.3.4.

Empirical research

There appears to be no empirical research describing subtitle output in terms of the number of lines used per subtitle. The only exception within the area of interlingual subtitling for hearing viewers is the study by Lomheim (1999), who provides statistics showing the distribution of one- and two-line subtitles in the three programmes of his corpus. According to the data presented, in two of the programmes the percentage of one-line subtitles was approximately 35%, while in the third programme it was only 17%. This difference is interpreted in terms of the third programme's literary tone, which accounts both for the high number of words in the original dialogue and the low reduction rate in the subtitles (ibid, p. 192).

3.3.4 Line segmentation

As mentioned in section 3.3.3, many interlingual guidelines state that when possible it is preferable to write the text of any subtitle on a single line. However, this is not always possible or even desirable and quite often subtitles have to continue over a second or third line or even over a new subtitle. In all cases, what is important is to present the text in such a way as to enhance readability. Reading text is a highly complex cognitive activity, which becomes even more demanding when reading subtitles (see section 2.4.6, p. 34). For this reason, it is essential to carefully segment subtitle text as "only in this way can the cognitive process of reading the subtitles and watching the action proceed with the least effort" (Perego, 2008, p. 35). Careful segmentation entails following syntactic and grammatical considerations (Díaz Cintas and Remael, 2007, p. 172) so that words intimately connected by grammar, semantics or logic are written on the same line (Ivarsson and Carroll, 1998, p. 90) and the content of each line and/or subtitle reads naturally and as fully as possible (Neves, 2005, p. 214). If such considerations are not taken into account, the result is poor segmentation which is assumed to make text-processing particularly demanding with disruptions spreading beyond poorly broken subtitles into subsequent well-structured ones (Perego, 2008, p. 42).

Before presenting the relevant guidelines, we need to point out that segmentation applies both within and across subtitles with rules thought to be basically the same in both cases (Díaz Cintas and Remael, 2007, p. 172). In this thesis, focus has been exclusively on segmentation within subtitles, also known as line breaks. It is also important to note that segmentation is not only a linguistic matter, but also a matter of geometry; geometry is here understood as the need to have both lines of a subtitle as equal in length as possible (Karamitroglou, 1998, no page). Many of the guidelines available suggest that segmentation should be a compromise between linguistics and geometry and that linguistic considerations should only take precedence if the result is a bad line-break (see for instance Baker et al., 1982, p. 12; Karamitroglou, 1998, no page; Neves, 2004, p. 41). This thesis has focused on line breaks considered inadequate in linguistic terms and guidelines mentioned below are presented from such a linguistic perspective.

SDH guidelines

SDH recommendations on line breaks in the UK are rather general. Baker et al. point out that "linguistically coherent segmentation of text can significantly improve readability" (1982, p. 11) and suggest that lines should end at natural linguistic breaks, ideally at clause or phrase boundaries (ibid, p. 11, also Baker, 1981, p. 5). If that is not possible, they recommend breaking the line after other major sentence constituents, such as the main verb. They also advise against lines that break up closely integrated phrases (e.g. splitting an article from a noun) and in particular against line-breaks within a word, as they are especially disruptive to the reading process (ibid, p. 12). The ITC guidelines are effectively the same, although there is a further recommendation to balance linguistic considerations with eye movement when using left, right and centre justification to identify speaker position (1999, p. 9, 1997, p. 7-8). Similarly, the BBC guidelines also state that line breaks should occur at logical points in a sentence, e.g. at the end of a phrase or clause (1998, p. 38). If that is not possible, then it is recommended to avoid splitting the following parts of speech:

- article and noun

- preposition and rest of prepositional phrase

- conjunction and following phrase/clause

- pronoun and verb

- parts of a complex verb

(ibid, p. 28)

At the same time, the BBC Subtitling Guide slightly differs from other SDH and interlingual guidelines in that although it underscores the importance of good line breaks for reading and comprehension, it also specifies that segmentation should give way to good editing and timing when they are mutually exclusive (ibid). It is not, however, clear whether good editing also includes no editing and the preference for verbatim subtitling.

In Ireland, the BCI has adopted the recommendations of the ITC guidelines (2005, no page). In Australia, the Deafness Forum only suggest that "line breaks and caption breaks should reflect the natural flow of the sentence and its punctuation" (2004, no page). More detailed recommendations appear in the guidelines published by the DCMP in the USA (2007). They are as follows:

- Do not break a modifier from the word it modifies (e.g. adjective from noun)
- Do not break a prepositional phrase
- Do not break a person's name or titles from a personal name
- Do not break a line after a conjunction
- Do not break an auxiliary verb from the word it modifies

(ibid, p. 10-11)

Even more detailed are the recommendations by Neves in Portugal:

- Do not begin a second sentence on the same line
- Do not separate a conjunction from the phrase it introduces
- Do not separate an article from the noun

- In a noun phrase do not separate the noun from the modifier
- When there is a relative clause, break the sentence before the relative pronoun[19]
- When possible, keep the verb and its complements on the same line
- Do not separate a transitive verb from its direct complement
- Do not separate a compound verb phrase
- Do not separate a preposition from the accompanying noun phrase
- Do not separate titles from a name
- Do not segment a subtitle within a word (even if it is hyphenated)

(2004, p. 41-45, my translation)

Interlingual guidelines

According to Karamitroglou (1998, no page), segmentation should occur at the highest syntactic node possible. He provides the following example to make his point:

The destruction of the
city was inevitable.

The destruction of the city
was inevitable.

and goes on to support his claim:

Out of the two segmentations, it is the second that flows as more readable. This occurs because the higher the node, the greater the grouping of the semantic load and the more complete the piece of information presented to the brain.

---

[19] Although Neves refers to relative clauses and relative pronouns, her example involves a complementizer *So quero dizer que o embrulho ja chegou* (*I only want to say **that** the parcel has already arrived*).

Ivarsson and Carroll also place emphasis on the need to break lines in sense blocks and/or grammatical units so that words connected by logic, semantics or grammar appear on the same line (1998, p. 77; 1998b, no page). Although they do not use linguistic terms as such, they advise against splitting an adjective from the noun it modifies, a preposition from its noun phrase, as well as against starting a new sentence on the same line. They also introduce a sense of gradation, as they provide a series of examples each deemed less acceptable than the others. Thus:

Having said that, he put the green shirt on
the shelf.

is considered unacceptable and worse than

Having said that, he put the green
shirt on the shelf.

which in turn is considered less acceptable than

Having said that, he put
the green shirt on the shelf.

which again is considered less acceptable than the ideal version:

Having said that,
he put the green shirt on the shelf.

Similar considerations are voiced by Díaz Cintas (2001, p. 120-121). He points out that "la división de la frase debe ser lo más lógica posible" and provides the following examples:

- Never separate a noun from its article
- In the case of a complex verb phrase, never separate the auxiliary verb from the participle
- In the case of a conditional sentence, use one line for the protasis and another for the apodosis

These guidelines are further elaborated by Díaz Cintas and Remael (2007, p. 176-177), who base their discussion on Karamitroglou's general guideline of segmenting text at the highest syntactic node possible and proceed to make the following specific recommendations:

- Do not hyphenate words
- If a subtitle consists of two or more sentences, put one sentence on each line
- If a subtitle consists of a sentence with two subordinated or coordinated clauses, use one line for each clause (if it is impossible to have both on the same line)
- Do not separate adjective from noun or adverb
- Do not separate adverb from verb
- Do not separate article from noun
- Do not separate preposition from prepositional phrase
- Try not to split to-infinitives
- Try not to split phrasal verbs
- In the case of compound verbal forms, do not separate the auxiliary from the lexical verb (past participle or infinitive)
- Avoid separating a verb from its direct or indirect object.

Empirical research

Line-break segmentation is one of the few subtitling issues that are "disregarded, though not neglected, both in theory and in practice" (Perego, 2008, p. 37). Very

little research exists and only two relevant studies seem to be available. The first one comes from the field of SDH. As part of an extensive research, Kyle and his associates analysed four two-minute extracts from two programmes (*The Nine O'Clock News* on BBC and ITN's *News at Ten*). Both programmes were live, but two of the four extracts, one from each channel, came from parts of the programme "where the story ha[d] already gone out earlier in the day and it [was] therefore possible to prepare sections or all of the subtitle script" (1992, p. 79). The main finding of this part of the research was that "there is a great deal of improvement to be achieved" as far as line breaks are concerned (ibid, p. 90). In particular, focusing on the extracts that could have been prepared in advance, the researchers found that in a total of 21 and 19 subtitles respectively, 77% of line breaks were correct for the BBC story and only 45% for the ITN story. Most common errors involved splitting articles or adjectives from nouns, splitting verbs or adverbs, splitting prepositions from their appropriate phrase and splitting conjunctions from their clause, which might or might not be an error (ibid, p. 83-84).

The second research into segmentation is still in progress (Perego, 2008). Apart from being one of the two studies investigating the issue of line breaks, this research is interesting in that it discusses both interlingual subtitles for hearing viewers and intralingual subtitles for D/deaf and HoH people. After offering theoretical evidence that linguistically coherent segmentation of text can improve processing, Perego presents the results of "a qualitative analysis of text distribution over different subtitle lines [undertaken] in order to detect the various types of undesirable line-breaks that may occur" (2008, p. 36). The qualitative analysis is based on an extensive corpus consisting of different film genres and different source and target languages[20] and uses the traditional micro- to macro-levels of language organisation. Thus, segmentation is studied both at the level of the phrase (intra-clausal segmentation) and at the level of the clause/sentence (inter-clausal segmentation). At

---

[20] The corpus consists of 15 films covering a wide variety of genres and release dates (from the early 1960s until today). It comprises original soundtracks in English, Dutch, German, Italian and Hungarian, interlingual subtitles in English, Italian and Hungarian, and intralingual subtitles in English and Italian.

intra-clausal level, the following types of inappropriate line breaks are identified grouped together under three categories:

- Noun phrases
- Pre-modifier (determiner) + Noun: e.g. *a / four-hour documentary, some / life*
- Proper Nouns: e.g. *the "Susan / Marie"*
- Title + Proper Noun: e.g. *Miss / Sherwood*
- Noun-noun sequences (in English): e.g. *construction / work*
- Nominals: crystallised forms (e.g. *credit / card*), collocations (e.g. *animal / urges*), idiomatic expressions (e.g. *dato / i numeri* 'go mad')

- Prepositional phrases
- Preposition + Noun: e.g. *della / barca, from / France*
- Ambiguous structures: e.g. *The night before / he left* when intended meaning is [The night] [before he left]

- Verb phrases
- Lexical verb + Auxiliary: e.g. *had / been playing*
BUT: *we could / talk* and *I just want / to talk* are considered less distracting
- Fixed verb patterns, such as the causative structure *fare /* infinitive
- Highly specific verb-particle combinations: e.g. *get along / with*
- Constructions with infinitive markers: e.g. *try to / figure out*

Of the three categories, prepositional phrases are thought to be easier to present on the same line, as "both simple and complex prepositions are short enough to be moved around conveniently without upsetting the established norms governing the maximum number of [characters per line]" (2008, p. 50). When that is not possible, Perego suggests rephrasing the sentence or even exceeding the maximum number of characters per line. However, it can be said that both solutions should only be applied with great care. Particularly in the case of intralingual SDH, rephrasing the original would be a debatable solution and would need to be used cautiously given many

viewers' preference for verbatim subtitling (see subsection *Reduction and verbatim subtitling in SDH* in section 2.4.6).

At the inter-clausal level, instances of poor segmentation are detected at the following levels:

- Coordination

    - Coordinating conjunction + clause/sentence: e.g. *Come puoi fare la marmellata e / fingere che non sia successo niente?*

    - Negative constructions: e.g. *If the boys are half English it does not / mean they're bad Pakistanis.*

- Subordination

    - Subordinating conjunction + subordinate clause: e.g. *I felt that / you just wanted ...*

    BUT the degree of acceptability increases if also part of the subordinate clause is retained in the upper line

    - If / protasis: e.g. *If / you agree ...*

    - Comparative constructions: e.g. *similar to the / wife*

Looking at the categories identified by Perego, many are similar to those previously suggested in relevant guidelines, while there are also some interesting additions. However, there appear to be some problems with some of the categories. For instance, it is not clear why negative constructions are classed under coordination or comparative constructions under subordination. More interestingly, collocations and idiomatic expressions are grouped under nominals and noun phrases. Nevertheless, many of the examples provided involve phrases of other types:

Collocations: *facevamo / una festa* ('given a party'), *girare a destra* ('turn right')
Idiomatic expressions: *dato / i numeri* ('go mad')

It would also be highly informative to have some quantitative data in addition to the qualitative analysis. Are the categories presented here those where errors are most common? How common? And how are errors distributed among the different categories? Is poor segmentation consistently a problem? Are there any differences in error rates according to the medium (DVD vs. cinema) or release date? Given that the corpus is multilingual and includes both interlingual and intralingual subtitles, are there any differences in the types and frequency of line-break errors?

Finally, it is interesting to note along with Perego that poor and satisfactory segmentation very often co-exist (2008, p. 55). Potential reasons include:

> a lack of sufficiently strict in-house guidelines in subtitling companies, scant attention to or underestimation of the problem on the part of the subtitler, or indeed excessive reliance upon subtitling programmes which split lines exclusively according to pre-defined technical instructions.

To these factors, which are all quite plausible, we need to add that poor segmentation might also be a conscious choice (for instance when the subtitler is unwilling to reformulate the original sentence) or even a necessity (for instance if it is impossible to rephrase the subtitle or exceed the maximum number of characters per line). Furthermore, we need to take into account the fact that quite often subtitlers work to very strict deadlines and under great time pressure, in which case it might be impossible to pay attention to "details", such as the application of proper line breaks. Last but not least, it is also possible that different subtitlers evaluate segmentation errors differently with the same line break being classified as proper by some and as poor by others (for more on this issue, see section 6.2).

3.3.5. Reduction

As mentioned in section 2.4.6, reduction has come to be regarded as one of the defining features of subtitling and is often discussed in subtitling literature. This is

particularly true in the case of SDH, where verbatim subtitling versus editing is a highly charged issue. Focus here will be mainly on deletion rather than condensation.

SDH guidelines

Both Baker (1981) and Baker et al. (1982) take editing for granted. In their view, "verbatim subtitling rates usually prohibit comfortable subtitle reading, and leave no time for lip-reading" (Baker et al., 1982, p. 34). Accordingly, "the job of the subtitler is to preserve as much as possible of the content and flavour of a ... sound-track while still maintaining a manageable reading rate" (ibid, p. 33). They put forward two methods for editing subtitle text: the propositionalising or re-writing technique, "which involves taking a text (or utterance) and reproducing it in one's own words" and the "red pencil" technique, which involves simply deleting words from the text (Baker, 1981, p. 12). Baker seems to prefer the propositionalising technique or a combination of the two methods; he also advises that the propositionalised version should remain as close to the original script as possible, since many viewers might use their residual hearing or try to lip read (Baker, 1981, p. 12). Later, however, and for the same reasons, Baker et al. prefer red-pencil to rewrite edits, confining the latter "to intransigent cases" (1982, p. 34-35).

Regarding omission, two categories of linguistic items that could be omitted are suggested:

- Fillers, such as *well, you know, as I say, if I may say so, really, sure, so, eh?, as I've indicated*, etc: Such expressions are used to pad speech and can be readily omitted. However, when time permits, it is a good idea "to pepper a dialogue with such fillers, to convey something of the informal tone of a conversation or the idiosyncratic speech habits of an individual speaker" (Baker, 1981, p. 14)

- Adjectives, especially in tautologous expressions such as *ordinary everyday things, great big box*: These can often be omitted without much loss of meaning (Baker et al., 1982, p. 35).

The researchers also advise against omitting redundant items, such as *that* in sentences like *The man that we saw was drunk*. Omitting such items only increases the complexity of the sentence thus requiring more time in order to be processed successfully (Baker, 1981, p. 15; Baker et al., 1982, p. 36-37). Moreover, it is recommended that swear words should not be censored unless they are frequently repeated and omitting some of them is necessary for spatiotemporal reasons (Baker et al., 1982, p. 39).

Similarly, the ITC guidelines also mention the need for editing due to temporal constraints:

> Careful and sensitive editing is ... needed in order to produce subtitles which will suit the intended audience, while still conveying the full meaning of the dialogue or commentary within the limitations set by the pace of the programme. (1997, p. 2; 1999, p. 4).

However, the only recommendation provided is too general:

> Without making unnecessary changes to the spoken word, construct subtitles which contain easily-read and commonly-used English sentences in a tidy and sensible format. (1997, p. 3)

Editing is only discussed in further detail with regard to subtitling for children with reduction figuring as one of the three editing devices suggested (ITC, 1997, p. 18; ITC, 1999, p. 19). The same guidelines have been adopted by the BCI (2005, no page); again, editing is not discussed in the context of general SDH.

The BBC Subtitling Guide, on the other hand, offers more detailed recommendations regarding editing. Attributing the need for editing to the fact that people generally speak much faster than it is possible to read the text of their speech, it specifies that "the subtitler must always edit according to the amount of time available" and the recommended standard timings (1998, p. 12). However, it also states that if there is plenty of time for verbatim or near-verbatim subtitles, then no editing or minimum editing respectively should be used, as the aim should be to provide viewers with as much access to the programme's soundtrack as possible. In addition, subtitlers are advised to keep editing faithful to the original by not rewording unnecessarily, by retaining the start and end of a speaker's utterance when they are visible on screen and their words can be lip-read, and by remaining close to the speaker's style of speech taking into account register, nationality, era, etc. (ibid, pp. 12-13). Moreover, it is recommended to avoid editing the following:

- words like *but*, *so* or *too*, as they are often essential for meaning

- phrases such as *you know*, *well*, *actually*, as they often add flavour to the text

- names used to address people, as they can be essential for following the plot

- verb tenses, as the change may produce a nonsense sentence

- words that can be clearly lip-read, as viewers may feel cheated

- strong language, unless it is absolutely impossible to edit elsewhere in the subtitle

- sloppily constructed or incoherent sentences, when the incoherence is the desired effect – this is more applicable to dramas rather than factual programmes, such as the news and documentaries

- hesitations, such as *um* and *er*, if they are important for characterisation or the plot; if the hesitation is incidental, and this is again more likely to be the case in factual programmes, then they can be edited out

- characters' catchphrases

- song lyrics; these should be verbatim, particularly in the case of well-known songs (some freedom, however, is allowed with unknown songs that have been especially written for a programme)

<div align="right">(ibid, pp. 11, 12-13, 18, 20, 31)</div>

The guidelines by the Deafness Forum of Australia also point out the need for editing by conceding that "the pace of a television program, movie, DVD or video sometimes means that it would be impractical to include every word or sound in caption form on the screen" (2004, no page). However, they also differ from the majority of guidelines mentioned above in that they explicitly state that "where time allows, a caption should be verbatim" (ibid). Moreover, like in Baker (1981), Baker et al. (1982) and BBC (1998), it is recommended that reduction should remain faithful to the original script preserving vocabulary and sentence structure as much as possible (Deafness Forum of Australia, 2004, no page). The guidelines by the DCMP are along the same lines:

> Editing is performed only when a caption exceeds the specified presentation rate limit... Proper editing should maintain both the original meaning/content and meet presentation rate requirements. (2007, p. 13)

The only exceptions to this general guideline are "when any person is quoted, a well-known person is speaking on-screen, poems and other published works are quoted, and/or song lyrics are sung" (ibid). These must be rendered verbatim even at the expense of presentation rate requirements. In addition, it is recommended not to subtitle information that is already shown on the screen.

Questioning the adequacy of verbatim subtitling for hearing-impaired, but also hearing viewers, Neves seems to be the most fervent supporter of editing or adaptation in SDH with reduction being only one of the adaptation techniques she discusses in her guidelines (2004). To her, the original script consists of three types of elements:

- indispensable items, whose presence in the subtitle is obligatory
- partially dispensable items, that can be condensed
- dispensable items, that can be readily omitted

(ibid, p. 32)

Focusing on omission, it is further specified that it only involves redundant or superfluous elements or elements that are present in other semiotic components, such as the image. Examples include interjections, comment phrases that are obvious from the context, repetitions and fixed phrases, such as greetings (ibid, p. 34). Two general recommendations are provided:

i) Omission must be applied with great care, as what might be obvious to hearing viewers, might be indispensable to Deaf and HoH people
ii) Omission must only be used when there is a lack of reading time

<div align="right">(ibid, my translation)</div>

Interlingual guidelines

Editing is also present in guidelines regarding interlingual subtitling. In Karamitroglou's words:

> The subtitler should not attempt to transfer everything, even when this is spatio-temporally feasible. The subtitler should attempt to keep a fine balance between retaining a maximum of the original text (essential for the comprehension of the linguistic part of the target film), and allowing ample time for the eye to process the rest of the non-linguistic aural and visual elements (essential for the appreciation of the aesthetic part of the target film). (1998, no page)

Using the categories proposed by Baker (1981) and Baker et al. (1982), he suggests three categories of linguistic items that can be omitted:

- Padding expressions (such as *you know, well, as I say*, etc.): Such expressions are usually empty of semantic load and can be readily omitted
- Tautological cumulative adjectives/adverbs (such as *great big*, *super extra*, *teeny weeny*): Since the first adjective in each pair is used emphatically, they can be replaced by single-word equivalents (e.g. *huge*, *extremely*, *tiny*)

- Responsive expressions (*yes, no, ok, please, thanks, thank you, sorry*): These words have been found to be recognised by most European people and can therefore be omitted. The only exceptions are when they are not uttered clearly or when they are used in a slang, informal or colloquial version.

In addition, it is recommended to retain and translate word-for-word items such as proper nouns (e.g. geographical names such as *Los Angeles*), words that the source and target languages have in common, and words that the target language has lent to or borrowed from the source language (e.g. *μαθηματικά* and *mathematics*). This is deemed necessary, given that if such words are omitted, viewers might suspect that the subtitles are inaccurate or incorrect.

For Ivarsson and Carroll, reduction is necessary in cases where the dialogue is too fast to fit in the limited space available on screen and/or when many people are talking at the same time (1998, p. 85). Although they acknowledge that there can be no specific rules for reducing text, they prefer omission to paraphrase, as it is easier and less irritating to viewers (ibid, p. 86). To the two categories suggested by Baker (1981), Baker et al. (1982), and Karamitroglou (1998), i.e. fillers/padding expressions and tautologies, they add a third category of omissible elements, repetitions. However, they point out that words belonging to these categories cannot be ignored without careful consideration. In the case of speech containing confused syntax, slips of the tongue and many repetitions (such as that found in interviews), their recommendation is to give "the gist of what the person has said in a reasonably coherent form, including just enough of the speaker's inarticulate diction to give the viewers a hint of the delivery" (ibid, p. 87). They also advise against omitting exclamations, simple forms of address or greeting, and names called out, as these phrases are not readily perceived by the hundreds of thousands or even millions of people with hearing impairment (ibid, p. 93).

The same advice features in the "Code of Good Subtitling Practice" proposed by Ivarsson and Carroll (1998b, no page). Two further guidelines are provided:

- Where compression of dialogue is necessary, the results must be coherent
- Obvious repetition of names and common comprehensible phrases need not always be subtitled[21]

Díaz Cintas also believes that "los subtítulos no son ni pueden ser una traducción integral de los diálogos de la versión original" (2001, p. 123). On the basis of Ivarsson and Carroll (1998b), he provides two guidelines:

- Dialogue compression must respect coherence and cohesion
- Obvious repetitions of proper names and expressions easily understood on account of similarities between the two languages need not be always subtitled

(2001, p. 115, my translation)

He also points out that in reducing the original, the subtitler should always strive to respect the syntax and style of the original (ibid, p. 126).

The belief that subtitling cannot be a complete and detailed rendering of the original is also present in Díaz Cintas and Remael (2007, p. 145). Not only that, but like Karamitroglou (1998), they argue that "a complete translation is, in fact, not required" (ibid). As a result, reduction both in the form of condensations and omissions is unavoidable. No specific guidelines are provided, as the authors point out that no foolproof reduction rules exist (ibid, p. 148). The only general advice offered is that "the subtitler must act on the principle of relevance" and should first watch the entire programme in order to be in a better position to judge the relevance of particular items for understanding the programme and thus in a better position to decide whether these items should be omitted or included in the subtitles (ibid).

---

[21] This guideline is actually contradictory to the previous guideline available in Ivarsson and Carroll (1998) to avoid omitting exclamations, simple forms of address or greeting, and names called out.

Empirical research

Reduction is a much researched area in comparison to other subtitling issues. Research has mainly focused on two questions: how much is reduced and what is reduced.

How much is reduced

Within intralingual SDH, on the basis of their corpus consisting of 12 programmes but only 262 subtitles, De Linde (1997) and De Linde and Kay (1999) found that on average 4.14 words per subtitle were omitted from the original dialogue. While the degree of editing was marked for all programme types apart from film, it varied widely between different types with the average ranging from 1 to 8 words per subtitle. For instance, while the figure for the two documentaries analysed (both from BBC2) was among the lowest at 2.62 words per subtitle, the respective figure for games/chat shows was 7.82 words per subtitle (De Linde and Kay, 1999, p. 50). The difference was attributed to the difference in discourse modes: programmes with large sections of interactive dialogue have higher speech rates than narrative-based programmes (such as documentaries) and also have higher subtitle rates.

The next study within SDH and admittedly, the only large-scale research available to date is the one by Jensema et al. (1996). From a total of 183 programmes and 13 genres, two programmes per genre were randomly selected and a ten-minute segment was analysed from each. It was found that among the 26 programmes investigated, 94% of the original audio had been subtitled on average (ibid, p. 288). The figures for the two documentaries examined were among the highest at 100% and 99% respectively (ibid, p. 289).

Relevant studies have also been conducted within the field of interlingual subtitling for hearing viewers. On the basis of his small corpus of three programmes broadcast on Norwegian television (a detective series, a comedy series and a science fiction series), Lomheim reported that "a significant reduction takes place during the

transfer from dialogue to subtitles, and that the extent of this reduction varies between programmes" (1999, p. 191). Thus, 78% of the original had been retained in the detective series, 77% in the comedy series and only 60% in the science fiction series (ibid).

Variation was also one of the main results reported by Koljonen (1996). On the basis of six programmes broadcast on Finnish television[22], she found that quantitative cuts ranged from 19% to 40% (1995, p. 184). The reduction rate of 40% was also reported by Díaz Cintas, who analysed the Spanish subtitles of Woody Allen's *Manhattan Murder Mystery* (2003b, p. 202). This rather high percentage was in this case mainly attributed to the talkative nature of Woody Allen's films and the extremely fast speed of delivery.

A much larger interlingual corpus investigating reduction is the one created by Pedersen, who examined six episodes from a sitcom (*The Simpsons*)[23] and six episodes from a soap opera (*The Young & The Restless*) and their respective subtitles in Danish (2003, p. 3). A quantitative reduction of 31% was reported overall. The reduction rate was somewhat higher for sitcoms (35%) than soap operas (29%) and this small variation was thought to be "due to the more rapid dialogue that is one of the hallmarks of the witticisms of the sitcom genre and which forces the subtitler to condense slightly more" (ibid, p. 11). It was also found that reduction rates varied a lot within individual subtitles and segments (however, no relevant figures are provided).

The largest interlingual subtitling corpus with a view to studying reduction seems to be the one by Georgakopoulou (2003). Her corpus consists of 12 films/TV series belonging to different genres (drama, action films and comedies) and comprises both their English (American) scripts and respective Greek subtitles. It was created in order to investigate "the extent to which filmic characteristics may determine the

---

[22] The six programmes were: an Australian series, an Australian docudrama, a British drama based on Dickens' novel *Great Expectations*, an American docudrama about patients at a mental hospital, a British comedy series and a Scottish detective series.
[23] Although Pedersen classifies *The Simpsons* as a sitcom, some might find this classification questionable.

degree of reduction in a film" (Georgakopoulou, 2003, p. iii). Reduction percentages in the corpus were calculated using two metrics. The first, traditionally used in relevant research, involved calculating the subtitle rate vs. speech rate ratio (the number of words in subtitles vs. the number of spoken words). The second metric suggested by Georgakopoulou herself involved linguistically analysing the subtitled text on the basis of Lomheim's model of subtitling strategies (see section 3.3.6) in order to identify the subtitling strategy employed by the subtitler for each word of the original dialogue; the reduction percentage was, in this case, calculated by adding the recorded omission and compression percentages. This second metric is thought to be more informative, as it involves an in-depth analysis of the language used in the subtitles compared to that in the original dialogue, but cannot be used when comparing large numbers of programmes. Relevant figures from Georgakopoulou's corpus are shown in Table 8.

Regarding the two metrics themselves, on the basis of the figures in Table 8, it was found that variation between reduction percentages in the same film/programme varied from as little as ±0.15% to as much as ±7.93%. As a result, the suggestion was put forward that

> It is thus safe to assume that the subtitle/speech ratio percentage may be used as a gross indicator of the extention to which reduction strategies have been used in the subtitling process, which would be useful when comparing large numbers of subtitle files *originated* (i.e. created) to different company or national styles. (Georgakopoulou, ibid, 193, emphasis in original).

However, as Georgakopoulou herself points out much later (ibid, p. 283), the actual reduction percentages calculated on the basis of the subtitling strategies used are only roughly comparable to the subtitle/speech ratios of the films. More research is undoubtedly required to investigate the relationship between the two metrics.

Table 8. Reduction statistics by Georgakopoulou (2003, pp. 158 and 190)

Name of programme	Genre	Subtitle/Speech ratio	Reduction in translation
Casablanca	Drama	50%	45.82%
Kramer vs. Kramer	Drama	41.72%	49.65%
Titanic	Drama	29.24%	29.39%
As Good As It Gets	Situation comedy	37.73%	34.30%
Dracula: Dead and Liking it	Situation comedy	25.94%	19.42%
My Best Friend's Wedding	Situation comedy	46.38%	49.09%
Star Trek: Generation	Action film/ Science fiction	43.94%	36.63%
X Files: The Walk	Action film/ Science fiction	40.35%	40.23%
Blow Up	Action film/ Gangster, thriller	25.19%	17.55%
The Godfather	Action film/ Gangster, thriller	42.14%	46.68%
Pulp Fiction	Action film/ Gangster, thriller	44.42%	38.53%
Usual suspects	Action film/ Gangster, thriller	39.59%	32.61%

Regarding the effect of film genre on the degree of reduction, it was reported that there were no conclusive findings and that a generalisation might not actually be possible, "as most films would probably fall in the category of hybrids, making a [sic] classification difficult" (ibid, p. 282). While such a generalisation might not indeed be possible for different film genres (although further research, as acknowledged by Georgakopoulou herself, would be required), it might be possible when discussing different genres, such as films, documentaries, news, etc. Further research involving more (representative) films/TV series/programmes and genres would be required in that respect. In addition, it would be interesting to extend such

research in order to include intralingual SDH subtitles, too. Applying Georgakopoulou's metric in this context would also be expected to be more objective than in the case of interlingual subtitles, as it does away with the difference in lexical structures between different languages.

Subtitling strategies rather than mere word counts are also at the heart of Gottlieb's study (1992). In his analysis of the Danish subtitles for Mel Brook's *Young Frankenstein* and using a total of ten strategies, three of which are reductive in nature, he found that "only 16% of the original verbal segments ... suffer a loss of (semantic or stylistic) information in the process of translation" (ibid, p. 167). Since the basic unit of investigation seems to be the subtitle (the term "verbal segment" that is used is not defined), Gottlieb's findings cannot be compared with other studies. More importantly, as Jaskanen points out, the strategies suggested by Gottlieb are overlapping and subjective, even awkward at times, and the examples provided fail to shed any light (1999, p. 11).

<u>What is reduced</u>

There appear to be no relevant studies within SDH and only a few studies from the area of interlingual subtitling for hearing viewers. The only study available in the UK seems to be the one by De Linde (1995). Analysing two very different French films (*Delicatessen* and *La Lectrice*) and their subtitles in English, she found that the style of each film had a sharp effect on deletions both quantitatively and qualitatively. Indeed, *Delicatessen*, a film with a high degree of fast "oral coded" discourse, was found to contain more deletions than *La Lectrice*, a film, that as implied by its title, consisted mainly of slow "written coded" narrative. The statistics are quite telling: in a total of 300 subtitles per film, there were 246 deletions in *Delicatessen* and only 73 in *La Lectrice* (ibid, p. 16). Moreover, when deletions were categorised, it was found that their distribution and frequency also differed. Thus, although the largest category in both cases was "Markers of Interaction", deletions under the other headings differed greatly. Deletions were categorised under the following categories:

Markers of interaction – interactional terms, e.g. *écoutez-moi*, *tu sais quoi*, modals, tag questions, subject ellipsis, etc.

Gestural language – exclamations, affirmations/negations, fillers (e.g. those language items which indicate that the speaker wishes to continue speaking, e.g. *euh*, *ben*, *ah*, etc.)

Repetitions – repetition of words or propositions

Additional comment – supportive statements

Additional information – e.g. additional information of place, actions, deictic expressions such as *là-bas* and *maintenant*

Whole dialogue – e.g. background chatter, phatic communion, adjacency pairs

Proper names

and the statistics shown in Table 9 were obtained.

As pointed out by De Linde, "none of the categories [proposed] are mutually exclusive. For instance, proper names can feature in a mutual greeting, thus being described formally as an adjacency pair and classified under 'Whole dialogue'" (ibid, p. 16). Although most of the categories are rather straightforward and are further specified by examples (see definitions above), it would be extremely useful to have examples presented in their context, as well as to have the complete list of deletions from both films.

Table 9. Deletion statistics by De Linde (1995, p. 16-17)

	Delicatessen		*La Lectrice*	
TYPES OF DELETION				
Markers of Interaction	63	25.6%	21	28.8%
Gestural language	59	24.0%	15	20.6%
Repetitions	32	13.0%	5	6.9%
Additional comment	32	13.0%	12	16.4%
Additional information	29	11.8%	16	21.9%
Whole dialogue	17	6.9%	2	2.7%
Proper names	14	5.7%	2	2.7%

The only other research available is that of Kovačič (1992, 1994, 1998). Looking at English to Slovene subtitles and drawing on Halliday's system of language functions, she observed that the choice of deletions in subtitling is based on the function of their language sign. Thus, it was found that ideational elements tended to be preserved more than interpersonal and textual elements, while the difference between the latter two categories was not statistically significant (1992, quoted in 1998, p. 78). Unfortunately, as mentioned in section 2.4.6, the original study of 1992 (Kovačič's PhD thesis) is not readily available to consult, as it was written in Slovene. Regrettably, no details about the programmes studied and the figures obtained are quoted in subsequent papers. More importantly, it is not clear how Kovačič defines each of the linguistic functions adapted from Halliday and what linguistic elements each encompasses. Undoubtedly, as in the case of De Linde, it would have been highly informative to have access both to specific examples and subtitles, as well as to the entire list of deletions as categorised according to their linguistic function.

3.3.6 Overall strategies used

Deletion and condensation are only two of the strategies available to subtitlers when performing their task. Nevertheless, together, under the heading of reduction, they seem to have monopolised research interest (this is particularly true for deletion). Indeed, very little research has been dedicated to identifying the entire range of strategies at the subtitler's disposal. Furthermore, in most cases, this has been limited to identifying "an exhaustive taxonomy of strategies available to the subtitler when confronted with ECRs [extralinguistic culture-specific references] in the original dialogue" (Gottlieb, 2005b, p. 203); such taxonomies have been suggested by among others Nedergaard-Larsen (1993), Gottlieb (2005b) and Pedersen (2007).

Guidelines

In both intralingual SDH and interlingual guidelines, any reference to strategies is restricted to verbatim translation/transcription and reduction, in particular deletion (as discussed in section 3.3.5). This is hardly surprising, as this issue would

effectively cover the entire subtitling process. Apart from the difficulty, even impossibility of specifying such general guidelines (a difficulty well manifested in trying to compile guidelines regarding deletion alone), it would first be necessary to adopt a descriptive approach before making any attempt at prescriptivism.

Empirical research

As mentioned above, there is not much research into the range of strategies available to subtitlers. The only study relevant to SDH comes from the related field of respeaking (see definition in section 2.3.3). The rest of the research available and presented below comes from the field of interlingual subtitling.

Eugeni

Working in the area of respeaking, Eugeni (forthcoming, b) has suggested the taxonomy of strategies available to respeakers shown in Table 10. The suggested taxonomy is based on the concept of "idea unit" used by the ITC (now Ofcom) in its recommendations for real-time subtitling and defined as "where a proposition or key information is given" (1999, p. 25). Adopting the concept but finding the ITC definition lacking, as it is still unclear what an "idea unit" actually is, Eugeni (ibid) draws a distinction between two kinds of idea units providing the following definitions:

- micro-units: all relevant pieces of information in a clause (lexical units like subject group, verb group, adverbs, etc.) contributing to the global meaning of a bigger meaning unit
- macro-units: every clause providing a finite self-standing set of information (e.g. a defining relative clause, an embedded clause, an incidental clause, etc.)

Table 10. List of respeaking strategies according to Eugeni (forthcoming, b)

Strategies	*Definition*
Repetition	A micro-idea unit has been faithfully repeated in the target text (including features of orality like false starts, hesitations, etc.)
Omission	A macro-idea unit has not been rendered at all
Alterations	A macro-idea unit has been altered, that is slightly or heavily modified in form at a micro-unit or word level
--- *Reduction*	A macro-idea unit has been partially omitted or partially or totally reformulated
--- Omission	One or more micro-units or lexemes within a macro-idea unit have been deleted
--- Qualitative	Deletion of meaningful lexemes or phrases
--- Quantitative	Deletion of grammatical words, features of speech and lexemes not useful to the semantic meaning of the unit
--- Compression	The form of a macro- or micro-idea unit has been changed in order to make it shorter
--- Qualitative	Compression that alters the meaning of the rest of the macro-idea unit
--- Quantitative	Compression that has no real semantic relevance on the rest of the macro-idea unit
--- *Expansion*	A macro-idea unit has been explained/disambiguated by means of more characters than the source text
--- Qualitative	Information is added to the source text
--- Quantitative	Just characters are added to the source text
--- *Mistakes*	A macro-idea unit has be altered by a human or software mistake (Note: mistakes are not the result of voluntary strategies, but rather in the words of Eugeni "the result of an unconscious or mechanic (unpleasant) incident"

Regarding the strategies themselves, it is important to note that omission features twice, once as an independent category and again as a sub-category of reduction. This distinction has been drawn solely on the basis of what is reduced: macro-idea units in the first case and micro-idea units in the second. Although this distinction is sensible, it might be more accurate to classify omission of macro-idea units under reduction and alterations, as it too alters the source text.

A further distinction introduced by Eugeni is the one between quantitative and qualitative omissions, compressions and expansions. The difference is that while quantitative alterations do not have "a real semantic relevance on the rest of the macro-idea unit", qualitative ones do (ibid). Thus, for instance, in the case of micro-unit omissions, qualitative omission entails the deletion of meaningful lexemes or phrases, whereas quantitative omission involves the deletion of grammatical words, features of speech and lexemes not useful to the semantic meaning of the macro-idea unit (see Table 10). Although there is a major difference between the two categories and they should indeed be distinguished from one another, the terms "quantitative" and "qualitative" do not seem appropriate. Discussing omission in quantitative terms would entail counting every single deletion and not just instances that have no impact on meaning. To distinguish between deletions with semantic relevance and deletions without semantic impact, it would be more accurate to use instead of the terms "qualitative" and "quantitative", the terms "semantic" and "non-semantic". These terms are in fact used by Eugeni himself later in his article when he applies his taxonomy to the analysis of BBC News. The BBC News is a rather complex news genre, as it consists of five different moves (Headlines, News reports, Live reportages, Weather forecasts, News summary), ranging from written to be read to oral-oral and exhibiting different rates of grammatical intricacy, lexical density, reading speed/speed of delivery, orality features, etc. Accordingly, both the overall and inner distribution of strategies varies, as illustrated in Table 11.

Table 11. Distribution of strategies in BBC News by Eugeni (forthcoming, b)

Move	Omissions	Repetitions	Alterations													Total
			Mistakes			Expansions			Reductions						Total	
			Human	Software	Total	Semantic	Non Semantic	Total	Omissions		Total	Compressions		Total		
									Semantic	Non Semantic	Total	Semantic	Non Semantic	Total		
Headlines	18.2	48.2	40.0	60.0	3.0	75.0	25.0	8.7	13.7	86.3	44.4	65.0	35.0	55.6	88.3	33.6
News Reports	16.3	50.4	43.2	56.8	1.9	31.6	68.4	10.4	18.8	81.2	37.1	58.6	41.4	62.9	87.7	33.3
Live Reportages	21.0	45.1	68.3	31.7	7.5	30.1	69.9	1.1	4.6	95.4	37.6	43.9	56.1	62.4	91.4	33.9
Weather Forecasts	25.3	32.6	71.9	28.1	1.3	26.5	73.5	4.3	12.1	87.9	53.2	59.9	40.1	46.8	94.4	42.1
News Summary.	14.5	57.2	41.1	58.9	1.2	68.7	31.3	8.8	44.6	55.4	30.3	46.7	53.3	69.7	90.0	28.3
Average.	18.9	47.4	62.2	37.8	3.4	35.0	65.0	7.1	13.8	86.2	37.8	50.5	49.5	62.2	89.5	33.7

<u>Gottlieb</u>

Gottlieb seems to be the first researcher to have proposed a set of translation strategies used in subtitling (1992). His typology includes ten strategies some of which are particular to subtitling, while the rest are not as shown in Table 12.

Table 12. Typology of subtitling strategies according to Gottlieb (1992, p. 166)

Type of strategy	Character of translation	Media specific type?
1) Expansion	Expanded expression, adequate rendering (culture-specific references etc.)	No
2) Paraphrase	Altered expression, adequate rendering (non-visualised language-specific phenomena)	No
3) Transfer	Full expression, adequate rendering ('neutral discourse – slow tempo)	No
4) Imitation	Identical expression, equivalent rendering (proper nouns, international greetings etc.)	No
5) Transcription	Anomalous expression, adequate rendering (non-standard speech, etc.)	Yes
6) Dislocation	Differing expression, adjusted content (musical or visualized language-specific phenomena)	Yes
7) Condensation	Condensed expression, concise rendering (normal speech)	Yes
8) Decimation	Abridged expression, reduced content (fast speech of some importance)	Yes
9) Deletion	Omitted expression, no verbal content (fast speech of less importance)	Yes
10) Resignation	Differing expression, distorted content ('untranslatable' elements)	No

Of the suggested strategies, *decimation*, *deletion* and *resignation* are grouped together as those involving a loss of semantic or stylistic content; of the three, the most abortive is resignation and it is also the only one not particular to subtitling. *Condensation* is distinguished from these three strategies and in particular *decimation*, as according to Gottlieb it does not involve a loss of meaning or stylistic content of the original, but only a loss of redundant oral language features (ibid, 166-167). For this reason, it is grouped along with the remaining strategies as providing correspondent translations of the segments involved. Gottlieb then uses this typology

to analyse the Danish subtitles of Mel Brook's *Young Frankenstein* as broadcast on Danmarks Radio and provides the figures shown in Table 13.

Table 13. Distribution of strategies in Young Frankenstein adapted from Gottlieb (1992, p. 167)

Type of strategy	Number of verbal segments	Frequency
Transfer	439	44.9%
Paraphrase	142	14.5%
Condensation	126	12.9%
Decimation	79	8.1%
Deletion	72	7.4%
Expansion	48	4.9%
Transcription	44	4.5%
Dislocation	13	1.3%
Imitation	10	1.0%
Resignation	5	0.5%
	978	100.0%

When discussing these figures, Gottlieb focuses exclusively on the three reductive strategies (*decimation*, *deletion* and *resignation*). His main suggestion, as mentioned in section 3.3.5, is that only a small percentage of the original suffers a loss of semantic or stylistic information in the subtitles (16% in this case) and that this percentage is in reality even smaller, as "not all informative value of the original segments is lost" (1992, p. 168).

As mentioned earlier, Gottlieb's typology constitutes the first attempt to identify the strategies available when subtitling, and as such, its value cannot be denied. Yet it is also true that as both Jaskanen (1999, p. 11) and Lomheim (1999, p. 202) have pointed out, some of the suggested strategies are rather awkward, the distinction between them is not always clear and the examples provided are not very enlightening.

Lomheim

The second typology of subtitling strategies comes from Lomheim (1995, 1999). Lomheim's typology is based on the one suggested by Gottlieb, but has been

specially adapted. This adaptation was considered necessary, as when Lomheim attempted to classify the subtitles in his corpus according to Gottlieb's model, he encountered a number of difficulties mostly due to the large number of categories available and the somewhat fluid boundaries between them (1999, p. 202). As a result, the alternative typology Lomheim suggests consists of seven strategies only – *translation, omission, compression, expansion, generalisation, specification* and *neutralisation* – which are further reduced to six, since *omission* and *compression* can be grouped together under the heading of *reduction*, as they both result in shorter texts (1999, p. 205). As Lomheim points out, it is not always easy to distinguish between the categories both conceptually and in practice when categorising example subtitles. In addition, the strategies are very often used simultaneously and are not exclusively used at a time when formulating the subtitles (1999, p. 204).

For each strategy, Lomheim provides examples, which make the intuitive labels he uses even clearer (1999, p. 202-204). The only strategy he goes at some length to explain is the strategy of *translation*, which refers to "a linguistic transfer that meets the normal requirements of *equivalent translation*" as defined by Nida and Tabor (Lomheim, 1999, p. 200). According to this definition, translating for instance *day and night* into idiomatic Norwegian as *døgn*[24] would not be classed as *reduction* but as *translation* (ibid). Lomheim then proceeds to analyse his material on the basis of his typology, but interestingly only focuses on three of the strategies: *translation, reduction* (= *omission* and *compression*) and *expansion* (1999, p. 206). Nevertheless, the figures he presents in Table 14 are all percentages.

Table 14. Distribution of subtitling strategies according to Lomheim (1999, p. 206)

	Translation	Reduction	Expansion	
Maigret	68%	26%	6%	100%
Golden Years	50%	42%	8%	100%
Allo! Allo!	33%	61%	6%	100%

---

[24] *Døgn* is the term for an unspecified period that begins one day and ends about the same time the next day. http://www.proz.com/kudoz/norwegian_to_english/law:_taxation_customs/1904342-herav_diett:_7820_antdogn:_17.html

It is not obvious whether the remaining three strategies (*generalisation, specification* and *neutralisation*) have been completely ignored in analysing the three programmes (in which case, it would be reasonable to wonder why they were proposed in the first place) or whether they have been subsumed under any of the three categories in the table. If that is the case, there is no such indication in Lomheim's two papers (1995, 1999). The only information might come from his proposed model of subtitling strategies shown in Figure 1.

Figure 1. Lomheim's model of subtitling strategies (1999, p. 207)

Generalisation

Neutralisation

Reduction ⟵ ⟶ TRANSLATION ⟶ Expansion (form)

Specification
(content)
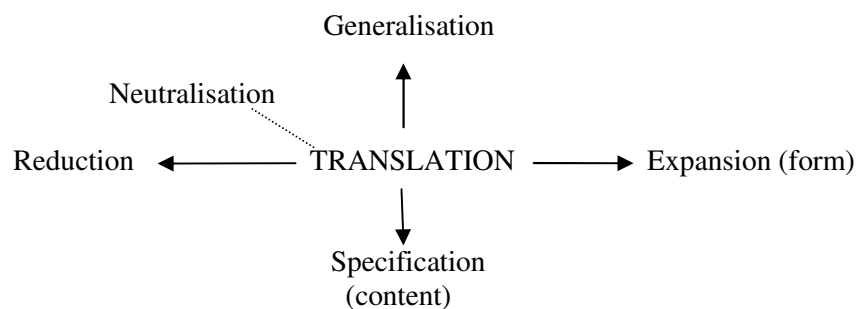
In this two-dimensional model, the horizontal axis represents form, while the vertical one represents content. *Translation* occupies the centre of the model, as equivalence is the ultimate goal of translation and is thus the point of reference. The purpose of the model is to describe the degree of departure from this point of reference, and its value is twofold, as it can be used both to describe the subtitling process and to "draw up a subtitle profile of an individual programme or subtitler" (1999, p. 209). Looking at Figure 1, it would seem as though Lomheim's analysis of his corpus presented in Table 14 has concentrated on the horizontal axis of the model. Nevertheless, this still does not explain the type of relationship between the two axes: does Lomheim's analysis suggest that it is possible to study the horizontal axis alone without taking account of the remaining strategies? Or has he actually subsumed *generalisation, specification* and *naturalisation* under the strategies of the horizontal axis? Without access to the material studied and without access to the way each example has been categorised by Lomheim it is difficult to know. Access to such information is deemed essential for any research of this type in order to elucidate the categories and metrics used by the researcher, as well as in order to

evaluate and corroborate or refute the results obtained. Such issues will be discussed in further detail in Chapter 4.

4. Research Corpus and Suggested Metrics

As mentioned in sections 2.5.4 and 3.2, very little research has been dedicated to identifying the characteristics of current subtitle output (for exceptions in the field of intralingual SDH see De Linde, 1997; De Linde and Kay, 1999; Jensema and McCann, 1995; Jensema et al., 1996). Descriptive research is, nevertheless, urgently required, as description, evaluation and the promotion of best practices go hand in hand and are indispensable in our quest for higher-quality SDH. Ideally, such research should be as large-scale as possible covering a large number of programmes, subtitles and characteristics. It should also be part of an even larger-scale research that would examine audience preferences and needs, as well as existing guidelines and would involve D/deaf and HoH people and associations, broadcasters, subtitlers, researchers and subtitling associations such as ESIST (see Chapter 1 and section 2.5.4 for further details).

After providing a brief overview of corpora used in AVT research, this chapter presents the corpus used in this research and suggests characteristics to measure and metrics to use in order to describe subtitle output. This is then followed by some preliminary results describing the characteristics of subtitles analysed from the corpus in Chapter 5. The present research is by no means large-scale, as it is restricted to two genres and a limited number of characteristics and subtitles. As a result, it is only intended to provide a blueprint for large-scale research of the type advocated above.

4.1 Corpora

For many years corpora have been a valuable resource in the study of language as attested by the increasing popularity of corpus linguistics both in terms of the relevant literature (see for instance Leech, 1991; McEnery and Wilson, 1996; Garside et al., 1997; Kennedy, 1998) and the proliferation and diversification of the

corpora produced, their users and uses. In translation studies too, corpus-based research has been an increasingly prevalent methodology for some time now (see for instance Baker, 1995; Kenny, 1998; Laviosa, 2002; Olohan, 2004). This is not surprising, as corpus-based research embodies an empirical perspective to translation in keeping with the principles of descriptive translation studies. Indeed, it involves "the direct observation of real-life examples, rather than ... speculations based on intuitive data or *a-priori* assumptions" (Laviosa, 2002, p. 16) and "is concerned with describing what translations and translators actually do, what translations are like" (Olohan, 2004, p. 5), what frequently and typically occurs in translation, etc., thus contributing to a better mapping of translation phenomena.

4.1.1 Definition and typology

"A corpus is a collection of texts selected and compiled according to specific criteria" (Olohan, 2004, p.1). The texts are held in electronic format so that they can be processed by various types of tools and may be either full running texts or text extracts. Different types of corpora have been used in translation research; the typology presented below although far from exhaustive is representative (for more details, see Laviosa, 2002, p. 34-38).

Monolingual, bilingual, multilingual corpora: A monolingual corpus consists of texts produced in one language, whereas a bilingual/multilingual corpus consists of texts in two or more languages.

Parallel and comparable corpora: A parallel corpus is "a corpus consisting of a set of texts in one language and their translations in another language" (Olohan, 2004, p. 24) or other languages. On the other hand, a comparable corpus consists of translations and similar original texts in the same language (monolingual corpus) or of similar original texts in two or more languages (bilingual or multilingual corpus respectively) (ibid, p. 35).

Full-text and sample corpora: "A full-text corpus contains unabridged texts, while a sample corpus is made up of portions of texts selected according to stated design principles concerning the size, the location of the sample within the full text and the method of selection" (Laviosa, 2002, p. 34).

Written, spoken and mixed corpora: A written corpus contains written texts. A spoken corpus, on the other hand, contains recorded spoken texts including, those written to be spoken (Laviosa, 2002, p. 35).

Synchronic and diachronic corpora: A synchronic corpus is made up of texts from a particular period of time, while a diachronic corpus consists of texts produced over a longer period.

4.1.2 Corpora in AVT

As mentioned above, corpora are considered powerful tools in better mapping the world of translation. This aspect of corpus-based research in describing the object of study has also been stressed within the field of AVT (see Delabastita, 1989; Díaz Cintas, 2004). In the words of Díaz Cintas:

> a systematic and detailed mapping of what really happens in the world of subtitling is *sine qua non* to be able to investigate both the norms that regulate the profession as well as those which underline the translator's behaviour. It would be very productive to carry out work, in research groups, that analyses sufficiently broad corpora and allows the derivation of substantial conclusions. (2004, p. 63)

Despite such calls, corpus-based research within AVT seems to be far from broad. Focusing on subtitling, most of the corpora that have been used so far have tended to include very few programmes (from one to twelve at best and only exceptionally) and quite often research has been restricted to the analysis of a particular film/programme and its subtitles. In addition, there has been a tendency to focus on

films and to a lesser extent related genres, such as soap operas and sitcoms at the expense of other types of programmes (for more see section 4.2.1). Obviously, such small corpora only constitute the first step in our effort to map the world of subtitling. Whatever their findings and contribution to research, they can only provide partial answers to our questions. This is even true for deceptively straightforward issues, such as subtitle rates and the level of editing. For instance, are current intralingual SDH subtitles verbatim? Are they edited? If so, how much editing is done? Even though verbatim subtitling vs. editing is a widely-debated topic in the context of SDH, there has been hardly any relevant large-scale research describing current subtitles in that respect (cf. De Linde, 1997; De Linde and Kay, 1999; Jensema and McCann, 1995; Jensema et al., 1996 discussed in sections 3.3.1 and 3.3.5). Undoubtedly, the research called for by Díaz Cintas is long overdue.

4.1.3 Adapting corpus typology for AVT purposes

The partial typology presented in section 4.1.1 includes the main types of corpora that have been traditionally used in translation studies. At first sight, it would seem that this typology may be transferred over as such in order to describe the types of corpora available in AVT. However, there is one problem: neither of the categories identified by the distinction between parallel and comparable corpora covers the concept of a typical intralingual SDH corpus, i.e. a corpus consisting of texts in one language and their subtitles in the same language (as is the case with the present research material). Apparently, such a corpus is not comparable; even though intralingual subtitles are not translations in the general sense of the term, they are still intralingual translations in Jakobson's sense (see section 2.2) and their relationship to the original dialogue is one of derivation rather than independence (there would be no need for subtitles were it not for the original dialogue, and subtitles need to mirror the original dialogue at least to a certain extent). Similarly, an intralingual SDH corpus is not a parallel corpus either, as according to the definition provided in section 4.1.1, a parallel corpus involves texts in different languages and is thus by definition either bilingual or multilingual. Nevertheless, there is nothing inherently bilingual or multilingual about parallel corpora. It would therefore seem reasonable

to introduce a further classification and distinguish parallel corpora into bilingual/multilingual and monolingual. Intralingual SDH corpora would then be conveniently accommodated under the latter heading.

In addition, it is important to note that while translation corpora are exclusively made up of written texts, subtitling and AVT corpora in general have the potential of being multimedia, i.e. of comprising not just a transcript of the original dialogue and subtitles, but also the actual programme in both its original and subtitled version. Only one such corpus seems to be available, though. The particular corpus, Forlixt 1, the Forlì Corpus of Screen Translation, comes from the field of dubbing and has been developed to promote the collection and study of data pertaining to film translation, as well as in order to be used as a pedagogical tool in the training of audiovisual translators (Valentini, 2006; Valentini, 2008; Heiss and Soffritti, 2008). It is composed of 11 Italian and 9 German films in both their original and dubbed versions, including the complete transcription of the dialogues of each film. The corpus has been aligned so that it is possible to extract concordances in the form of transcripts of film dialogues along with the original audiovisual scenes and has also been annotated according to a pre-established set of linguistic-pragmatic criteria, which includes categories such as *linguistic specificities*, *names of specific entities* and *communicative situations* (Valentini, 2006, no page). Given the polysemiotic nature of audiovisual products (see section 2.4.4), creating more corpora along the same lines as Forlixt 1 would be an invaluable asset to current AVT research.

4.2 Research corpus

The material used in this research comes from a small parallel corpus of six BBC TV documentaries and "newsy" current affairs programmes broadcast in 1998, 1999 and 2001. The corpus was provided by the MUSA project (see section 1.1) and consists of programmes' transcripts in English and their respective intralingual SDH subtitles along with time-in and time-out codes. Unfortunately, the material provided does not include the video for any of the programmes, although this is available in the original

MUSA corpus.[25] This undoubtedly limits the usefulness of the present research, as it made it impossible to include any metrics relating to speech (e.g. speech rates and synchronisation with subtitles), as well as to consult the information available on screen when analysing characteristics, such as subtitle display times, deletion and segmentation (see sections 5.1, 5.2, 5.3, 5.5, 5.6 and 6.1 for further details).

Of the six programmes provided by MUSA, only three were analysed, two documentaries and one current affairs programme as shown in Table 15. This is for two reasons. First of all, the main reason for using the corpus was to suggest characteristics to measure and metrics to use when describing subtitles and demonstrate their usefulness; neither the corpus nor the research were meant to be or could be large-scale in the sense advocated in this thesis and described in Chapter 1 and section 2.5.4. Second, both the annotation/mark-up and analysis of the corpus were performed for the most part manually. Moreover, in order to cover more programmes, it was decided to analyse a selection of subtitles from the three programmes rather than analyse one documentary in full. Thus, from each programme a selection was made of the first, middle and last 40 subtitles.

Table 15 includes all information available regarding the three programmes. Unfortunately, there is no information about the director, while the total length provided for each programme (shown in minutes : seconds : frames) was calculated on the basis of the available time-codes and is thus only an approximation.

---

[25] The BBC Horizon and Panorama websites provide information about two of the three programmes in the research corpus including transcripts (http://www.bbc.co.uk/science/horizon/archives.shtml and http://news.bbc.co.uk/1/hi/programmes/panorama/1540307.stm).Both websites also provide links containing information about how to watch the programmes. On the Horizon site, a link "Can I get a copy?" is included. This, however, leads to a page, where it is stated that Horizon programmes are not sold to the public, but tapes of many of them can be ordered through the BBC Active page for use in education and training (http://www.bbc.co.uk/science/horizon/about.shtml#videos). The Panorama website also refers interested parties to the BBC Active page for older programmes (programmes broadcast recently are available on the BBC iPlayer, while a range of programmes dating from 30/7/2007 is available on the Panorama subtitled archive http://www.bbc.co.uk/accessibility/win/hearing/alt/panorama/sub_3.shtml). Unfortunately, the BBC Active page explicitly states that BBC programmes are not sold to individuals for research or other purposes (http://www.bbcactive.com/broadcastlearning/asp/home/home.asp).

Table 15. Research corpus

Name	Programme/ Type	First aired	Length	Total subtitles	Subtitles analysed
Atlantis Uncovered (A)	Horizon Documentary	28/10/1999	43:59:06	399	120
Chimps on Death Row (B)	Horizon Documentary	1/10/1998	43:55:20	419	120
The World's Most Wanted (C)	Panorama Current affairs	16/9/2001	27:04:12	360	120

In addition, as shown in Table 15, the two documentaries come from BBC Horizon, BBC Two's award-winning science documentary series that has been on air since 1964, while *The World's Most Wanted* comes from BBC Panorama, the world's longest-running current affairs programme broadcast on BBC One since 1953.[26] *Atlantis Uncovered*, which was first broadcast on October 28th 1999, is the first part of a Horizon special that explores the myth of Atlantis and investigates whether this popular, yet controversial idea could be true. *Chimps on Death Row*, first broadcast on October 1st 1998, looks at the use of chimpanzees in scientific research and the responsibility society bears towards experimental animals. Finally, in *The World's Most Wanted*, which was first aired on September 16th 2001, reporter Jane Corbin investigates how the 9/11 attacks on New York and Washington were carried out and examines the evidence that Osama Bin Laden was behind them.

4.2.1 Genre

As mentioned above, the corpus analysed for the present research consists of documentaries and current affairs programmes. Both genres belong to the informative text type as defined by Newmark (1988) and adapted for screen translation purposes by Georgakopoulou (2003, p. 101) as shown in Table 16.

---

[26] Further information is available on the two programme's websites
http://www.bbc.co.uk/science/horizon/about.shtml and
http://news.bbc.co.uk/panorama/hi/about_panorama/default.stm

Table 16. Newmark's informative function adapted to screen translation by Georgakopoulou (2003, p. 101)

Function	Informative
Core	Truth
Author's status	Anonymous
Type	Areas of knowledge or events
Examples from screen translation	Documentaries, news reports, educational-purpose programmes

Both genres consist to a great extent of text that has been written to be read by a narrator (always off-screen for documentaries) or to be spoken by a journalist (in front of the camera or sometimes off-screen for current affairs programmes). Being written to be read, such text might be expected to be characterised by a high rate of lexical density and a low rate of grammatical intricacy (see Halliday, 1989). Furthermore, speech rates seem more likely to be constant and average to low rather than high. In addition to such pre-prepared text, both documentaries and current affairs programmes also contain parts belonging to other genres, such as interviews. These parts are oral, many of them spontaneous and exhibit high grammatical intricacy, lower lexical density, higher and varying speech rates and a high presence of oral features (Eugeni, forthcoming, b).[27]

The decision to focus on documentaries and to a lesser extent current affairs programmes was taken for the following reasons:

a) Factual programmes, in particular documentaries, are consistently among the favourite types of programmes viewed by D/deaf people and among the programmes they believe should be subtitled irrespective of viewing preference (see Kyle, 1992; RNID, 1999; NSM Research, 2001).

b) Like many other types of programmes, documentaries and current affairs programmes have been considered inferior to films and have been largely neglected

---

[27] Corpus-based research is required to investigate whether these expectations, in particular those regarding speech rates, are actually valid. However interesting, such investigation was outside the scope of this thesis mainly due to the lack of access to the video of each programme.

by subtitling research (Díaz Cintas, 2004, p. 67). Díaz Cintas's argument is quite persuasive:

> Most of the films analysed tend to be rather similar. They are highly considered socially, are aimed at a particular audience and have been directed by acclaimed filmmakers. This attitude can be blamed to some extent on insecurity, or an inferiority complex, on the part of the researchers themselves. We believe that our area of study is marginal and we try to rectify this by resorting to prestigious, highly respected cineastes and programmes. We are only managing to map out a very partial and limited reality of subtitling that cannot hope to be representative of the world of subtitling. (ibid)

The tendency of subtitling research to focus on fiction films has also been interpreted as a reflection of "the prevailing orientation in translation theory, which is still strongly dominated by literary translation" (Gambier, 1994, p. 275). Whatever the predominant reason for this state of affairs, it is hoped that the research undertaken for this thesis might contribute to the effort to redress the balance.

c) Last but not least, a major constraint that influenced the decision to focus on documentaries and current affairs programmes had to do with the nature of the material that was available. Unfortunately, there are not many subtitling corpora readily available for research purposes. Building a corpus from scratch was an alternative considered, especially so because it would also ensure access to the video of each programme. However, in order to construct such a corpus, it would be necessary to either record programmes from TV or buy programmes on DVD. The first scenario would involve transcribing the audio, copying subtitles from screen and recording time-in and time-out codes in a rather crude and approximate manner, a tedious and time-consuming process overall. In the second scenario, it would be possible to rip the subtitles and their time-in and time-out codes, yet it would still be necessary to transcribe the

audio.[28] In addition, ripping the DVD and subtitles even for research purposes would very likely constitute copyright infringement in many countries, including the UK[29]. Copyright is a very complex issue, yet seems to have been hardly addressed in subtitling literature.

4.2.2 Limitations and objectives

Undoubtedly, the present research is far from large-scale. It is restricted to two genres, three programmes, 120 subtitles from each programme and a limited number of characteristics. In addition, the research corpus only contains programme transcripts and subtitles and not the actual video. As a result, any findings and suggestions presented in this thesis are only indicative and will need to be corroborated by large-scale research covering more genres, programmes, subtitles and characteristics.

Although such limitations are admittedly significant, the present research was not meant to be large-scale in the first place. Nor could it be, as it seems that truly large-scale research can only be conducted by research groups (rather than individual researchers) in close co-operation with broadcasters and possibly subtitlers themselves. For practical reasons then, the research reported here is instead intended to provide a blueprint for large-scale corpus-based research of the type advocated above by suggesting characteristics to measure and metrics to use when describing subtitle output. In addition, it offers some preliminary results, compares findings with previous research and tentatively relates them to existing subtitling guidelines.

---

[28] We need to note, however, that even transcribing the audio might not be necessary, given that many programme/film transcripts are readily available on-line.
[29] The UK Copyright Service mentions that according to the Copyright, Designs and Patents Act 1988, the principal copyright legislation in the UK, copyrighted works can be used for research purposes to a certain extent without copyright being infringed (this policy is known as fair dealing) (http://www.copyrightservice.co.uk/copyright/p01_uk_copyright_law). However, fair dealing only covers literary, dramatic, musical or artistic works and does not extend to films, broadcasts and cable programmes (http://www.opsi.gov.uk/acts/acts1988/ukpga_19880048_en_3#pt1-ch3-pb2-l1g29).
 (Both pages last accessed on October 28th 2008)

As mentioned in section 4.1, corpus-based research is par excellence part of the descriptive approach to translation. Using a corpus, albeit a small one, has accordingly allowed us to identify to a certain extent some of the characteristics of the programmes included in the present material. Description is not however the only valuable outcome of corpus-based research, as corpora can also be used for evaluation and prescriptive purposes. This is particularly true for large-scale corpora describing current subtitle output. Building such corpora and comparing findings to both a) existing guidelines and b) viewers' needs and preferences as recorded in large-scale studies would enable us to better evaluate current subtitles and guidelines. Evaluation would then ideally help promote best practices, point out insufficient solutions and suggest better alternatives and overall improve current subtitle output and recommendations. In such a context, there is little doubt that description, evaluation and prescriptive guidelines would all be powerful tools inextricably linked in our quest for higher-quality subtitling.

4.3 Corpus analysis

As mentioned in section 4.2, each programme provided by the MUSA project consists of a transcript file and a separate subtitle file. These files formed the basis for analysing the three programmes included in the present research corpus. The analysis proceeded in three steps:

(1) The original files were used to note a variety of features (see section 4.3.1)

(2) For each programme, the transcript file and an updated subtitle file without time codes were aligned/merged using MS Word's "Compare Documents" facility (see section 4.3.2)

(3) The resulting merged files were manually marked up/annotated (see section 4.3.3).

For steps (1) and (3), it was first essential to decide on and define: a) the features to record and b) the categories/metrics to use. These are specified in the sections that

follow. Suffice it here to say that this was an important part of the research for two reasons. First of all, defining the categories and metrics used as clearly as possible is an integral part of explicitly reporting one's research so that it may be replicated and findings compared with those of other relevant projects. Moreover, clearly defining the categories and metrics used allows other researchers to adopt and/or adapt them thus promoting the establishment of a set of widely-accepted categories and metrics. For instance, subtitle rates have so far been calculated using two different methods, that is either by dividing the number of words in each subtitle by the subtitle's presentation time (e.g. De Linde and Kay, 1999) or by dividing the total number of subtitles with the total duration of the programme (e.g. Georgakopoulou, 2003). For comparison and accuracy reasons, promoting the use of a single method seems preferable (see also section 4.3.1.5).

4.3.1 Features recorded

For each programme the following data was recorded:

- Strategy/strategies used
- Time-in and time-out codes
- Duration of each subtitle
- Characters per subtitle
- Subtitle rate
- Whether time constraints are respected
- Time interval between consecutive subtitles
- Words per subtitle/ Words per respective speech segment
- Number of words deleted per subtitle
- Number of lines per subtitle
- Line-break errors

Each will be discussed in detail in the remainder of this section.

4.3.1.1 Strategies

As mentioned in section 3.3.6, very little research has been dedicated to identifying the range of strategies available when subtitling. Instead, attention has focused on a limited number of issues, namely reduction and in particular deletion (and related aspects), and the issue of verbatim subtitling vs. editing in the case of SDH. Drawing from the typologies suggested by Delabastita (1989) and Gottlieb (1992), and in particular Lomheim (1995, 1999), Table 17 includes the basic strategies available when producing intralingual subtitles for D/deaf and HoH viewers.

Table 17. Subtitling strategies (editing strategies in shading)

	Same words		Different words – Substitution
All words	Verbatim (same order)	Transposition (different order)	
More words	Addition/Expansion		
Fewer words – Reduction	Deletion		Condensation
Amount of words not relevant			Paraphrase
			Specification
			Generalisation

With the exception of verbatim subtitling, all other strategies are editing strategies, as they involve some kind of change to the original version.

Verbatim subtitling: All the words of the original are used in the subtitle in exactly the same word order. For SDH, it seems that this is the main strategy used, the default strategy. A subtitler will revert to other strategies only if they have any reason to believe that verbatim transcription will be inappropriate in a particular instance.

Transposition: All the words of the original are used, but there is a change in word order. This strategy is equivalent to Delabastita's *transmutatio* (1989).

Addition/expansion: All (or some of) the words of the original are used, while new words are also added to the subtitle.

Deletion: Only some of the words of the original dialogue are used, as certain words are omitted.

Condensation: The original or part of the original is replaced by a shorter term, e.g. *perhaps they'll find* becomes *they may find*. As mentioned in section 2.4.4, condensation and deletion are often grouped together under the heading of *reduction*.

Paraphrase: This is an umbrella term, as it seems to encompass all instances not covered by other substitution strategies. That is, a paraphrase is a substitution where the new term is neither shorter nor more specific or more general than the original. More often than not, paraphrase involves simplification, either syntactic or lexical. In syntactic simplification, a syntactic structure considered difficult is replaced by another considered easier to understand (e.g. a relative clause might become a main clause, or a passive construction may become active), whereas in lexical simplification a word or phrase considered difficult is replaced by a simpler word/phrase.

Specification: The original or part of the original is replaced by a more specific term, e.g. *they* becomes *these peaks*.

Generalisation: This is the opposite strategy to specification. The original or part of the original is replaced by a more general term, e.g. *Mavrodafni* becomes *wine*.

As can be seen in Table 17, strategies have been grouped in two main categories. Thus, strategies under the heading "Same words" result in subtitles that basically include the same words as the original (all or some of them), while strategies under the heading "Substitution" result in subtitles where the original or part of the original is replaced by another word, phrase or sentence. In addition, strategies have also been grouped according to their relation to the number of words in the original. Thus,

verbatim subtitles and subtitles with transposition include the same number of words, subtitles with reduction or condensation include fewer words, and subtitles with addition include more words. For the rest of the strategies, the number of words of the original seems to be of secondary importance, as applying each of them may result in a subtitle with the same number of words, fewer or more words than the original.

Furthermore, as might be expected, with the exception of verbatim subtitling, strategies are quite often combined within the same subtitle. To a certain extent, this also means that for instance although a subtitle that only exhibits addition will by default be longer than the respective speech segment, a subtitle that exhibits addition along with other strategies might be longer, shorter or equal in length to the original utterance.

4.3.1.2. Time-in and Time-out codes

Time-in and time-out codes indicate the exact time when each subtitle should appear and disappear from screen. The time codes used in the programmes analysed are of the standard form used in the industry. Thus, the first subtitle in programme A appears at 10:0:17:8.[30] The numbers indicate "hours : minutes : seconds : frames". Both time-in and time-out codes are among the features recorded in the original MUSA project and are included in the subtitle file of each programme before the beginning of each subtitle.

4.3.1.3 Duration of each subtitle

The duration of each subtitle was calculated on the basis of the time-in and time-out codes already available and is expressed in "seconds : frames". Calculation can be easily made either manually or automatically by inputting the subtitles and their

---

[30] The only difference is that in the time codes used in the MUSA corpus numbers 0 to 9 are not preceded by 0. Thus, the time code is 10:0:17:8 rather than 10:00:17:08.

respective time-in and time-out codes into subtitling software, such as SysMedia's WinCAPS (see Figure 2).

4.3.1.4 Characters per subtitle

For each subtitle, the number of characters was recorded. This was initially done manually, but at a later stage a tailor-made version of WinCAPS was used. [31] With WinCAPS, it suffices to click on any subtitle line and the software instantly displays the number of characters for that line. In the example shown in Figure 2, the second line of subtitle 01 has 32 characters (shown in green colour).

Figure 2. WinCAPS screenshot

---
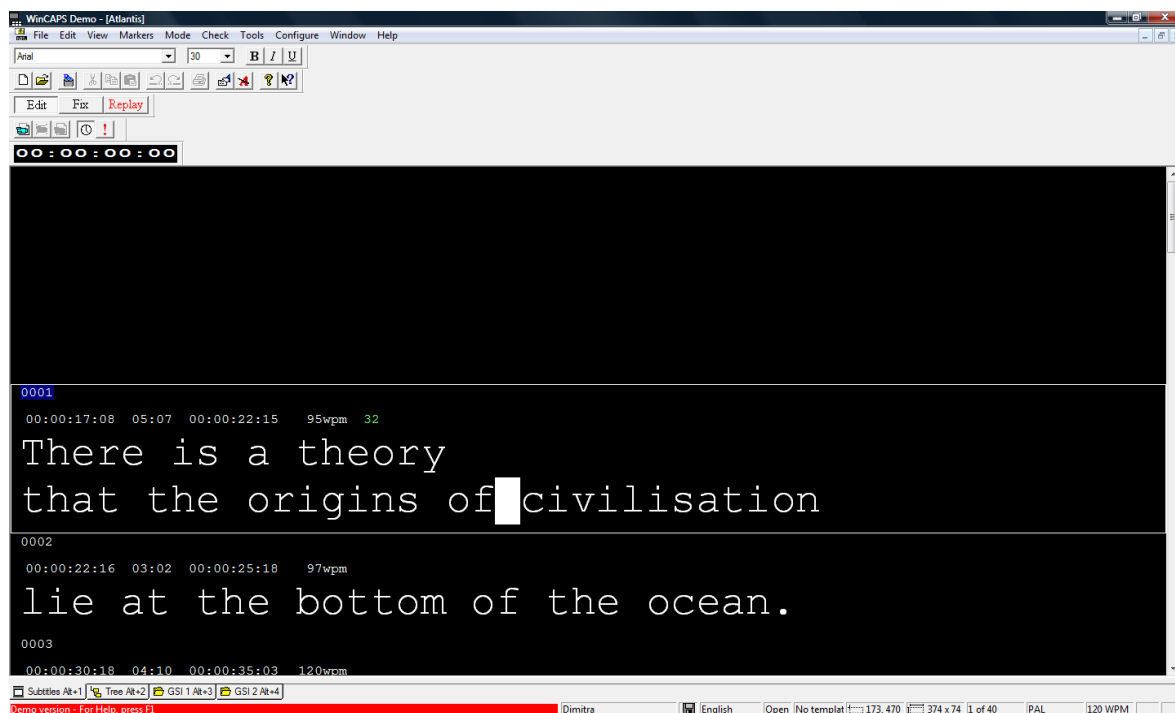
[31] The demo version of WinCAPS used for the analysis of my corpus came with Diaz Cintas and Remaels' book *Audiovisual Translation: Subtitling* (2007). Additional information can be found at www.sysmedia.com

4.3.1.5 Subtitle rate

For each programme, the subtitle rate was calculated by dividing the number of words in each subtitle by the subtitle's presentation time:

$$x = \frac{\text{number of words per subtitle}}{\text{subtitle's presentation time}}$$

Equation 1

This metric was used following de Linde and Kay (1995, p. 45) in the belief that as the two researchers point out, it is more appropriate than measuring the total number of subtitles per minute of film, as has sometimes been done, since the latter method "excludes the time when no verbal information (oral or written) is transmitted" (ibid).

When access to WinCAPS became available at a later stage, the software was used to confirm the results obtained on the basis of Equation 1. This was rather straightforward, as the software automatically calculates and displays the subtitle rate for each subtitle in wpm (see Figure 2) and all that was required was to calculate the mean rate for the entire programme or sample analysed. Although at first results from the two methods seemed to differ only marginally with figures calculated on the basis of WinCAPS slightly higher than those calculated on the basis of Equation 1, when the process was completed, it became apparent that differences were in fact quite substantial (see sections 5.1 and 5.9). Despite not knowing the calculation formula used by WinCAPS and even though Equation 1 had seemed to be a reasonable metric, the relevant data provided in Chapter 5 was calculated on the basis of the subtitle rates provided by WinCAPS.[32] The main reason behind this decision was that this would be what a professional subtitler would be expected to

---

[32] The only exception to this was when comparing figures with the relevant findings reported by De Linde and Kay (1999), see section 5.1.

do, i.e. to calculate subtitle rates using the in-built facility provided by the subtitling software they are using.[33]

4.3.1.6 Whether time constraints are respected

This variable was only recorded on subtitle (rather than programme) level and had two values: "respecting" and "not respecting". In order to provide an answer, it was necessary to first decide on a maximum presentation rate. In accordance with the literature and research so far conducted, this was set to 120 wpm. However, due to the lack of large-scale empirical research and since the trend nowadays is for faster subtitles, it was also decided to set the maximum higher at 130 wpm and 140 wpm.

It should be pointed out that this seems to be the least objective of the metrics used for the purposes of this project. For one thing, there has been no large-scale research into identifying the maximum subtitle rate for D/deaf and HoH viewers. More importantly, even if it were possible to establish such a maximum rate, there are so many factors influencing the appropriate subtitle rate for each programme and subtitle (see sections 2.4.3 and 3.3.1) that any decision should be made on an ad-hoc basis thus making any general analysis tentative. Furthermore, the present analysis is even more tentative given that it is not possible to take into account factors such as the speech rate, the amount of information on screen and camera cuts, since the corpus material does not include the video for each programme.

4.3.1.7 Time interval between consecutive subtitles

This was calculated on the basis of the time-in and time-out codes in the subtitle file provided for each programme. Like subtitle duration, this was expressed in "seconds : frames".

---

[33] Nevertheless, it needs to be added that not all subtitling is done using dedicated subtitling software. In such cases, subtitlers need to calculate themselves how long their translation will be often on the basis of a table with the subtitling company's desired subtitle rate (information provided by Fotios Karamitroglou, personal correspondence, 25/09/2008).

4.3.1.8 Words per subtitle/Words per respective speech segment

These were measured manually from the transcript and subtitle files provided by MUSA. The figures for subtitles did not include words identifying the speaker, e.g. *NARRATOR (DILLY BARLOW)*. In addition, compound words connected with a hyphen and expressions such as *we've* were counted as one rather than two words.

4.3.1.9 Number of words deleted per subtitle

On the basis of the figures provided by 4.3.1.8, it was possible to establish the number of words deleted per subtitle and programme. It is important to note that since the focus of this study is intralingual subtitling, this is more of an objective measure than it is the case of interlingual subtitling, given the difference in lexical structure between different languages (Lomheim, 1999, p. 191, see also section 5.6).

4.3.1.10 Nature of deletions

All deletions identified in 4.3.1.9 were subsequently categorised according to type. Initially, the intention was to use the typology suggested by Kovačič (1992, 1994, 1998) whereby deleted linguistic elements are categorised according to their function, i.e. ideational, interpersonal or textual (see also sections 2.4.6 and 3.3.5). However, in attempting to classify the deletions in the present sample according to these categories, it soon became evident that this was not as easy and implementing the typology was thus abandoned. There were three main reasons for this:

1) As mentioned in section 2.4.6, the linguistic terms involved (*ideational*, *interpersonal* and *textual*) are not intuitively clear and do not seem to have been defined in the subtitling literature.

2) One of the main reasons for undertaking this research was not merely description for its own sake, but description as a means of enabling subtitlers to better know their

trade and to provide higher-quality subtitling. Using terms which are not intuitively clear and remain rather vague to subtitlers is not in line with such goals.

3) With the exception of Kovačič and other researchers referring to her work, of the three linguistic terms involved, only the term "interpersonal" seems to have been used more widely (see for instance De Linde and Kay, 1999; Díaz Cintas and Remael, 2007). More often than not, both researchers and subtitlers have opted for different terms. For instance, De Linde (1995, p. 16) has used the categories discussed in section 3.3.5 and partly repeated here for ease of reference:

- Markers of interaction
- Gestural language
- Repetitions
- Additional comment
- Additional information
- Whole dialogue
- Proper names

Somewhat overlapping categories have been used by the European Captioning Institute (ECI) to define easily omitted items when subtitling:

- Repetitions
- Names in appellative constructions
- False starts / ungrammatical constructions
- Certain internationally known words, such as *yes* and *no*
- Expressions followed by gestural language to denote salutation, politeness, affirmation, negation, surprise, telephone responses, etc.
- Exclamations
- Phatic phrases, such as *naturally*, *of course*, *understandably*, repeat implied superlatives (*basically*, *fundamentally*), prepositional phrases (*in view of the fact that*), rhetorical flourishes, sonorous phrases used for sound-effect (*might and main*, *ways and means*)

(Georgakopoulou, 2003, p. 216)

Using these categories and others suggested by Díaz Cintas and Remael (2007), Gottlieb (2005), Karamitroglou (1998), and Ivarsson and Carroll (1998)[34], the following list of easily omitted items served as the starting point for the analysis of the present corpus:

1. False starts, hesitations, repetitions, pauses, self-corrections

2. Ungrammatical constructions, slips of the tongue, self-contradictions, ambiguities, nonsense

3. Overlapping speech, background chatter

4. Interjections/exclamations: e.g. *oh, o, ah, hmm, um, eh, uh, uh-huh, dear, ouch, wow*

5. Backchannels, i.e. acknowledgements that the listener is following what is being said and encourages the speaker to continue, and clauses that are intended to elicit support from the listener:

   a. Backchannels: e.g. *yes, yeah, no, ok, exactly, good, right, of course, sure, I see, that's right, that's true, really?, you think so?, have you?*, interjections, restatements (i.e. restating part of what the speaker has said)

   b. Clauses that are intended to elicit support from the listener: e.g. *as you know, you see*

6. Vocatives

7. Greetings, formulas of courtesy: e.g. *hello, nice to meet you, thank you, please*

8. Expressions followed by gestural language to denote salutation, politeness, affirmation, negation, etc.

9. Discourse markers (fillers/padding expressions): e.g. *well, anyway, now, you know, as I say, so, like*, etc.

10. Emphatic repetition and other emphatic devices and phrases (e.g. *might and main*, clefts)

11. Hedges, e.g. *it seems to me, I think, sort of, kind of, slightly, somewhat, more or less*

12. Intensifiers, particularizer and exclusive adverbs:

   a. Intesifiers: e.g. *fairly, pretty, quite, rather, really, very*

---

[34] See detailed presentation in section 3.3.5

b. Particularizer advebs: e.g. *especially, particularly, at least, primarily*

c. Exclusive adverbs: e.g. *just, simply, only*

13. Numerals and quantifiers: e.g. *many, more, most, much, several, enough, (a) few, (a) little, a bit, a lot, a couple, some, half, every, all*

14. Conjunctions and conjuncts:

a. Conjunctions: e.g. *and, or, but, because, that*

b. Conjuncts: e.g. *first, firstly, then, similarly, also, for example, therefore, on the other hand*

15. Disjuncts:

a. Style disjuncts: e.g. *approximately, briefly, honestly, personally, in general, to be fair, to speak frankly, if I may be candid*

b. Content disjuncts: e.g. *admittedly, certainly, obviously, actually, basically, fortunately, sadly, ironically, reasonably, wisely, wrongly,* etc.[35]

16. Additional information: e.g. additional information of place, time, frequency, manner, comparison, quality (expressed by adverbs, prepositional phrases, attributive adjectives, etc.)

17. Grammatically redundant items or redundant information that is obvious or can be inferred from the co-text, context or visual channel, etc.

18. Less relevant information or additional comment, particularly when there is lots of text to transfer and not much time (could be a whole phrase or dialogue)

As can be seen, categories 1-9 include so-called oral features, so it is possible to group them all together under one large group. We need to point out that repetition features twice in the list: once under category 1 along with pauses and hesitations and again under 10 as a separate category. Initially, repetition was envisaged as a single category as defined in relevant taxonomies (e.g. De Linde, 1995; Georgakopoulou, 2003). However, on analysing the present material, it became obvious that there were two rather different types of repetition, as exemplified by the following examples:

---

[35] All grammatical terms here (such as content and style disjuncts, conjuncts, intensifiers, particularizer and exclusive adverbs) and relevant examples are from Greenbaum, 1996.

Transcript: who went out <u>to, to</u> some horrible place...

Subtitle A29: He went out to some horrible place...

Transcript: and after all, if you have <u>people finding</u> forgotten tombs in Ancient Egypt and *people finding* lost civilisations in the middle of the tropical rainforest of Mexico

Subtitle A30: And, if you have people finding

                 forgotten tombs in ancient Egypt

Subtitle A31: and lost civilisations in the middle

                 of a tropical rainforest in Mexico,

There is little doubt that the two repetitions are quite different. In subtitle A29, repetition is an instance of hesitation. In subtitles A30-31, on the other hand, it seems to serve as a rhetorical device used to create emphasis and amplification.

Initially, apart from using categories 1 to 18, the intention was to classify all deletions in the present corpus using Eugeni's distinction between semantic (qualitative) and non-semantic (quantitative) omission described in section 3.3.6. This idea was, nevertheless, abandoned due to the time limitations applying to this thesis, as a draft of Eugeni's paper (forthcoming, b) only became available to the present author in July 2008. In addition, according to Eugeni, semantic omission entails the deletion of meaningful lexemes or phrases, while non-semantic omission involves the deletion of grammatical words, features of speech and lexemes not useful to the semantic meaning of the macro-idea unit. Determining, however, what is and what is not useful to the semantic meaning of an utterance is not always straightforward nor always determined in the same way by different people. Moreover, although oral features might not contribute to the semantic meaning of an utterance, as mentioned in section 2.4.6, they might still be integral to a character's style of discourse and/or the portrayal of their character (De Linde, 1995, p. 12; Kovačič, 1998b, p. 127). Such reservations do not invalidate the usefulness of Eugeni's distinction, yet need to be taken into account when using it to classify

deletions. Had more time been available, it would have been highly informative to apply the distinction while mapping it against categories 1 to 18 suggested above.

4.3.1.11 Number of lines per subtitle

The number of lines was counted and recorded for each subtitle on the basis of the subtitle file provided.

4.3.1.12 Line-break errors

All errors in segmentation, that is all line breaks considered poor in terms of linguistics, were recorded for each subtitle analysed. Focus was on segmentation errors within rather than across subtitles. Although this seemed to be an objective metric, there is some evidence suggesting that this might not actually be the case (see section 6.2).

4.3.2. Alignment/Merge

The corpus was aligned and merged using MS Word 2007 and the "Compare files" function. This function automatically detects and marks insertions and deletions, as well as other changes between an original document and a revised version. In this case, the transcript file was set as the original document and the subtitle file as the revised version; it was first necessary to delete all time-codes from the subtitle file and create a new file containing only the subtitles selected for analysis, as well as a new transcript file containing the corresponding speech segments (the new transcript files and subtitle files including time codes for each programme are provided in Appendices 1-3). Deletions were marked by strikethrough, while additions by underline. Since the transcript file was set as the original document, the resulting merged file had to be divided into subtitles. In addition, all alignments and merges had to be manually checked, as it soon became evident that they were not always

accurate. For instance, in programme A, part of subtitle A13 was aligned and merged as follows:

Transcript: I knew that it needed to be investigated ...

Subtitle A13: I knew Nakbe had to be investigated ...

Merge: I knew <u>Nakbe had</u> ~~that it needed~~ to be investigated ...

whereas a more appropriate version might be:

Alternative: I knew ~~that~~ <u>Nakbe</u> ~~it~~ <u>had</u> ~~needed~~ to be investigated ...

Similar problems were encountered across subtitles. For example, in the same programme, part of subtitle A5 was aligned and merged with subtitle A6:

Transcript: around 12,000 years ago, but a few survivors ...

Subtitle A5:   around 12,000 years ago.

Subtitle A6:   But some survivors ...

Merge: around 12,000 years <u>ago. But some</u> ~~ago, but a few~~ survivors...

A more accurate merge would be as follows:

Alternative: around 12,000 years ago<u>.</u> ~~,~~ <u>But</u> ~~but~~ <u>some</u> ~~a few~~ survivors...

Moreover, it also became evident that the original transcript and subtitle files were not accurate themselves. The transcript file, for instance, contained errors, such as:

Transcript: Could it possible [sic] *be true?* (corresponding subtitle A6)

Without access to the original video, it cannot be confirmed whether this is an error in the original audio or the transcription itself (although the latter seems more likely, as this phrase is spoken by the narrator).

In addition, it was soon evident that there were differences between the original transcript and subtitle files. For instance, in programme A, the time-out code for subtitle A96 is 10:41:35:23 and the time-in code for subtitle A97 is 10:41:35:24. The transcript file, however, includes an extra six lines between the speech segments corresponding to the two subtitles. Without recourse to the original video, it is difficult to tell whether the problem lies with the transcript or the subtitle file. Contrary to this, in the same programme, the time-out code for subtitle A111 is 10:43:15:8, while the time-in code for subtitle A112 is 10:43:19:17. In this case, the problem seems to be with the subtitle file, as not only is this gap rather long, but more importantly the transcript file again includes an extra six lines between the two subtitles. In both cases, it is not clear whether the discrepancies come from the original files provided by the BBC or whether they were inadvertently introduced in the process of creating the MUSA corpus. On the other hand, they might only be present in the copies provided for this research.

4.3.3 Mark-up and Annotation

After the merged files had been divided into subtitles and all alignments/merges had been checked, the text was manually annotated (the annotated merged files for the three programmes are provided in Appendices 1-3). Tags were provided for the following:

- Subtitle number
- Line breaks
- Line-break errors (segmentation)
- Part-of-speech (POS) tags (partially only)
- Strategy/strategies used
- Duration of each subtitle
- Whether time constraints are respected
- 3-line subtitles

Some of the tags are standard HTML tags, such as *
*, while others were modelled on HTML tags, such as *_{* **subtitle text** *}*. POS tags were modelled on standard SGML tags of the form *<w POS>* where *w* stands for *word* (see Leech, 1997, p. 30); in the present corpus, these tags follow rather than precede the word to which they apply. The same form was also used as a model for tags, such as *<duration min:frames>*, e.g. *<duration 7:20>*.

Subtitle number: All subtitles analysed have been assigned a number tag to make identification easier. Thus, subtitle 1 has been marked-up using an HTML-style tag as follows:

<sub1>
There is a theory
that the origins of civilisation
</sub1>

Line breaks: All subtitles in the merged text have been divided according to the original line breaks. Line breaks are only indicated with a *
* tag for clarity reasons when a subtitle line becomes too long due to the incorporation of the transcript text and tagging, e.g.

<sub11>
which set them apart from <DELETION> ~~monkeys~~ </DELETION> <DELETION> ~~and~~ </DELETION> the rest
of the animal kingdom.
</sub11><strat DELETION, DELETION><duration 4.24><TC respecting>

Line-break errors and POS tags: All errors in segmentation within the same subtitle have been marked-up using an empty tag similar to the <br> tag: *<LINEBREAKERR>*. In addition, all relevant words have been POS tagged (the tagset used is available in Appendix 4).

<sub6>
But ~~but~~ <u>some</u> ~~a few~~ survivors escaped in ships
and brought civilisation to the <w DET> <LINEBREAKERR>
primitive <w AJ> peoples <w NN> around the world.
</sub6>

<u>Strategy/strategies used</u>: For every editing strategy used, both subtitle-internal and subtitle-external tags have been assigned:

<sub115>
to give up the fugitive <DELETION> ~~terrorist~~ </DELETION>
or face <CONDENSATION> <u>America's</u> ~~the~~ wrath ~~of America~~ </CONDENSATION>.
</sub115><strat DELETION, CONDENSATION>

Quite often, the same strategy is used twice within the same subtitle and is accordingly tagged twice:

<sub46>
<DELETION> ~~I guess~~ </DELETION> <DELETION> ~~at that point in time~~ </DELETION> <CONDENSATION> <u>Euthanasia was considered</u> ~~there was consideration of euthanasia~~ </CONDENSATION>
as a population control method.
</sub46><strat DELETION, DELETION, CONDENSATION>

If no editing strategy has been used, the subtitle only receives a strategy-related tag externally:

<sub14>
They show characteristics like
altruism, compassion and so forth.
</sub14><strat VERBATIM>

<u>Duration of each subtitle/Whether time constraints are respected</u>: Both the duration of each subtitle and whether it respects time constraints have been marked-up externally. As pointed out in section 4.3.1.6, the question of whether time constraints

are respected is not easy to answer, as it depends on a wide variety of factors. For this part of the analysis, the maximum display rate was set to 120 wpm.

<sub6>
Within hours,
suspicion fell on one man.
</sub6><strat VERBATIM><duration 3.18><TC respecting>

3-line subtitles: All subtitles consisting of three lines have been assigned an external tag, *<3-liner>*:

<sub2>
Now <CONDENSATION> ~~they're forcing~~ scientists <u>are facing</u> ~~to face~~
</CONDENSATION>
the consequences of experimenting
on our closest living relatives.
</sub2><strat CONDENSATION><duration 7.03><TC not respecting><3-liner>

4.4. Conclusion

In this chapter, we have provided a brief overview of corpora in general and in translation studies in particular. Corpus-based research has been an integral part of translation studies for some time now and its application to AVT can only benefit the field by helping us provide a detailed mapping of what happens in subtitling and SDH in particular and more importantly helping us implement the tripartite research plan advocated in this thesis. In addition, we have also presented the corpus used in this research including limitations, such as the lack of access to the video of the programmes analysed. Finally, we have suggested some basic characteristics to measure and categories/metrics to use in order to describe current subtitle output (see Table 18). These were used to analyse the present material and relevant preliminary results will be presented and discussed in Chapters 5 and 6.

Table 18. Overview of features, categories and metrics used

Feature	Metric/Categories used
1. Strategies used	- Verbatim subtitling - Transposition - Deletion - Condensation - Addition - Paraphrase - Specification - Generalisation
2. Duration of each subtitle	x = Time-out code – time-in code Seconds : frames
3. Subtitle rate	- WinCAPS - $x = \dfrac{number\ of\ words\ per\ subtitle}{subtitle's\ presentation\ time}$ Words per minute (wpm)
4. Whether time constraints are respected	- Yes - No Baseline maximum rate 120/130/140 wpm
5. Time intervals between subtitles	x = Time-in code of subtitle z – time-out code of subtitle y Seconds : frames
6. Number of words deleted per subtitle	Words per speech – words per subtitle (compound words connected with a hyphen and expressions such as *we've* count as one word)
7. Nature of deletions	1) False starts, hesitations, repetitions 2) Ungrammatical constructions 3) Overlapping speech, background chatter 4) Interjections 5) Backchannels and clauses intended to elicit support from the listener 6) Vocatives 7) Greetings, formulas of courtesy 8) Expressions followed by gestural language to denote affirmation/negation, etc. 9) Discourse markers (padding expressions) 10) Emphatic repetition and other emphatic devices 11) Hedges 12) Intensifiers, particularizer and exclusive adverbs 13) Quantifiers and numerals 14) Conjunctions and conjuncts 15) Disjuncts 16) Additional information (e.g. of place, time, as expressed by adverbs, prepositional phrases, etc.) 17) Grammatically redundant items or redundant information 18) Less relevant information or additional comment
8. Number of lines per subtitle	-One-/two-/three-line subtitle
9. Line-break errors	- Syntactic categories (phrase-structure trees) - Grammatical functions

5. Results and Discussion

This chapter offers some preliminary results, compares findings with previous research and tentatively relates them to any relevant existing guidelines.

As mentioned in section 4.2, three programmes were analysed:

Programme A: *Atlantis Uncovered* (documentary)

Programme B: *Chimps on Death Row* (documentary)

Programme C: *The World's Most Wanted* ("newsy" current affairs programme)

From each programme, only the first, middle and last forty subtitles were analysed. Accordingly, each programme was divided in three parts, for instance programme A into parts A1, A2 and A3 respectively. Results are presented separately for each part and overall for each programme.

5.1 Subtitle rate

Subtitle rate is one of the most common metrics used in subtitling. Table 19 presents the subtitle rates for each programme and separately for each part as recorded by WinCAPS.

Table 19. Subtitle rate (wpm) - WinCAPS

	Programme A				Programme B				Programme C			
	A1	A2	A3	**All**	B1	B2	B3	**All**	C1	C2	C3	**All**
Mean	121.9	108.4	118.1	**116.1**	115.6	117.9	121	**118.2**	105.8	121.4	112.2	**113.1**
Standard Deviation	17.8	18.5	16.6	**18.4**	13.9	16.4	14.7	**15**	37.2	20.5	24.9	**28.9**
Maximum	147	139	150	**150**	139	141	141	**141**	194	153	152	**194**
Minimum	72	73	76	**72**	61	60	77	**60**	23	28	28	**23**
Range	75	66	74	**78**	78	81	64	**81**	171	125	124	**171**

Regarding this feature, there seems to be a difference between the two documentaries and the current affairs programme. Indeed, while *Atlantis Uncovered* and *Chimps on Death Row* exhibit similar subtitle rates with *Chimps on Death Row* being slightly faster-paced (at 118.2 wpm vs. 116.1 wpm), *The World's Most Wanted* adopts a somewhat lower rate at 113.1 wpm.[36] This pattern is not, nevertheless, repeated when we look at the separate parts analysed, as both parts A1 and C2 are faster than part B3 (which is the fastest part in programme B) and part A3 is also faster than parts B1 and B2).

Variation is also present within each programme. Apart from programme B where all mean values cluster more or less around the overall mean of 118.2 wpm, in both programmes A and C, values are more widely dispersed. In the case of *Atlantis Uncovered*, the value for part A2 (at 108.4 wpm) is considerably lower than the values for parts A1 and A3 and seems to be the odd one out. In the case of *The World's Most Wanted*, on the other hand, the mean values from each part are so widely dispersed that it is not possible to pick any out; the value from part C1 at 105.8 wpm, however, immediately draws one's attention as it is the lowest in the entire corpus.

This sense of variation in the programmes is also confirmed by the figures provided in Table 19 regarding the standard deviation (SD),[37] maximum value, minimum value and range for the subtitle rates recorded. What these figures reveal is that in programme C, and to a lesser extent programmes A and B, there is great variation and many individual subtitle rates are far from the mean. Without access to the video of each programme, it is difficult to determine whether this variation in subtitle rate, particularly for parts A2 and C1, is due to differences in speech rates in comparison

---

[36] However, this difference was not found to be statistically significant (p-value > 0.05) The p-value was calculated with a Welch t-test (variance was not equal) using an online calculator available at http://www.graphpad.com/quickcalcs/ttest1.cfm?Format=S (for more on t-tests, see Pym, 1999; Motulsky, 1999). It is important to note that although the difference was not found to be significant, this finding only applies to the particular three programmes. More programmes would be needed to draw conclusions with greater statistical accuracy.

[37] Standard deviation is a statistical measure of the dispersion of all the values in a data set, measuring how far away they are from the arithmetic mean (Pym, 1999, no page). If many values are close to the mean, then the standard deviation is small; if many values are far from the mean, then the standard deviation is large.

to other parts and/or whether it is due to any other factors. However, the extremely low subtitle rate recorded for part C1 seems to be due at least to a certain extent to the presence of very slow subtitles containing extralinguistic information; seven such subtitles (out of a total of ten in the entire programme) are included in part C1 with rates ranging from a minimum of 23 to a maximum of 50 wpm (and *The World's Most Wanted* is the only programme in the corpus to include such subtitles). For part A2, on the other hand, it is more difficult to account for the considerably lower subtitle rate, as this part does not seem to differ from parts A1 and A3. Indeed, in all three parts, there is a mixture of text narrated by the narrator and text belonging to the subgenre of interviews (see section 4.2.1) with various experts talking (only the experts' replies are included in each instance, not the questions). Only two differences can be discerned and are both shown in Table 20.

Table 20. Distribution of text (narration/interview/conversation) – Programme A

	Part A1	Part A2	Part A3	**Total**
Narration	29 (72.5%)	21 (52.5%)	27 (67.5%)	**77 (64.2%)**
Interview	11 (27.5)	12 (30%)	13 (32.5%)	**36 (30%)**
Conversation	0	7 (17.5%)	0	**7 (5.8%)**
Total	**40**	**40**	**40**	**120**

a) Whereas in parts A1 and A3, approximately 70% of subtitle text is narration, this percentage drops to 53% for part A2. It is not obvious how this might affect the mean subtitle rate of each part. In fact, the association of predominantly narrated parts with higher subtitle rates might even be puzzling, as one might assume that parts where narration is predominant would tend to be slower than parts containing more spontaneous speech, such as interviews and conversations (cf. Tables 23, 24 and 25, Figures 3, 4 and 5 and relevant analysis).

b) Part A2 includes a scene where various archaeologists are talking amongst themselves at an excavation site (subtitles A53 and A63 to A68) and is thus the only example of a conversation part. Although one might expect these subtitles to display

a higher subtitle rate than the mean, in fact the average subtitle rate here is extremely low at about 85 wpm (88 wpm if we include subtitle A53 which is rather independent from the rest). This seems to be due to the deletion of a number of oral features, including exclamations (*Oh boy.* between subtitles A65 and A66), false starts (in subtitle A67) and what seems to be overlapping speech (*That's a dog. That's a dog.* between subtitles 68 and 69) and dialogue heard in the background (*...use anything that's in the corner of this wall...* between subtitles A51 and A52). However, subtitles A53 and A63 to A68 are only seven of the 40 subtitles included in part A2. There are still another 20 subtitles with a subtitle rate of less than 120 wpm, of which 14 exhibit a rate of less than 110 wpm (see Table 22). To further examine the reasons why so many subtitles in part A2 exhibit such low rates particularly in comparison with parts A1 and A3, we would need access to the video in order to establish among other things speech rates and the complexity of visual information on screen.

More importantly, such information is indispensable in order to account for the different mean subtitle rates in the three programmes and identify any patterns if applicable. Unfortunately, this information is not available for the present corpus and any analysis is thus tentative and regrettably incomplete. For instance, it would appear that the two programmes where narration is predominant (*Atlantis Uncovered* and *The World's Most Wanted*) are those with the lowest mean subtitle rate (see Tables 20, 21, 22 and 19).

Table 21. Distribution of text (narration/interview/conversation) - Programme B

	Part B1	Part B2	Part B3	**Total**
Narration	19 (47.5%)	20 (50%)	18 (45%)	**57 (47.5%)**
Interview	21 (52.5%)	20 (50%)	22 (55%)	**63 (52.5%)**
Conversation	0	0	0	**0**
Total	**40**	**40**	**40**	**120**

Table 22. Distribution of text (narration/interview/conversation) - Programme C

	Part C1	Part C2	Part C3	Total
Narration - Reporter	25 (62.5%)	29 (72.5%)	18 (45%)	**72 (60%)**
Interview	6 (15%)	10 (25%)	15 (37.5%)	**31 (25.8%)**
Statement	2 (5%)	0	5 (12.5%)	**7 (5.8%)**
Sound trans.	7 (17.5%)	1 (2.5%)	2 (5%)	**10 (8.4%)**
Total	**40**	**40**	**40**	**120**

Assuming a causative relationship between narration and low subtitle rates, however, would be based on the hypothesis that i) a narrator's speech rate would be slower than the speech rate of a person interviewed or people in a conversation and b) accordingly subtitles in narration parts would exhibit slower rates than subtitles in parts with interviews or conversations. Unfortunately, without access to the speech rates of the three programmes the first part of the hypothesis cannot be tested and the second part can only be tested partially. In order to test the second part, the subtitle rates for each programme were divided according to the part of the programme the subtitle came from: narration, interview, statement/public address, conversation, sound transcription; the term "statement/public address" was used for parts of the programme where a public figure, in this case President George W. Bush, addresses the public. The relevant results are presented in Tables 23, 24 and 25 and Figures 3, 4 and 5 and can be summarised as follows:

Table 23. Distribution of subtitle rates according to which part of the programme subtitles come from - Programme A

	Narration				Conversation				Interview			
	A1	A2	A3	**All**	A1	A2	A3	**All**	A1	A2	A3	**All**
<100 wpm	4	4	5	**13**	0	4	0	**4**	1	2	1	**4**
100–110wpm	3	5	7	**15**	0	3	0	**3**	1	3	2	**6**
111–120 wpm	7	5	4	**16**	0	0	0	**0**	2	3	2	**7**
121–130 wpm	4	4	5	**13**	0	0	0	**0**	2	4	4	**10**
131–140 wpm	7	3	3	**13**	0	0	0	**0**	3	0	4	**7**
> 140 wpm	4	0	3	**7**	0	0	0	**0**	2	0	0	**2**

Figure 3. Subtitle rates according to which part of the programme subtitles come from - Programme A
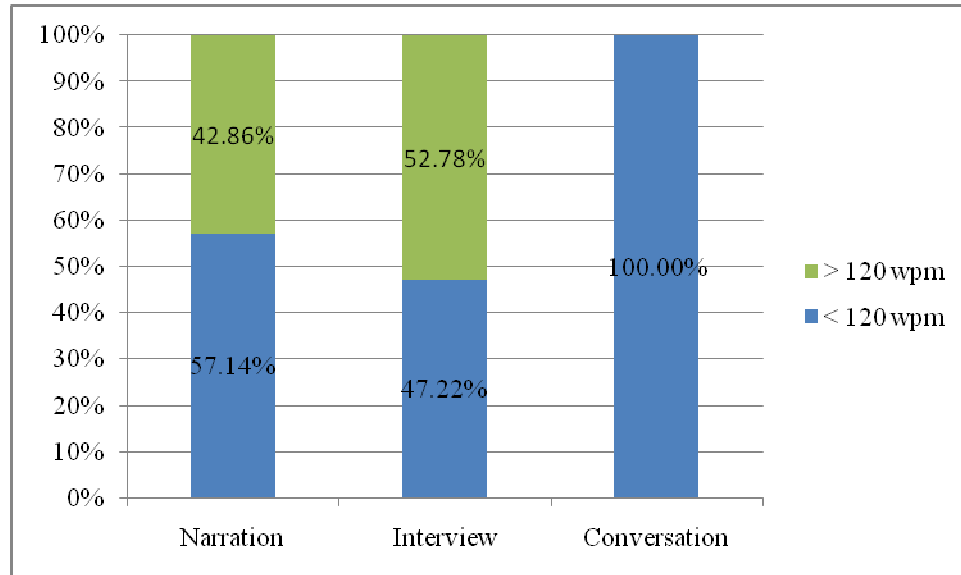
Table 24. Subtitle rates according to which part of the programme subtitles come from - Programme B

	Narration				Conversation				Interview			
	B1	B2	B3	**All**	B1	B2	B3	**All**	B1	B2	B3	**All**
< 100 wpm	2	1	2	**5**	0	0	0	**0**	1	2	2	**5**
100–100 wpm	3	2	2	**7**	0	0	0	**0**	5	3	1	**9**
111–120 wpm	8	7	1	**16**	0	0	0	**0**	8	5	5	**18**
121-130 wpm	3	7	10	**20**	0	0	0	**0**	5	7	10	**22**
131–140 wpm	3	1	3	**7**	0	0	0	**0**	2	3	3	**8**
> 140 wpm	0	2	0	**2**	0	0	0	**0**	0	0	1	**1**

Figure 4. Subtitle rates according to which part of the programme subtitles come from - Programme B
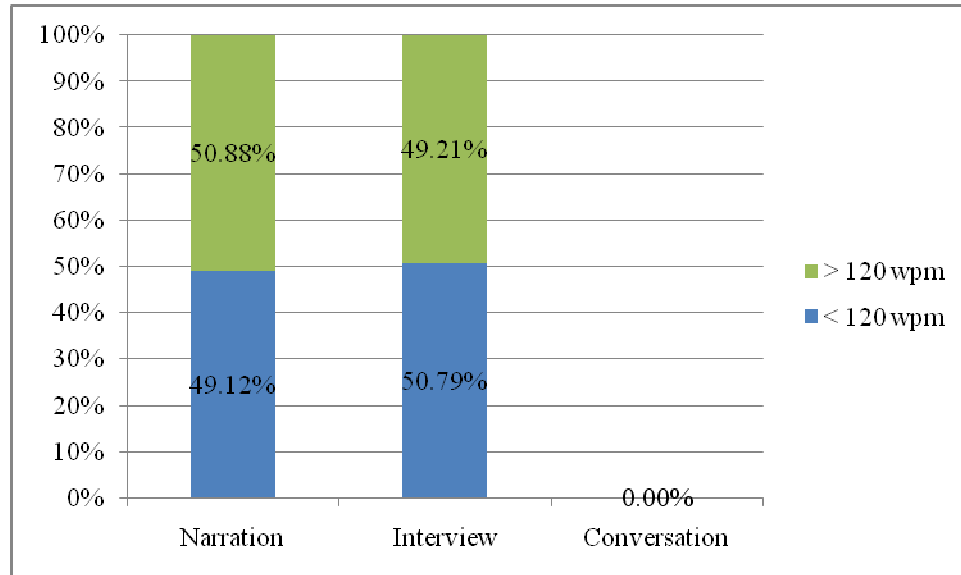
Table 25. Distribution of subtitle rates according to which part of the programme subtitles come from - Programme C

	Narration				Interview				Statement				Sound transcription			
	C1	C2	C3	**All**	C1	C2	C3	**All**	C1	C2	C3	**All**	C1	C2	C3	**All**
<100	0	1	1	**2**	2	1	2	**5**	1	0	3	**4**	7	1	2	**10**
100-10	2	5	2	**9**	1	1	1	**3**	0	0	1	**1**	0	0	0	**0**
111-20	11	6	5	**22**	0	1	5	**6**	0	0	1	**1**	0	0	0	**0**
121-30	4	9	5	**18**	3	3	6	**12**	1	0	0	**1**	0	0	0	**0**
131-40	6	3	3	**12**	0	3	0	**3**	0	0	0	**0**	0	0	0	**0**
>140	2	5	2	**9**	0	1	1	**2**	0	0	0	**0**	0	0	0	**0**

Figure 5. Subtitle rates according to which part of the programme subtitles come from - Programme C
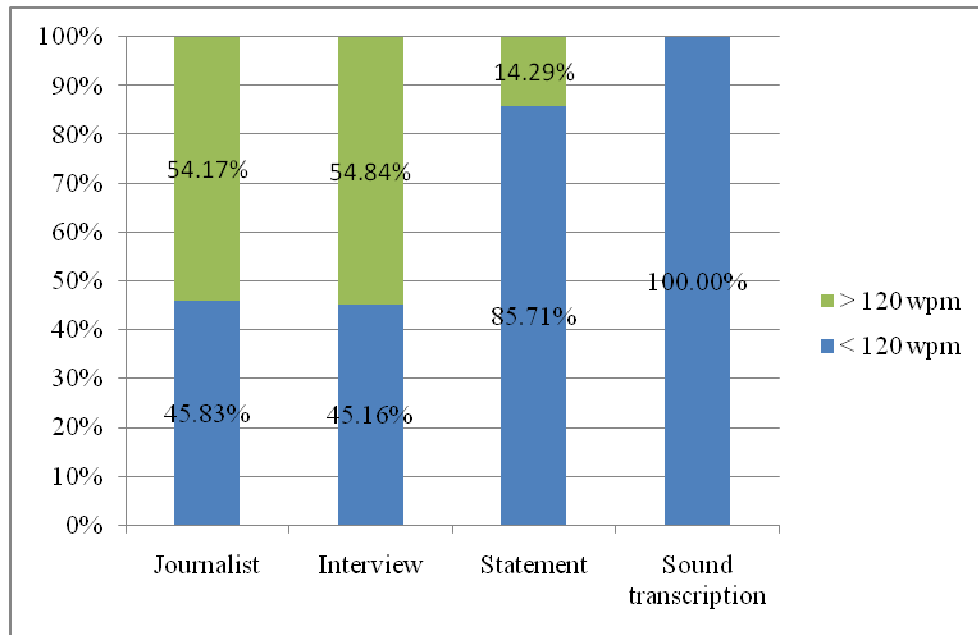
1) There is no evidence supporting the part of the hypothesis that narration parts are slower than interview parts. In fact, there seems to be no distinguishing pattern between narration and interviews. Across the three programmes, both narration and interview parts exhibit quite similar percentages of subtitle rates with the majority of subtitles displayed at over 120 wpm (the only exception is programme A where the majority of subtitles in narration parts are below 120 wpm). More importantly, as shown in Tables 23, 24 and 25, subtitles over 130 wpm tend to be more numerous in narration than interview parts (apart from programme B where there are nine subtitles in both cases). This might even suggest that interview parts can be or indeed are occasionally slower, possibly because the interviewee has consciously or unconsciously adopted a slow speech rate.

Unlike what happens with interview parts, a pattern can be discerned with subtitles from public addresses, as the overwhelming majority of subtitles (85.71%) are displayed at below 120 wpm. This would seem reasonable, as one might expect a politician delivering a formal speech to speak at a slower rate than people in an interview, let alone a conversation. This hypothesis would need to be confirmed, however, by checking the speech rates corresponding to the relevant subtitles. In

addition, since all instances come from *The World's Most Wanted*, further research is required to confirm whether this is really a tendency.

2) The figures provided in Table 23 suggest that the part of the hypothesis that narration parts are slower than conversation parts might be false. Indeed, as mentioned on pages 128-129, subtitles A53 and A63 to A68 which come from a part of the programme where some archaeologists are talking among themselves, all display rather low rates below 100 wpm with an average of 88 wpm. This low rate seems more likely to be due to the deletion of a high number of oral features (see section 5.6) rather than a decrease in speech rates, since we would expect people in a conversation to talk faster than a narrator on television; this is nevertheless merely a hypothesis that would need to be checked against the actual speech rates from the programme. In addition, since this is the only instance of conversation in the three programmes, further research involving more examples and access to the video is required before any conclusions can be reached.

3) Another possible pattern discerned concerns subtitles that include extralinguistic information, such as:

Subtitle C40:   GUNFIRE

Subtitle C105: CHEERING

There are only ten such subtitles in the corpus and they all come from programme C. The average rate for these subtitles is a mere 34.7 wpm with rates ranging from as low as 23 wpm to a maximum of 50 wpm. Since all examples come from a single programme, further research is again required to confirm this particular trend for such exceptionally low rates.

In order to compare the present findings with the findings reported by De Linde and Kay (1999), all subtitle rates were measured again using Equation 1 rather than WinCAPS (see section 4.3.1.5). The figures obtained are shown in Table 26:

Table 26. Subtitle rate (wpm) – Equation1

	Programme A				Programme B				Programme C			
	A1	A2	A3	**All**	B1	B2	B3	**All**	C1	C2	C3	**All**
Mean	121.2	114	122.4	**118.8**	122.4	127.8	127.8	**126**	121.2	118.8	117	**118.8**
SD	21	21.6	21	**21.6**	22.8	22.8	27	**24**	37.8	30	32.4	**33.6**
Maximum	159.6	160.2	168	**168**	156.6	171.6	169.8	**171.6**	220.8	178.2	170.4	**220.8**
Minimum	85.8	63.6	81.6	**63.6**	45	75	82.2	**45**	21	20.4	17.4	**17.4**
Range	73.8	96.6	86.4	**104.4**	111.6	96.6	87.6	**126.6**	199.8	157.8	153	**203.4**

Before anything else, we need to point out that the figures in Table 26 differ considerably from those calculated on the basis of WinCAPS in Table 19. As explained in section 4.3.1.5, the figures provided by WinCAPS were the ones used for the purposes of this research. Apparently, however, further research is required to identify and agree on a suitable metric for subtitle rate.

According to the figures in Table 26, the average subtitle rate for *Atlantis Uncovered* and *Chimps on Death Row* (122.4 wpm) is somewhat lower than the respective overall figure of 126 wpm reported by De Linde and Kay for the two documentaries included in their corpus (1999, p. 50, see also section 3.3.1). However, a closer look at the figures provided for each of the two documentaries (*Horizon* and *In Search of Our Ancestors*) reveals that the subtitle rates for the two programmes differed considerably with the mean rate for *Horizon* a mere 113 wpm compared to 139 wpm for *In Search of Our Ancestors* (ibid, p. 78 and 80). We also need to remember that De Linde and Kays' figures were calculated on the basis of just 39 subtitles (22 from *Horizon* and 17 from *In Search of Our Ancestors*). Given the small number of subtitles analysed, the figures provided cannot necessarily be assumed to provide an accurate indication of the average subtitle rate of each programme.

In addition, the subtitle rates reported in this thesis (calculated using either WinCAPs or Equation 1) are much lower than the mean rate of 139 wpm reported for documentaries by Jensema et al. (1996, p. 287-288, see also section 3.3.1). Although the figures are not directly comparable, as Jensema et al. do not specify the method

used to calculate subtitle rates, their findings seem to be indicative of the trend for faster subtitling in the USA compared to the UK and the rest of Europe (for an indication of this see also the guidelines by the DCMP discussed in section 3.3.1, Erard, 2001 and Neves, 2005, p. 142-143). Figures 6 and 7 present this information schematically.

Figure 6. Subtitle rate for documentaries - My corpus, De Linde & Kay, Jensema et al.
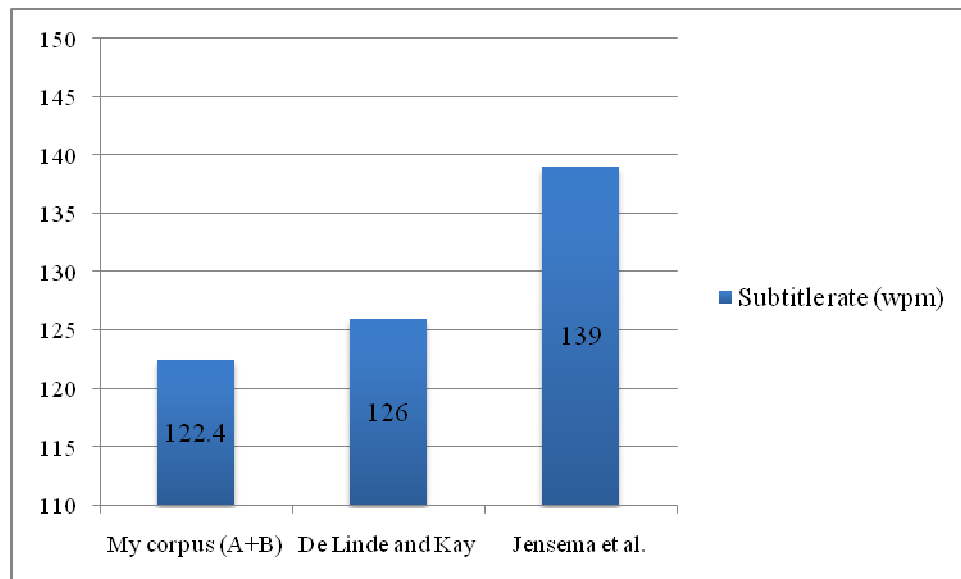
Figure 7. Subtitle rate for documentaries- My corpus, De Linde & Kay
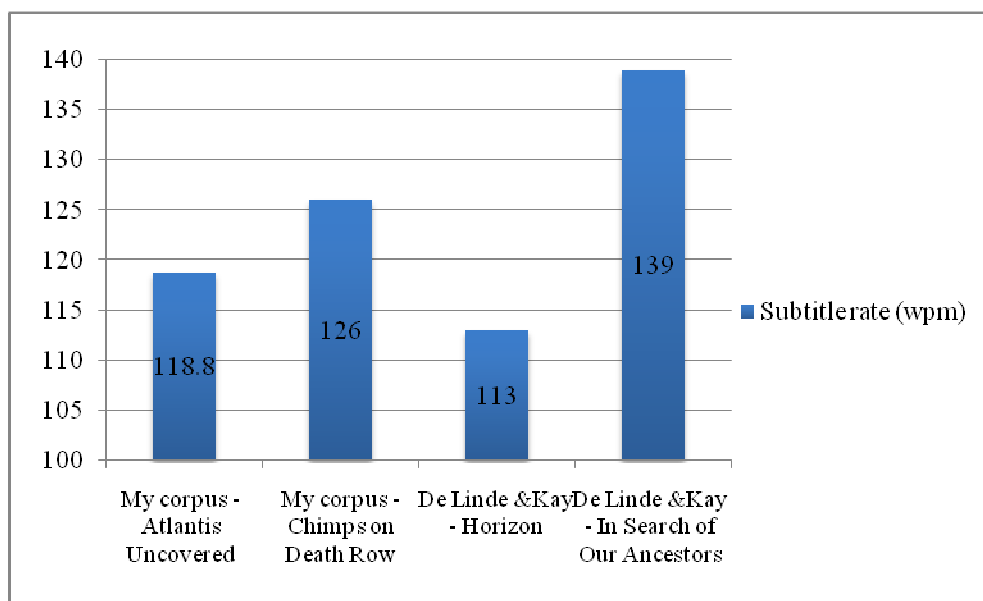
5.2 Display time

Table 27 and Figure 8 present the distribution of subtitles in all three programmes according to their display time on screen.

Table 27. Display time

Display time	Programme A				Programme B				Programme C			
	A1	A2	A3	**All**	B1	B2	B3	**All**	C1	C2	C3	**All**
<2 sec.	0	0	0	**0**	0	0	0	**0**	0	0	0	**0**
2-3 secs	1	6	3	**10**	3	2	4	**9**	14	6	12	**32**
3-4 secs	6	5	4	**15**	2	4	1	**7**	4	6	12	**22**
4-5 secs	8	4	8	**20**	12	13	14	**39**	14	18	9	**41**
5-6 secs	16	17	12	**45**	9	11	10	**30**	5	7	7	**19**
> 6 secs	9	8	13	**30**	14	10	11	**35**	3	3	0	**6**

Figure 8. Display time per programme
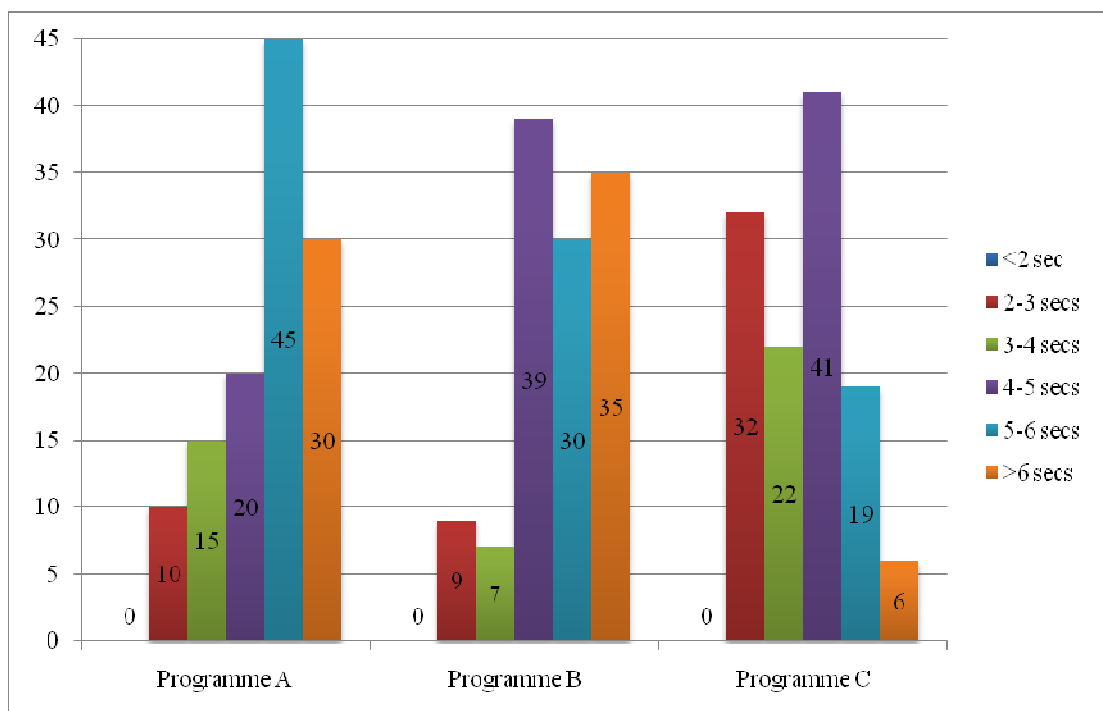
According to the figures recorded, subtitles in the two documentaries tend to appear on screen for more time than those in the current affairs programme. Nevertheless, if

we take a closer look, a slightly more complex picture emerges with the three programmes forming a sort of continuum. Thus, although in both *Atlantis Uncovered* and *Chimps on Death Row*, the majority of subtitles, nearly 80% and 87% respectively, are displayed for more than four seconds, in *Atlantis Uncovered* most of these subtitles are shown for over five seconds, whereas in *Chimps on Death Row*, they tend to be marginally briefer remaining on screen for 4-5 seconds. In *The World's Most Wanted*, on the other hand, subtitles tend to be displayed for even less time with the majority, approximately 80%, appearing on screen for two to five seconds. Although most of these subtitles are displayed for 4-5 seconds (almost to the same extent as in programme B), there are also many subtitles that are only shown for 2-3 (26%) or 3-4 seconds (18%) unlike what happens with both documentaries, where such brief subtitles are only sparingly used.

In previous research, this metric has only been used by Lomheim (1999). As his figures were based on the analysis of three TV series and their interlingual subtitles for hearing viewers (see relevant discussion in section 3.3.1 and in particular Table 6) it would not make much sense to compare his results with the figures provided here. What is important to note, however, is that Lomheim used his results to categorise the three programmes in terms of how fast viewers have to read and in terms of how much they have to read; thus, the following conclusions were reached (1999, p. 198):

- in *Allo! Allo!* (where the highest percentage of subtitles was in the 2-3 seconds category), it was thought that viewers have to read quickly,
- in *Maigret* (4-5 seconds), it was concluded that viewers have to read a lot, but do not need to read so quickly, and
- in *Golden Years* (3-4 seconds), the conclusion reached was that viewers have to read neither quickly nor a large amount of text

Although Lomheim does take into account the number of characters displayed in a subtitle (longer subtitles should be displayed for more time than short ones), his conclusions are based on an idealised view of the relationship between display time

and amount of text. This relationship, however, cannot be taken for granted, as quite often, subtitles are not timed according to the number of characters they include (for the factors influencing the duration of a subtitle, see sections 2.4.3 and 3.3.1). This entails for instance that a subtitle that might be considered slow on the basis of exposure time alone may actually have to be read very quickly by viewers on account of the high number of characters included. For example, let us consider subtitle B66:

It was the Americas' first-known
plant domestication - prompted by
practicality rather than hunger.   (97 characters)

This subtitle appears on screen for 6.24 seconds and could thus be considered very slow. Yet, given the high number of characters included, this subtitle is actually extremely fast reaching a rate of 148 wpm (as calculated by WinCAPS) and is potentially too fast for many viewers. Applying Lomheim's interpretations to this subtitle, we would classify it as a subtitle where viewers have to read a lot, but do not need to read so quickly. Apparently, this is not the case, as in this instance not only do viewers have a lot to read, but they also need to read it very quickly. This is an indication of how this metric is insufficient in itself and needs to be combined with information about subtitle rates (see section 5.1) and crucially with information about whether time constraints are respected or not (see section 5.3).

5.3 Time constraints respected or not

As pointed out in section 4.3.1.6, this seems to be the least objective of the metrics used for the purposes of this research. This is mainly due to two reasons:

1) There is no generally accepted maximum subtitle rate for D/deaf and HoH viewers. In the UK, older guidelines recommended a rate of 120 wpm (see Baker, 1981 and Baker et al., 1982), which was later raised to 140 wpm (ITC, 1997 and

ITC, 1999) which now seems to be under revision again (Ofcom, 2005). In other countries, recommendations for optimum rates range from 120-130 wpm (Portugal, Neves, 2005) to 180 wpm (Deafness Forum of Australia, 2004) to near-verbatim rates reaching up to 235 wpm (DCMP, 2007). None of these rates, however, seem to have been arrived at on the basis of large-scale research involving D/deaf and HoH people. For this reason, the maximum rates considered for the purposes of this research were set to 120, 130 and 140 wpm.

2) The decision as to whether a subtitle respects time constraints or not was solely based on the amount of text (measured in characters) included in the subtitle and the subtitle rate set as the maximum in each case (120, 130 or 140 wpm). However, although this might be appropriate for some or even many subtitles, as mentioned in section 3.3.1, the amount of text displayed is only one of the factors influencing the amount of time a subtitle remains on screen (other factors including among other things the speech rate, the amount of visual information on screen and camera cuts). Regrettably, this information was not available for the purposes of this research, as the present corpus does not contain the video of each programme. Nevertheless, the information provided in Tables 28, 29 and 30 is thought to be indicative to a certain extent of what happens in each programme.

For Table 28, the maximum subtitle rate was set to 120 wpm (as calculated by WinCAPS). Interestingly, although all three programmes have a mean subtitle rate of less than 120 wpm (see section 5.1), they were all found to contain a high number of subtitles not respecting time constraints. The majority of such subtitles were recorded for programme B, which is also the programme with the highest mean subtitle rate at 118.2 wpm; thus, 50% of subtitles in *Chimps on Death Row* do not respect time constraints vs. 48% for *The World's Most Wanted* and 42% for *Atlantis Uncovered*.

Table 28. Subtitles not respecting time (rate at 120wpm)

	Programme A				Programme B				Programme C			
	A1	A2	A3	**All**	B1	B2	B3	**All**	C1	C2	C3	**All**
Subtitles not respecting time	21	11	18	**50**	13	20	27	**60**	16	24	17	**57**
Verbatim	8	3	9	**20**	6	7	11	**24**	7	9	4	**20**
Edited	13	8	9	**30**	7	13	16	**36**	9	15	13	**37**

Although this might be hardly surprising, what is surprising is that when the subtitle rate is set to 130 and 140 wpm (Tables 29 and 30 respectively), a totally different picture emerges with programme A being the one with the highest number of subtitles not respecting time constraints at 130 wpm and programme C at 140 wpm.

Table 29. Subtitles not respecting time (rate at 130wpm)

	Programme A				Programme B				Programme C			
	A1	A2	A3	**All**	B1	B2	B3	**All**	C1	C2	C3	**All**
Subtitles not respecting time	16	3	10	**29**	5	6	7	**18**	8	12	6	**26**
Verbatim	5	2	4	**11**	3	0	3	**6**	2	4	1	**7**
Edited	11	1	6	**18**	2	6	4	**12**	6	8	5	**19**

Table 30. Subtitles not respecting time (rate at 140wpm)

	Programme A				Programme B				Programme C			
	A1	A2	A3	**All**	B1	B2	B3	**All**	C1	C2	C3	**All**
Subtitles not respecting time	6	0	3	**9**	0	2	1	**3**	2	6	3	**11**
Verbatim	2	0	1	**3**	0	0	0	**0**	1	2	1	**4**
Edited	4	0	2	**6**	0	2	1	**3**	1	4	2	**7**

This entails that it is in fact the programmes with the lowest subtitle rates (programme A at 116.1 and programme C at 113.1 wpm) that contain more subtitles

not respecting time constraints. However paradoxical this might seem, it might be explained by combining the information from Tables 28, 29 and 30 with the data presented in Table 19. Thus, according to the figures provided in Table 19, although the mean subtitle rate for programmes A and C is lower than that of programme B, both standard deviation and maximum value are higher. As a result, more subtitles in these two programmes are far from the mean and many of them are also considerably higher than the mean (as attested by the data provided in the tables in the present section). There is little doubt that this is an important indication of why however valuable the information provided by the mean, it needs to be complemented with other metrics.

As seen in Tables 28, 29 and 30, all subtitles not respecting time constraints were categorised according to whether they were verbatim or edited (see sections 4.3.1.1 and 5.7). This information is also presented in Figures 9, 10 and 11. In all cases, the majority of subtitles recorded were found to be edited rather than verbatim. This appears somewhat strange, as one might expect the exact opposite to happen. It would seem to indicate that while editing has been used, it has not been applied sufficiently and further reduction should have been called for and/or that spotting was not adequate.

Figure 9. Subtitles not respecting time - verbatim vs. edited (rate at 120wpm)
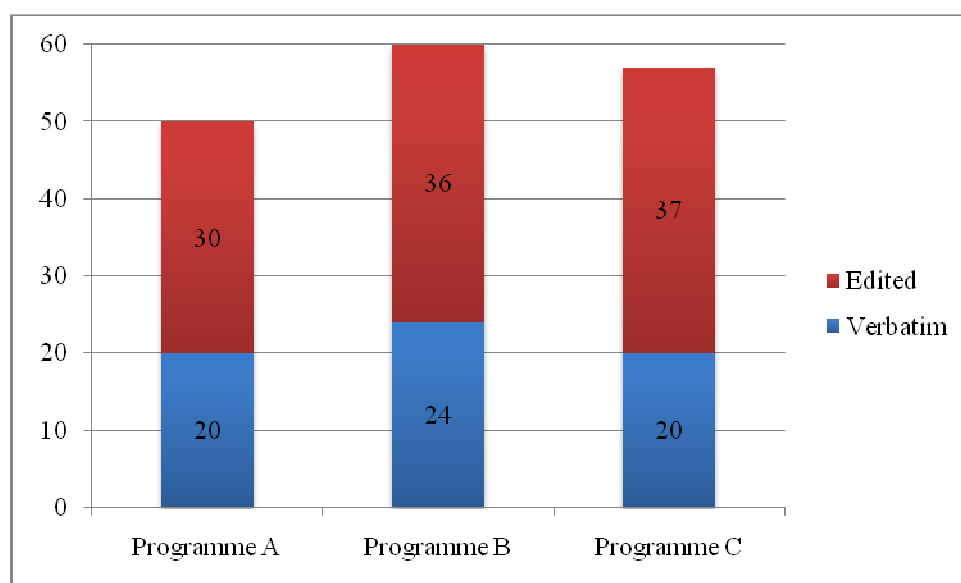
Figure 10. Subtitles not respecting time - verbatim vs. edited (rate at 130wpm)
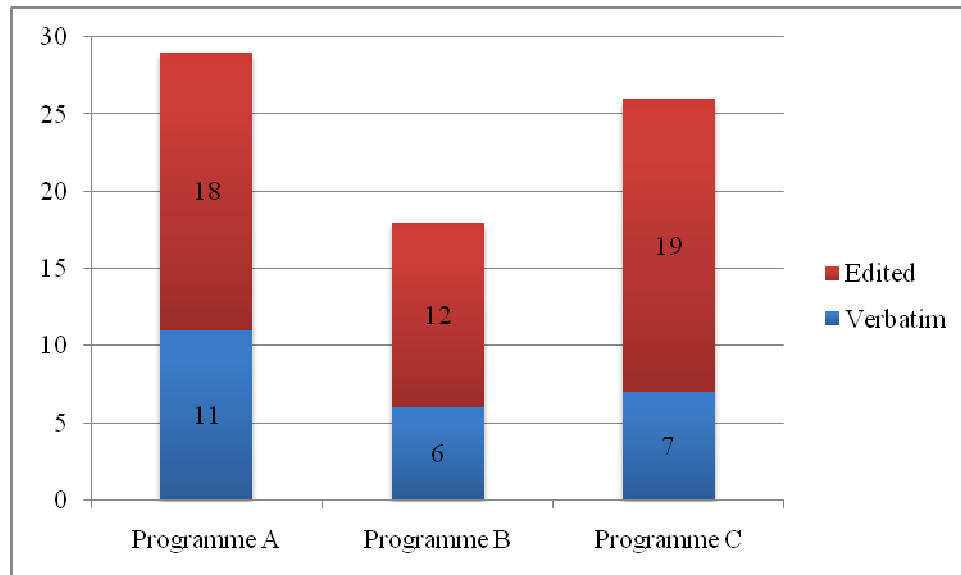
Figure 11. Subtitles not respecting time - verbatim vs. edited (rate at 140wpm)
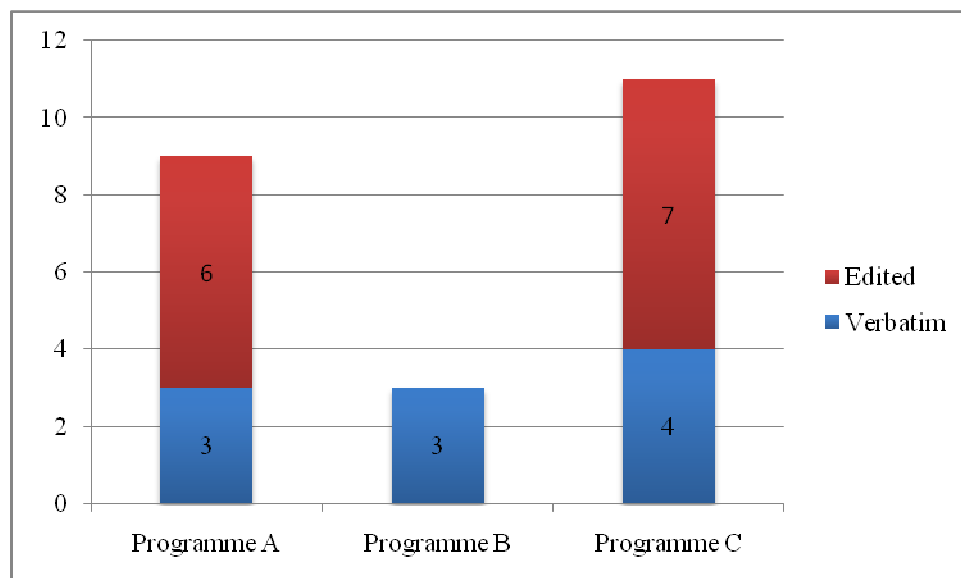
5.4 Subtitle intervals

The subtitle intervals for all three programmes can be broken down as shown in Table 31.

Table 31. Subtitle intervals

Subtitle intervals	Programme A				Programme B				Programme C			
	A1	A2	A3	**All**	B1	B2	B3	**All**	C1	C2	C3	**All**
1 frame	21	30	31	**82**	27	33	35	**95**	31	36	39	**106**
2 frames	0	0	0	**0**	0	0	0	**0**	1	0	0	**1**
3 frames	0	0	0	**0**	0	0	0	**0**	0	0	0	**0**
4-12 frames	0	0	0	**0**	0	0	0	**0**	1	0	0	**1**
13-24 frames	0	0	0	**0**	0	0	0	**0**	0	0	0	**0**
1-2 secs	5	1	0	**6**	5	3	0	**8**	3	1	0	**4**
2-4 secs	6	6	1	**13**	2	2	3	**7**	2	1	0	**3**
> 4 secs	7	2	7	**16**	5	0	1	**6**	1	1	0	**2**

This is the only feature with consistent results in all three programmes (including their parts), as the majority of subtitles (80.81%, 283 out of 350) are separated by only one frame. Even though this might be acceptable in the case of continuous subtitles, such subtitles only account for a mere 36.04% (102 out of 283 subtitles). In all, looking at these figures and the figures provided in Table 27, the impression created is of three programmes with a more or less rapid rhythm of subtitling. Since neither documentaries nor current affairs programmes are usually associated with a fast pace and dramatic action, the predominance of short subtitle intervals might be seen as a confirmation of Lomheim's conclusion that "a rapid pace in the subtitles (short intervals) does not necessarily go hand in hand with a fast pace and dramatic action in the programme" (1999, p. 199). However, this hypothesis would need to be confirmed by watching the programmes themselves.

We should point out that although the total of subtitle interval times is 117 per programme, only 116 intervals have been recorded for programme C. This is because the interval between subtitles C79 and C80 is not accurate, since there seem to be

some subtitles missing (the interval between the subtitles is more than 2 minutes long and there are lines in the transcript file not present in the subtitle file).

The figures reported here seem to be in accordance with the somewhat vague recommendation in the BBC Subtitling Guide to only separate subtitles that immediately follow each other by one frame only (see section 3.3.2). At the same time, they are in direct contradiction to interlingual subtitling guidelines, which recommend a minimum of 2-3 frames between any subtitles. They are also quite different from the results presented by Lomheim, who found no subtitles with an interval of less than two frames in a sample of more than 1,000 interlingual subtitles (1999, p. 199).

5.5 Lines per subtitle

Table 32 shows the distribution of subtitles according to the number of lines they contain:

Table 32. Distribution of one-, two- and three-line subtitles

Lines	Programme A				Programme B				Programme C			
	A1	A2	A3	All/ %	B1	B2	B3	All/ %	C1	C2	C3	All
3-liners	9	7	13	29/ 24.2%	13	11	11	35/ 29.2%	0	0	0	0
2-liners	27	22	23	72/ 60%	24	26	25	75/ 62.5%	25	33	25	83/ 69.2%
1-liners	4	11	4	19/ 15.8%	3	3	4	10/ 8.3%	15	7	15	37 30.8%

According to the figures recorded, the majority of subtitles in all three programmes and their parts are two-liners, (230 subtitles out of the 360 analysed or 63.9%). With regards to three-line subtitles, what is noteworthy is that while both documentaries contain approximately 25% of three-liners, the current affairs programme does not include any at all. This is even more striking, if we consider that *The World's Most*

Wanted is the only programme that contains any subtitles with extralinguistic information such as the following:

Subtitle C15:   GUNSHOT

Subtitle C34:   GUNFIRE

Subtitle C110: will hear all of us soon.
                            CHEERING

One would expect the inclusion of such information to promote the use of three-line subtitles, but actually no such instances were recorded. On the other hand, even without including such information, approximately 30 subtitles from each of the two documentaries include three lines of text. This tendency might be accounted for in terms of an attempt at verbatim subtitling, while trying to fit everything within the same subtitle. This would indeed seem to be the case, because although the current affairs programme is the one with the highest rate of words from the original dialogue in the subtitles (see Table 35), it is also the programme with the lowest number of words in the original.

As far as the BBC Subtitling Guide and other intralingual SDH guidelines are concerned, *The World's Most Wanted* is the programme which abides most closely to relevant recommendations by favouring a maximum length of two lines per subtitle. Unlike most guidelines available (see section 3.3.3), however, it completely avoids three-line subtitles. At the other extreme, both *Atlantis Uncovered* and *Chimps on Death Row* seem to use three-liners more freely.

It would appear that there are two main reasons why most of the SDH guidelines mentioned here prefer two-line subtitles rather than three-liners. The first reason has to do with the image: the more text there is on the screen, the bigger the part of the visuals that is covered. The second reason, on the other hand, has to do with time; very often three-line subtitles seem to disappear from the screen quite fast making it

hard for viewers to read the entire subtitle within the limited amount of time available, let alone to actually watch the programme. Given that the present corpus does not include the video for any of the programmes, it was only possible to investigate the second reason. In order to do that, all three-line subtitles included in programmes A and B (there are no three-liners in programme C) were categorised according to whether they respected time constraints or not. For the reasons outlined in section 5.3, the maximum subtitle rates were again set to 120, 130 and 140 wpm and relevant findings are displayed in Table 33, while Table 34 (a partial reproduction of Table 32) reminds us of the total number of three-liners in the two programmes.

Table 33. Three-line subtitles not respecting time constraints

	Programme A				Programme B			
	A1	A2	A3	**Total**	B1	B2	B3	**Total**
120 wpm	7	3	11	**21 (72%)**	5	9	8	**22 (63%)**
130 wpm	7	1	8	**16 (55%)**	3	0	1	**4 (11%)**
140 wpm	2	0	3	**5 (17%)**	0	0	0	**0**

Table 34. Number of three-line subtitles in programmes A and B

Lines	Programme A				Programme B			
	A1	A2	A3	**Total**	B1	B2	B3	**Total**
3-liners	9	7	13	**29**	13	11	11	**35**

As can be seen from the figures provided, the majority of three-line subtitles in the two programmes (72% and 63% respectively) do not indeed respect time constraints when the maximum subtitle rate is set to 120 wpm. Although it is true that the duration of a subtitle is not only determined according to the amount of text (characters) it includes (see section 3.3.1), both percentages are quite high to suggest that many of the three-liners analysed would potentially be a problem for viewers in terms of time. When the subtitle rate is increased, though, to 130 wpm, the percentage of three-liners not respecting time constraints remains high (55%) for

only one of the programmes (*Atlantis Uncovered*) as opposed to a mere 11% for *Chimps on Death Row*. When the maximum rate is set even higher (140 wpm), the percentage for programme A drops considerably to just over 17%, while programme B becomes completely error-free. Overall, the figures presented in this section seem to suggest that the answer to the question whether three-line subtitles respect time constraints depends on the actual programme analysed and the maximum subtitle rate used as a metric. Thus, when the maximum rate is low (120 wpm), three-liners tend to flout time constraints (although this is also true for two-line and one-line subtitles too as recorded in the present corpus). For higher rates, three-line subtitles may or may not be displayed for a sufficient amount of time. This again underlines the need for wide-scale empirical research into the optimum subtitle rate for D/deaf and HoH viewers. In addition, further corpus research including more programmes is required to determine whether three-line subtitles generally tend to flout time constraints or not.

5.6 Reduction rates

As mentioned in section 4.1.2, verbatim subtitling vs. editing is a hotly debated and highly charged issue in the context of SDH. Nevertheless, there has hardly been any relevant large-scale research describing current subtitles in that respect: are current subtitles verbatim, are they edited? And if so, how much editing is done? These issues will be investigated in this section and section 5.7. For the purposes of this section, deletion was measured in two ways: Table 35 presents the overall subtitling and reduction percentages for each programme, while Table 36 shows the mean number of words deleted per subtitle.

Table 35. Comparison of original word count and subtitle word count

	Word count - transcript				Word count - subtitles				Subtitles as % of transcript/ Reduction rate			
	Part1	Part2	Part3	**All**	Part1	Part2	Part3	**All**	Part1	Part2	Part3	**Total**
A	466	503	490	**1459**	421	391	447	**1259**	90.34	77.73	91.22	**86.29// 13.71**
B	499	507	493	**1499**	441	433	435	**1309**	88.38	85.40	88.24	**87.32// 12.68**
C	332	438	334	**1104**	313	354	301	**968**	94.28	80.82	90.12	**87.68// 12.32**

Table 36. Number of words deleted per subtitle

Number of words deleted	Mean				Standard Deviation			
	Part1	Part2	Part3	**All**	Part1	Part2	Part3	**All**
Programme A	3.21	4.71	3.31	**3.94**	3.29	4.02	1.80	**3.40**
Programme B	3.58	3.95	3.16	**3.56**	2.53	2.72	2.39	**2.52**
Programme C	1.93	3.65	2.25	**2.72**	1.39	2.87	1.55	**2.25**

Overall, the extent of reduction is quite similar in all three programmes hovering around 13% (Table 35) or 3 words per subtitle (Table 36). This impression of homogeneity, however, is not in line with the findings from each of the separate parts of the programmes. The second part of each programme (and this is particularly true for *Atlantis Uncovered* and *The World's Most Wanted*) displays an evident decline in the percentage of original words appearing in the subtitles and a rise in the mean number of words deleted per subtitle. The reason why this might happen is not obvious and we would need to explore the issue further while acquiring information about, among other things, speech rates and the complexity of visual information on screen. The only part for which an explanation seems available is the scene in part A2 where various archaeologists are talking amongst themselves at an excavation site (see section 5.1). Since this is a conversation, a high number of oral features are deleted, including exclamations, false starts and overlapping speech. In all, a total of 32 words are deleted from this conversation with an average of 4.57 words per subtitle, when the average for the entire programme is 3.94 words per subtitle.

Table 36 also presents the standard deviation for the number of words omitted in each of the three programmes. In our case, *Atlantis Uncovered* is the programme with the highest standard deviation (3.40), as well as the programme with the highest mean of omissions (3.94 words per subtitle). The total figure for standard deviation here is rather high compared to the figures from the other two programmes suggesting that in the particular programme, the number of words deleted per subtitle varies to a great extent this being particularly true for part A2. On the other hand, the total figure for standard deviation for *Chimps on Death Row* and *The World's Most Wanted* is lower than in *Atlantis Uncovered* suggesting that values are more closely grouped and variation is less marked. The reason why programme A differs in that respect is not obvious and further research is required.

It is important to note that the percentages in Table 35 are only indicative of the extent to which verbatim transcription has been used. That is because even though counting the words of the original and comparing them to the subtitle word count may be more objective in intralingual rather than interlingual subtitling (see section 4.3.1.9), it still does not take into account alterations, such as paraphrases, specifications, generalisations, expansions and the use of alternative spellings. For instance, although the word count here is the same for both the original and subtitle, this is by no means a verbatim subtitle:

Transcript: we provide 'em a high quality life at a stabilised level of funding. (13 words)

Subtitle B116: by giving them a high quality life
at a stable level of funding.   (13 words)

On the other hand, this is without a doubt a verbatim subtitle:

Transcript: they could be in this bio containment housing for another 20 years.
                (12 words)

Subtitle B77: they could be
                in this bio-containment housing
                for another 20 years.   (11 words)

Still, counting the words of the original and comparing this figure with the number of words in the subtitles, as well as counting deletions and calculating the average number of words deleted seem to be the only straightforward and to a certain extent reliable methods of determining the extent of reduction for each programme (cf. last paragraph in this section).

In comparison with previous research, the figures presented in Table 36 for *Atlantis Uncovered* and *Chimps on Death Row* (average: 3.75 words per subtitle), are somewhat higher than the figure reported by De Linde and Kay for the two documentaries in their corpus (1999, p. 50). The two figures (presented in Figure 12) are nevertheless much closer than the figures reported by De Linde and Kay regarding the rest of the genres in their corpus (ibid).

Figure 12. Words deleted per subtitle – Documentaries in present corpus vs. De Linde and Kay
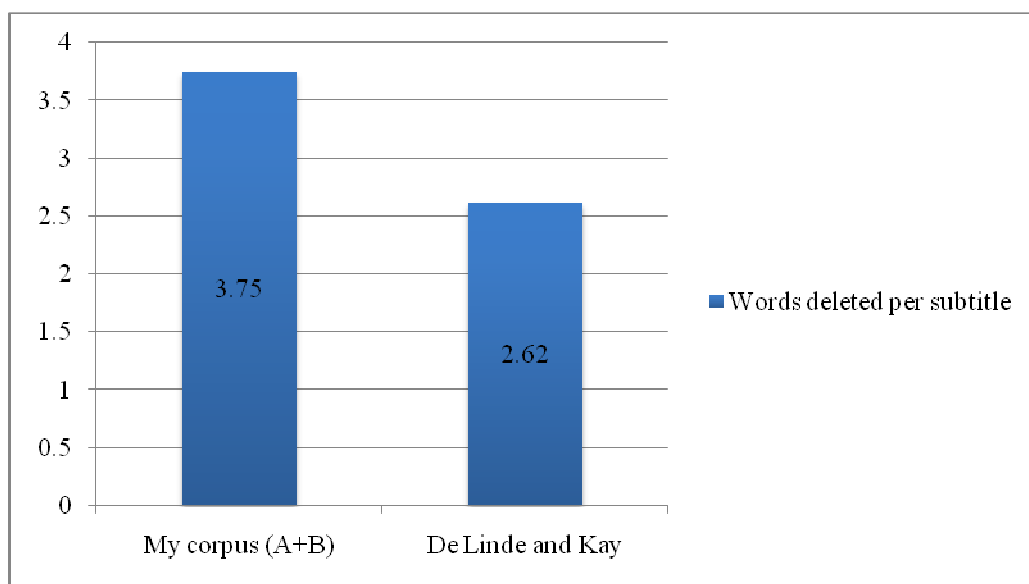
However, as mentioned in section 5.1, we need to remember that the figure provided by De Linde and Kay was calculated on the basis of just 39 subtitles. In addition, it seems that the extracts analysed only include narration and there are no parts with interviews, etc. (ibid, p. 78, 82). Any further comparisons or generalisations would thus be inopportune.

On the other hand, the subtitling percentages reported in Table 35 for *Atlantis Uncovered* and *Chimps on Death Row*, 86.29% and 87.32% respectively, are considerably lower than the figures reported by Jensema et al. for the two documentaries analysed in their corpus, i.e. 100% and 99% respectively (1996, p. 289). The average figures for the two corpora are displayed in Figure 13.

Figure 13. Percentage subtitled - Present corpus vs. Jensema et al.
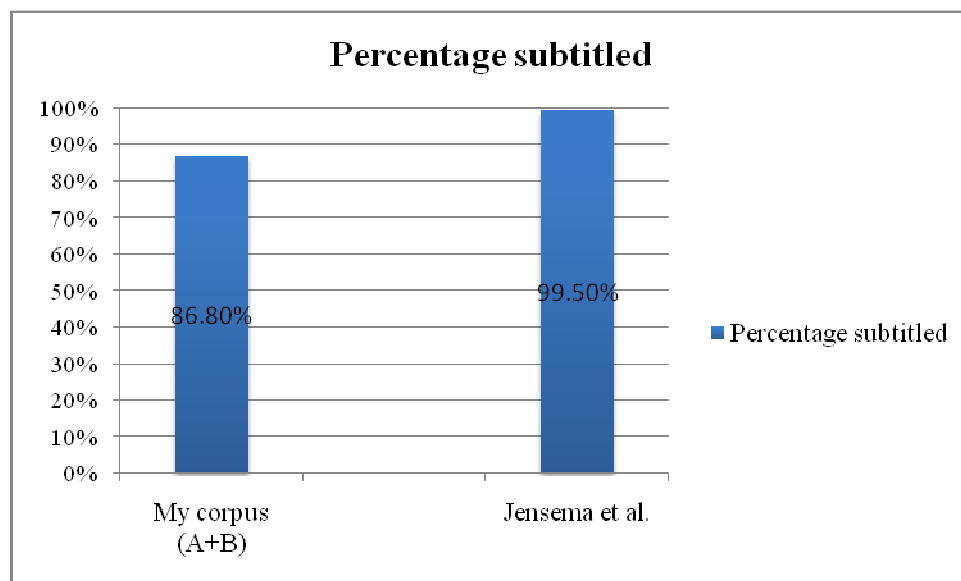
In addition, the figures for both *Atlantis Uncovered* and *Chimps on Death Row* are much lower than the average figure (94%) reported by Jensema et al. for all 13 genres, including drama series and children's programmes (ibid, p. 288). It would seem thus that while television subtitles (captions) broadcast in the USA are near-verbatim, subtitles broadcast on UK television tend to delete more while still striving to transfer as much as possible from the original. Last but not least, the reduction rates presented in Table 35 (13.71%, 12.68% and 12.32%) are considerably lower

than the figures reported by researchers working in the field of interlingual subtitling where the average reduction rate has been found to be approximately 30% with figures ranging from 19% to 40% (see section 3.3.5, pages 80-84). This is very likely due to the difference between intralingual and interlingual subtitling, but also due to the difference between genres: the present corpus consists of programmes belonging to the informative genre, whereas the interlingual figures are all based on the analysis of films and television series. Undoubtedly, analysing further genres and programmes from each genre in both intralingual and interlingual subtitling is required to further explore this aspect of reduction.

There is little doubt that the type of analysis provided here is valuable only to a certain extent, as it is only of a quantitative nature. A qualitative analysis would be required in order to determine whether the quantitative reduction of 13% is accompanied by a similar qualitative loss of information. Qualitative loss is here taken to mean loss of information other than oral language features (such as repetitions and false starts, see section 2.4.6 pages 35)[38] and information that viewers can pick up due to the positive feedback provided by the rest of the semiotic channels, in particular the image (Gottlieb, 1992, p.165). Since only the transcript and subtitle files were available for analysis, it was not possible to investigate the impact of intersemiotic redundancy on deletion. Regarding the nature of deletions, on the other hand, a partial classification of deletions was undertaken using the categories/descriptive labels suggested in section 4.3.1.10 in order to determine whether oral language features are indeed those deleted most in the process of subtitling (see section 5.8).

We also need to point out that the figures from Tables 35 and 36 bring to the fore the danger of analysing a limited sample from one's material, particularly the danger of arriving at generalisations on the basis of a limited number of subtitles. The findings reported here are only valid for the particular parts analysed, are indicative of the extent of reduction for the three programmes involved and can only be generalised both for the programmes themselves and their respective genres by analysing a

---

[38] See however the same section for the potential importance of such oral features for character portrayal, etc.

carefully selected proportion or all of the subtitles of each programme and by comparing them with findings from other documentaries and current affairs programmes. Nevertheless, as mentioned in section 4.2.2, the main purpose of the present research was to set out a methodology for analysing subtitles that could be used by large-scale research projects.

5.7 Strategies used

For the purposes of this part of the research, all subtitles were categorised according to the strategies suggested in section 4.3.1.1. Thus, two main groups were identified: verbatim subtitles (exact and complete transcription of speech) and edited subtitles, i.e. subtitles where at least one editing strategy (deletion, condensation, paraphrase, specification, addition or transposition) had been used. During the analysis stage, however, it became evident that the following adjustments to the original categories had to be made:

1) The category of verbatim subtitles was expanded to include:

a) Subtitles where minor errors have been amended, e.g.

Transcript: Could it <u>possible</u> be true?

Subtitle A8:   Could it <u>possibly</u> be true?

Transcript: In the 40 years they've been used, in research, the lives of captive chimps have <u>been proved</u>.

Subtitle B80:   In the 40 years they've been used
                in research, the lives
                of captive chimps have <u>improved</u>.

Nine instances of error correction were found in programmes A and B (five and four instances respectively). Although without access to the video of each programme, it is difficult to tell where such errors originate from, it seems more likely that they were introduced in the transcription rather than that they originate in the original audio.

b) Subtitles that include words which have been spelled differently than in the transcription, such as:

Transcript: But the Maya are only part of the <u>Mezzo-American</u> story.

Subtitle A43: But the Maya are only part
               of the <u>Mesoamerican</u> story.

Transcript: they could be in this <u>bio containment</u> housing for another 20 years.

Subtitle B77: they could be
               in this bio-containment housing
               for another 20 years.

2) A third category had to added, as apart from verbatim and edited subtitles, programme C also includes subtitles containing extralinguistic information, such as:

Subtitle C20: GUNSHOT

Subtitles like the following were also included under this category:

Subtitle C5:   Oh, my God!

This exclamation is not included in the transcript. Although this might just be an omission, it seems likely that the phrase was not part of the original script. Indeed, this exclamation (along with other two in subtitles C3 and C4) is uttered right after

the reporter has introduced the topic of the programme: *It stood proud on the New York skyline, a symbol of the American dream. On Tuesday it was shattered. Within hours suspicion fell on one man.* It would seem reasonable to assume that at that point footage of the destruction of the World Trade Center to which the reporter has been referring is shown and that the exclamations all come from that footage.

With these adjustments to the original categories in mind, Table 37 presents a breakdown of all subtitles according to the overall strategy used in each case.

Table 37. Distribution of subtitles according to overall strategy used

	Programme A				Programme B				Programme C			
	A1	A2	A3	**All**	B1	B2	B3	**All**	C1	C2	C3	**All**
Verbatim subtitles	23	17	25	**65**	18	20	19	**57**	15	16	14	**45**
Edited subtitles	17	23	15	**55**	22	20	21	**63**	18	23	22	**63**
Sound transcription	0	0	0	**0**	0	0	0	**0**	7	1	4	**12**

Looking at the figures from the two documentaries, a rather mixed picture emerges; thus, while in programme A, there are more verbatim rather than edited subtitles, in programme B, it is the other way around. Programme C is more like programme B, although here there are even fewer verbatim subtitles due to the presence of subtitles with sound transcription. Focusing on the two documentaries, the percentages for verbatim and edited subtitles are very close, as shown in Figure 14:

Figure 14. Distribution of verbatim vs. edited subtitles in programmes A and B
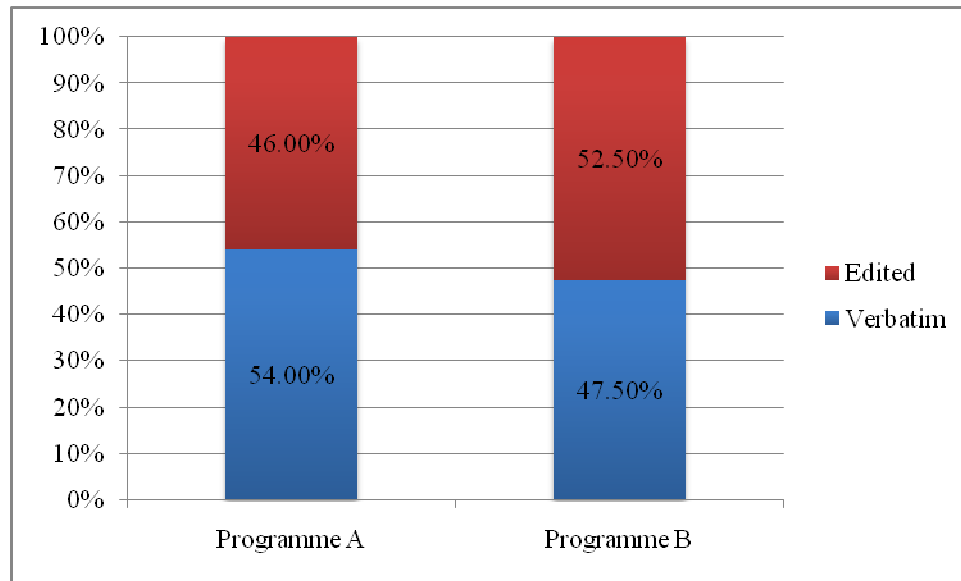
These percentages might seem surprising, as one might expect programme B to contain more verbatim subtitles, given that it has the lowest reduction rate/highest subtitling rate of the two (see Table 35). The fact that this assumption is not actually confirmed by the data makes us aware of the following:

i) As mentioned in section 5.6, reduction rates as calculated here (i.e. by comparing the word count of the original with the subtitle word count) are only indicative of the degree of verbatim transcription that has been used. That is because this metric does not take into account alterations, such as paraphrases, generalisations, etc. As a result, a low reduction rate/high subtitling rate does not necessary equal a high rate of verbatim transcription.

ii) Even in the ideal case where a high subtitling rate equals a high rate of verbatim transcription, a distinction needs to be drawn between verbatim subtitling at the level of the programme and verbatim subtitling at the level of the subtitle. For instance, a programme might be more verbatim on the whole, but contain more edited subtitles. This is possible because quite often individual subtitles are categorised as being edited on account of minor alterations; consider for instance the following subtitle:

Transcript: a life of foraging <u>high</u> in the mountains, where water was a long way
                off.

Subtitle A86:   A life of foraging in the mountains,
                where water was a long way off.

On account of the deletion of *high*, this subtitle is classified as edited. Yet, overall it is very close to verbatim status.

Looking at the high figures for verbatim subtitles in Table 37, it would be reasonable to assume that verbatim transcription can take us a long way when subtitling for D/deaf and HoH viewers. However, taking a closer look at the verbatim subtitles in the present corpus reveals that many of them are problematic, as they either do not respect time constraints (see section 5.3) or contain bad line breaks (see Chapter 6) or indeed present both problems at the same time (see Tables 38, 39 and 40). This is particularly true when the maximum subtitle rate is set to 120 wpm (Table 38), where almost half of the verbatim subtitles in each programme are displayed at a higher rate and/or include bad line breaks. Things improve greatly when the maximum rate is set to 130 wpm (see Table 39) or 140 wpm (see Table 40), although this is not the case for programme A, where approximately 38% of verbatim subtitles are still problematic.

Table 38. Good verbatim subtitles (rate = 120 wpm)

	Programme A				Programme B				Programme C			
	A1	A2	A3	**All**	B1	B2	B3	**All**	C1	C2	C3	**All**
Verbatim (total)	23	17	25	**65**	18	20	19	**57**	15	16	14	**45**
Not respecting time	8	3	9	**20**	6	6	11	**23**	7	9	4	**20**
Inc. bad line breaks	6	8	8	**22**	3	4	3	**10**	4	2	1	**7**
Total of good verbatim subtitles	11	8	13	**32**	10	11	7	**28**	7	7	10	**24**

Table 39. Good verbatim subtitles (rate = 130 wpm)

	Programme A				Programme B				Programme C			
	A1	A2	A3	**All**	B1	B2	B3	**All**	C1	C2	C3	**All**
Verbatim (total)	23	17	25	**65**	18	20	19	**57**	15	16	14	**45**
Not respecting time	5	2	4	**11**	3	0	3	**6**	2	4	1	**7**
Inc. bad line breaks	6	8	8	**22**	3	4	3	**10**	4	2	1	**7**
Total of good verbatim subtitles	13	9	15	**37**	13	16	13	**42**	10	11	12	**33**

Table 40. Good verbatim subtitles (rate = 140 wpm)

	Programme A				Programme B				Programme C			
	A1	A2	A3	**All**	B1	B2	B3	**All**	C1	C2	C3	**All**
Verbatim (total)	23	17	25	**65**	18	20	19	**57**	15	16	14	**45**
Not respecting time	2	0	1	**3**	0	0	0	**0**	7	2	1	**10**
Inc. bad line breaks	6	8	8	**22**	3	4	3	**10**	4	2	1	**7**
Total of good verbatim subtitles	15	9	16	**40**	15	16	16	**47**	11	13	12	**36**

5.7.1 Distribution of editing strategies

Table 41 shows the distribution of editing strategies in the three programmes analysed.

Table 41. Distribution of editing strategies

	Programme A				Programme B				Programme C			
	A1	A2	A3	**All**	B1	B2	B3	**All**	C1	C2	C3	**All**
Deletion	20	47	15	**82**	26	32	29	**87**	19	30	21	**70**
Condensation	10	10	8	**28**	15	16	10	**41**	6	9	8	**23**
Paraphrase	4	2	2	**8**	4	1	0	**5**	3	0	2	**5**
Specification	3	0	0	**3**	1	1	4	**6**	1	1	0	**2**
Addition	1	0	0	**1**	3	1	3	**7**	0	2	4	**6**
Transposition	0	0	0	**0**	1	0	1	**2**	0	1	1	**2**
Generalisation	0	0	0	**0**	0	0	0	**0**	0	0	0	**0**

The figures presented in this table suggest why reduction has come to be regarded as one of the defining features of subtitling. Indeed, in all three programmes, deletion is by far the most common among the editing strategies with condensation next. The rest of the strategies (paraphrase, addition, specification and transposition) are only sporadically used, while there are no instances of generalisation. This would seem to indicate a tendency to either use verbatim transcription (see Table 37) or edit subtitles in a way that remains as close to the original as possible (this would explain why deletion has been used more extensively than condensation).

It is also important to point out that as mentioned in section 4.3.1.1, with the exception of verbatim subtitling, all other strategies can and are often combined within the same subtitle. Not only that, but quite often the same strategy is used more than once in the same subtitle this being particularly true for deletions. For instance, consider the following example:

Transcript: I hope ~~some day that~~, ~~that when~~ we look back on this time we're, ~~we're~~ keeping chimpanzees in captivity ~~against their will~~

Subtitle B7:   I hope we look back on this time
              where we keep chimpanzees captive

In this single subtitle, there are six different instances of deletion (shown in strikethrough): two deletions of repetition/false start (second *that*, *we're*), one deletion of error (*when)*, one deletion of a subordinating conjunction (*that*) and two deletions of additional information (*some day, against their will*). This brings us to the issue of how to count deletions. In contrast to section 5.6, where deletions were counted in terms of words (by subtracting the sum of the words included in the subtitles from the sum of the words included in the transcript), this time deletions were counted so that deleted words belonging together only qualified as one deletion. For instance, in the previous example, the phrase *against their will* counts as one rather than three deletions.

In addition, although no examples of generalisation were found in the corpus, it was decided to keep it in the taxonomy, as not only it seems theoretically sound, but it is also a common strategy in interlingual subtitling (see for instance Lomheim, 1999 and Gottlieb, 2005b). More research would be required to determine whether the particular strategy is actually relevant to intralingual subtitling or whether it is exclusive to interlingual subtitling.

5.8 Nature of deletions

Table 42 presents the deletions in *Atlantis Uncovered* and *Chimps on Death Row* according to the categories suggested in section 4.3.1.10 (the full list of deletions including the category each was assigned is provided in Appendices 5 and 6).

Table 42. Deletion categories in the two documentaries

	Programme A	Programme B	Total
Additional information (adverbs, prepositional phrases, adjectives)	19	11	30 (17.75%)
Conjunctions/conjuncts	15/3	17/2	**37** (21.89%)
Disjuncts	1	4	5 (2.96%)
Less relevant information/additional comment when there is lots of text	2	8	10 (5.92%)
Grammatically redundant items/redundant information	10	15	25 (14.79%)
Emphatic repetition/other emphasis	7	1	8 (4.73%)
Item expressing similar meaning to adjoining item	4	2	6 (3.55%)
Quantifiers/numerals	4	2	6 (3.55%)
Intensifiers and exclusive adverbs (focus adjuncts)	2	6	8 (4.73%)
Hedges	1	1	2 (1.19%)
Oral:			
False start/repetition/hesitation	4	8	12 (7.1%)
Exclamation	4	0	4 (2.37%)
Discourse markers	2	3	5 (2.96%)
Overlapping speech	4	0	4 (2.37%)
Clause to elicit support	0	1	1 (0.59%)
Deictic expression/gestural language	0	1	1 (0.59%)
Error	0	5	5 (2.96%)
Total	82	87	169

According to the figures recorded, the following may be suggested:

1.   The linguistic items most likely to be omitted are conjunctions and conjuncts (37 instances) with adjuncts (adverbs, prepositional phrases) and adjectives following closely (30 instances).

2.   Despite the recommendation of the BBC Subtitling Guide to avoid editing items such as *but* and *so* (see section 3.3.5), the figures for conjunctions/conjuncts are consistently high in both documentaries (18 and 19 deletions respectively). This would seem to support De Linde and Kay's concern regarding the effects the deletion of such items might have on cohesion (1999, p 30). A separate study would be needed to further investigate this issue. What we need to point out, however, is that almost a third of the deleted conjunctions (10/32) in the two documentaries have been omitted because they have become redundant on account of other deletions, e.g.

Transcript: I could see Nakbe on the horizon ~~and~~ I knew that it needed to be investigated because of the implications that it had for the development of civilisation.

Subtitle A13: I knew Nakbe had to be investigated
                because of its implications
                for the development of civilisation.

Transcript: to a great ~~and~~ magnificent ancient civilisation

Subtitle A15: to a magnificent
                ancient civilisation?

In both examples, *and* is no longer required, since both *I could see Nakbe on the horizon* (subtitle A13) and *great* (subtitle A15) have been deleted.

3.   The third most common category of deleted items includes grammatically redundant items and items conveying redundant information. This is a rather diverse category, as illustrated by the following examples:

- Grammatically redundant items:
i) Subject deletion:

Subtitle B104: Booee escaped AIDS research,
              but ~~he~~ was infected with hepatitis,

ii) Deletion of relative pronoun and verb preceding past participle:

Subtitle A56: and a plaster floor ~~that was put~~ over that.

iii) Other examples, such as *lead us ~~down~~ to places* (subtitle A80) or *I felt ~~like~~ it was necessary to get to Nakbe and understand that* (subtitle A16).

In all examples, the words in strikethrough (*he*, *that was put, down*, *like*) can be deleted without affecting the structure or meaning of the sentence.

- Redundant information: This subcategory essentially covers all instances where the deleted items do not really add anything to the meaning of the sentence or can be inferred. Examples include:

Subtitle B4:   than the use of any other ~~non-human~~ animal
                for research.

Subtitle B30:   The chimpanzee will reach
                the honourable age of 50 ~~years~~ in its life

Subtitle B54:   ~~As it turns out,~~ The only animal that's known
                that gets infected with HIV 1

Subtitle B61:   We had to consider ~~the fact~~ that it may

lay dormant for years before

causing infection in the chimp.

4.   As mentioned in section 4.3.1.10 and can be seen in Table 42, there are two types of repetition: one where repetition is classified along with hesitations and false starts and another where it is used in order to create emphasis.

- Repetition as hesitation:

Subtitle B87:   To take these valuable animals and

throw them out in a field <u>in</u> ~~,in~~ Texas?! [sic]

- Emphatic repetition:

Subtitle A41:   This was part of a <u>long tradition</u>

<u>of </u>architectural development

and ~~a long tradition of~~ intense religious belief.

Subtitle B100: Look at him, he's laughing now.

<u>You see</u> ~~, you see~~?

In addition to these two categories, a third related category was established to cover instances where although repetition as such is not involved, the item (word, phrase or sentence) deleted partly reproduces or expresses a similar meaning to that of an adjoining item. Consider the following examples:

Subtitle A29:   He went out to some horrible place

and discovered some <u>ancient kingdom</u> ~~or lost civilisation~~.

Subtitle 108:   ~~The National Socialists in Germany,~~ <u>the Nazis</u>

had the notion of Aryan supremacy

Subtitle B83: They are like human beings kept

in a hospital ~~, in a clinic~~ as they get older.

It is important to note that repetition is the reason why it is possible to have the same element deleted twice in the same subtitle and categorise the deletion differently in each case. This is illustrated by the following example:

Transcript: I hope some day ~~that, that~~ when we look back on this time we're, we're keeping chimpanzees in captivity against their will

Subtitle B7: I hope we look back on this time

where we keep chimpanzees captive

In this example, the subordinating conjunction *that* has been deleted twice. In the first case, it has been treated as a conjunction, whereas in the second case as an instance of repetition-hesitation. Likewise, in subtitle B100 mentioned earlier, only the second occurrence of *you see* has been deleted; this is thus an instance of repetition rather than a discourse marker deleted.

5.   As seen in Table 42, a further "oral features" category was introduced to cover instances containing errors in the original. Five such instances were recorded (all in *Chimps on Death Row*) and they all come from parts of the programme where someone else other than the narrator is speaking. For instance:

Subtitle B22: which are a plague or a pest of ~~the~~ human kind

Transcript: ~~His named~~ is called Koo.
Subtitle B70: He is called Koon. [sic]

6.   According to the figures in Table 42, only 31 out of the 169 deletions recorded (approximately 18%) involve the omission of an oral feature. Given that

previous interlingual subtitling studies have reported that oral features are the usual candidates for deletion, this percentage seems rather low. Although this might be due to the difference between interlingual and intralingual subtitling, it might also be attributed to a difference in genres, as past studies, unlike the present research, have focused on the analysis of films/series (see sections 3.2, 3.3.5 and 4.1.2). These hypotheses, however, would need to be tested by using the categories suggested here to analyse a larger corpus including documentaries and films/series, as well as other genres from both intralingual/interlingual SDH and interlingual subtitling for hearing viewers.

7.   It is also important to note that all instances of deletion come from 94 subtitles (46 in *Atlantis Uncovered* and 48 from *Chimps on Death Row*). Of these subtitles, the majority come from parts of the programme with interviews or conversations, as illustrated in Figure 15.

Figure 15. Subtitles including deletions and part of the programme they come from
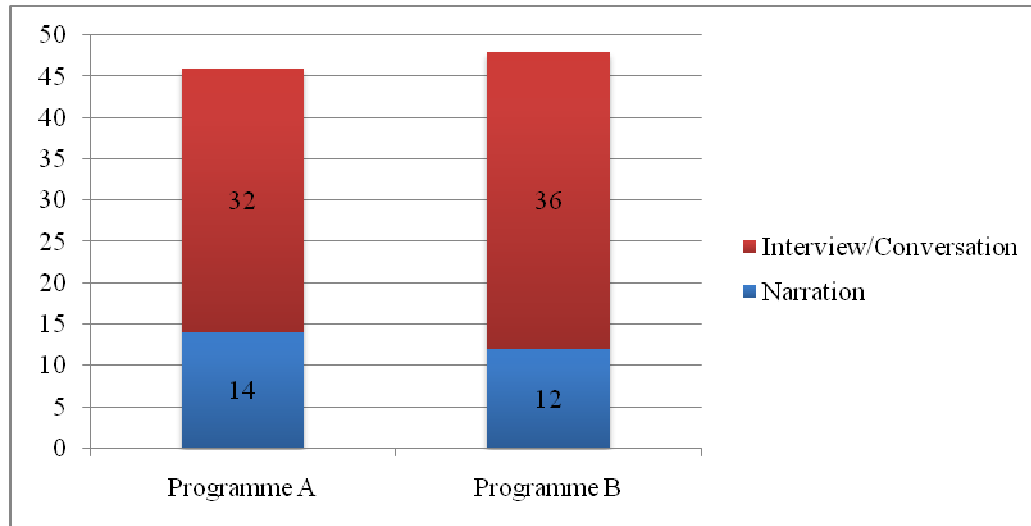
This would seem to suggest among other things that genres where narration is not predominant or even present might be expected not only to contain a higher number of deletions of oral features (see point 6 above), but also a higher number of deletions altogether. This hypothesis would again need to be tested.

5.9 Conclusion

This chapter has presented and discussed some preliminary findings on the basis of the present corpus.[39] As pointed out in section 4.2.2, this research was not meant to be large-scale and the findings and conclusions reported here are only tentative and need to be confirmed by compiling and analysing a considerably larger corpus including more genres and programmes. Further research would also confirm the value of the characteristics and metrics used here. This would indeed be most important, as one of the main goals of the present thesis was to provide a methodology for large-scale research dedicated to identifying the characteristics of current subtitle output.

What follows recapitulates the major issues raised and discussed in this chapter:

- The *need to have access to the original video* including information about among other things speech rates and the complexity of visual information on the screen repeatedly emerged while analysing the three programmes.

- When measuring characteristics, such as subtitle rate, *information about the mean should be complemented by other information, including standard deviation and range*. Indeed, although the mean is a very useful metric, it only provides a general idea of the subtitle rate adopted in a programme and may or may not be representative of the subtitle rates adopted by the programme's individual subtitles. Regarding the present research, in all three programmes analysed, the mean subtitle rate was rather slow (ranging from 118.2 to 113.1 wpm). Nevertheless, standard deviation was quite large suggesting that many values were far from the mean (many of these values were also later found to be above the mean, see section 5.3).

- Such variation within each programme should make us among other things *wary of arriving at conclusions and generalising on the basis of a limited number of*

---

[39] The only issue that has not been discussed so far is the issue of segmentation within subtitles. This will be the topic of chapter 6.

subtitles. In order to generalise, one would need to analyse the entire programme or at least a sufficiently large part of the programme and then replicate the process in a corpus consisting of a large number of programmes. Criteria for selecting extracts, if at all, need to be investigated.

• Subtitle rate is one of the most common characteristics measured in subtitling. The analysis in section 5.1 and in particular the figures in Tables 19 and 26, however, underscore the *need to agree on a suitable metric for subtitle rate*. That is, while the figures in Table 18 were calculated on the basis of WinCAPS, the figures in Table 25 were calculated on the basis of Equation 1; the figures in the two tables differ considerably to the extent that while according to Equation 1, programmes A and C display the exact same average subtitle rate (118.8 wpm), according to WinCAPS, programme A has an average rate of 116.1 wpm and programme C an average of 113.1 wpm.[40]

Although the formula used by WinCAPS is not known and the reasoning behind Equation 1 seems sensible (see section 4.3.1.5), the main analysis for the purposes of this research was carried out on the basis of the figures provided by WinCAPS, as in subtitling the calculation of the subtitle rate is usually left to the subtitling software and no-one would expect a professional subtitler to calculate the rate using Equation 1 or any other equation for that matter. For this reason, the figures from Equation 1 were only used so that they could be compared with the relevant findings reported by De Linde and Kay (1999). Apparently, further research is required in order to identify a metric that is suitable, commonly available and natural to subtitlers in order to describe this commonly used characteristic.

• It would appear that the two programmes where narration parts are predominant (*Atlantis Uncovered* and *The World's Most Wanted*) are those with the lowest

---

[40] In addition, when the subtitle rate was calculated by dividing the total number of words in the subtitles by the total number of minutes, as has sometimes been done in previous research, a different picture emerged. Mean subtitle rates were respectively a mere 93 wpm for programme A, 109 wpm for programme B and 108 wpm for programme C. This metric was not however adopted as it does not take into account times when no verbal information is transmitted (see section 4.3.1.5).

mean subtitle rate (see Tables 20, 22 and 19). This would suggest that i) a narrator's speech rate is generally slower than the speech rate of a person interviewed or people in a conversation, and ii) accordingly, subtitles in narration parts will exhibit slower rates than subtitles in parts with interviews or conversations. Of the two assumptions, only the latter was partially tested due to the lack of access to the video of the three programmes. Relevant results can be summarised as follows:

- There is no evidence supporting the hypothesis that narration parts are slower than interview parts. Across the three programmes, both narration and interview parts exhibit quite similar percentages of subtitle rates (with the majority of subtitles displayed at over 120 wpm) and not only that, but subtitles over 130 wpm tend to be more numerous in narration than interview parts.

- Interviews were distinguished from public addresses, as the overwhelming majority of subtitles from public addresses (85.71%) are displayed at below 120 wpm.

- Figures recorded seem to suggest that the hypothesis that narration parts are slower than conversation parts might be false. In the only instance of conversation in the three programmes (part A2), the average subtitle rate is 88 wpm with all subtitles displayed at a rate of below 100 wpm. This low rate seems more likely to be due to the deletion of a high number of oral features (see section 5.6) rather than a decrease in speech rates, however, this hypothesis would need to be confirmed by checking the actual speech rates from the programme.

- There also appears to be a tendency for subtitles containing extralinguistic information to be extremely slow. In *The World's Most Wanted*, which is the only programme in the corpus to include such subtitles, the average rate for these subtitles is a mere 34.7 wpm and the maximum is only 50 wpm.

- Unlike what Lomheim seems to suggest (1999), *display time alone cannot be used as a metric to categorise programmes in terms of how fast viewers have to read and in terms of how much they have to read.* His suggestion is based on an idealised view of the relationship between display time and amount of text (measured in characters), which is not to be taken for granted, as quite often, subtitles are not timed according to the number of characters they include (for other factors influencing the duration of a subtitle, see sections 2.4.3 and 3.3.1).

- Determining whether time constraints are respected or not seems to be the least objective of the metrics suggested, as i) there is no generally accepted maximum subtitle rate for D/deaf and HoH viewers, and ii) the optimum display time for each subtitle was calculated solely on the basis of the amount of text included in the subtitle (ignoring factors such as speech rates, etc.). However, this metric was considered potentially useful and was further explored by setting the subtitle rate to a maximum of 120, 130 and 140 wpm. Results can be summarised as follows:

  - Interestingly, although all three programmes have a mean subtitle rate of less than 120 wpm, they were all found to contain a high number of subtitles not respecting time constraints. This was particularly true when the maximum rate was set to 120 wpm with percentages of subtitles flouting constraints ranging from 50% to 42%.

  - Whereas at 120 wpm, it is programme B (the fastest programme) that contains more subtitles not respecting time constraints, when the subtitle rate is set to 130 and 140 wpm, it is respectively programmes A and C that contain more subtitles flouting constraints. However paradoxical this might seem, it might be explained if we consider that for both programmes A and C standard deviation and maximum value are higher than in programme B and as a result, more subtitles in these two programmes exhibit subtitle rates that are far from the mean.

- The majority of subtitles not respecting time constraints were found to be edited rather than verbatim. This seems somewhat strange, as one might expect the exact opposite to happen. It could be seen as an indication that while editing has been used, it has not been applied sufficiently and further reduction should have been called for and/or that spotting was not adequate.

- The majority of subtitles in the corpus appear on two lines. Regarding three-line subtitles, while both *Atlantis Uncovered* and *Chimps on Death Row* seem to use them rather liberally, *The World's Most Wanted* completely avoids them, even though it is the only programme to include extralinguistic information. In addition, when analysed more closely, the majority of three-line subtitles (72% and 63% respectively) were found not to respect time constraints when the maximum subtitle rate is set to 120 wpm. When the subtitle rate is increased, three-liners still remain a problem for only one of the programmes (although to a lesser extent).

- Overall, the extent of reduction in the three programmes is quite similar hovering around 13% or 3 words per subtitle. Again, variation within each programme is evident, as figures differ from part to part.

  Regarding documentaries, the mean number of words deleted per subtitle for *Atlantis Uncovered* and *Chimps on Death Row* (3.75) is somewhat higher than the figure reported by De Linde and Kay for the two documentaries in their corpus (2.62 words per subtitle) (1999). The two figures, nevertheless, are much closer than the figures reported by De Linde and Kay regarding the rest of the genres in their analysis (ibid).

  On the other hand, the mean reduction rate of 13% is considerably lower than the figures reported by researchers in the field of interlingual subtitling (ranging from 19% to 40%). However, these figures are not directly comparable given that:

i) in intralingual subtitling determining the number of words deleted is more of an objective measure than in the case of interlingual subtitling, given the difference in lexical structure between different languages, and

ii) the figures reported here are based on the analysis of programmes belonging to the informative genre, whereas the interlingual figures are mostly based on the analysis of films and television series.

- While the majority of subtitles in programme A are verbatim (54%), in programmes B and C, the majority of subtitles have been edited (verbatim subtitles still represent a high percentage of the total number of subtitles at 47.5% and 37.5% respectively). However, many of the verbatim subtitles either do not respect time constraints or contain bad line breaks or present both problems at the same time. This is particularly true when the maximum subtitle rate is set to 120 wpm (almost 50% of verbatim subtitles in each programme are displayed at a higher rate and/or include bad line breaks). Things improve greatly when the maximum rate is set to 130 or 140 wpm, although again a minimum of 27% and 18% of verbatim subtitles respectively still present problems.

- The distribution of editing strategies in the three programmes seems to support the prominence subtitling literature has given to reduction. Indeed, deletion is by far the most common among the editing strategies with condensation next, while the rest of the strategies are only sporadically used if at all. Overall, there appears to be a tendency to edit subtitles in a way that remains as close to the original as possible which would also be in keeping with the tendency to use verbatim transcription described above.

- The linguistic items most likely to be omitted are conjunctions and conjuncts (37 instances out of a total of 169) with adjuncts (adverbs, prepositional phrases) and adjectives following closely (30 instances). Although the deletion of oral features has been prominent in many previous studies, only 18% of deletions in the present corpus were found to involve the omission of an oral feature. This

seems to be most likely due to the difference between interlingual and intralingual subtitling and/or a difference in genres, as past studies, unlike the present research, come from the field of interlingual subtitling for hearing viewers and have mostly focused on the analysis of films and TV series. In order to confirm these hypotheses, the categories suggested in this thesis would need to be applied to a larger corpus containing more documentaries and current affairs programmes, as well as other genres, including films and TV series from the areas of both intralingual/interlingual SDH and interlingual subtitling for hearing viewers.

6. Results & Rules for Line Segmentation

This chapter presents findings relevant to segmentation within subtitles. Findings reported are then compared to previous research, as well as related to existing guidelines (see section 3.3.4). Given that current guidelines are found lacking in certain respects, a set of tentative rules is suggested. These rules are envisaged to be helpful both to professional and trainee subtitlers, but could also be elaborated on and used for the creation of a relevant subtitling tool. All linguistic abbreviations used in this part of the thesis are listed in Appendices 4 and 7.

6.1 Findings from research corpus

As mentioned in sections 4.3.1.12 and 4.3.3, all errors in segmentation were recorded for each subtitle analysed and all words involved received a POS tag (the tagset used is available in Appendix 4). In each case, the maximum number of characters per programme was taken into consideration (36 for programmes A and B, 37 for programme C). Preliminary results are displayed in Figure 16.

Figure 16. Distribution of line-break errors
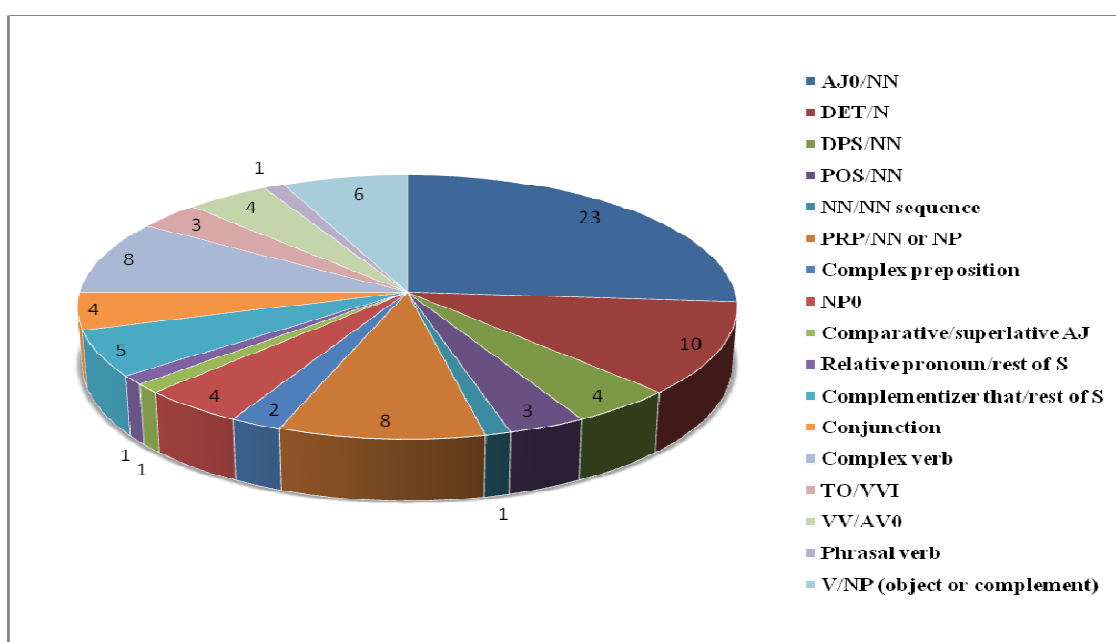
According to the figures shown, almost 30% of recorded line-break errors (23 out of a total of 88) involve separating an adjective from the noun it modifies. Other common errors involve separating a determiner from the noun (10 instances), breaking a preposition from the rest of the prepositional phrase or breaking complex verb forms (8 instances in each case). Each of these categories is exemplified by the following subtitles (including POS tags and the number of characters per line)[41]:

Subtitle A93: Then, they were growing different_AJ0   (33)

                 foods_NN to the Near Eastern farmers –   (35)

Subtitle B27: In this institute, we have the_AT0   (30)

                 policy_NN not to engage in experiments   (32)

Subtitle C10: If he thinks he can hide from_PRP   (29)

                 the_AT0 USA_NP0 or allies, he is mistaken.(34)

Subtitle A52: Gary and Linda Feinman are_VBB   (26)

                 uncovering_VVG a new site – El Palmillo. (36)

However, on examining all instances more closely, it soon became evident that the way some of the line-break errors had been distributed across the relevant categories in Figure 16 might not have been accurate. For instance, let us consider the following example:

Subtitle B8:   with the shame that I felt for my_DPS   (31)

               great-grandfather_NN as a slaveholder.   (35)

As indicated by the POS tags assigned, this was initially classed as an error involving a possessive determiner (*my*) and a noun (*great-grandfather*). However, this noun phrase (NP) is part of a prepositional phrase (PP), *for my great-grandfather*. This

---

[41] In this part of the thesis and in order to enhance readability, POS tags will be provided using the underscore rather than the SGML tags used in the annotated version of the corpus (see section 4.3.3 and Appendices 1, 2 and 3.

entails that even if it were possible to place *my* on the lower line, this would still leave us with a bad line-break involving the preposition *for* and the rest of the prepositional phrase (*my great-grandfather*). This seems to be the case with many line-break errors involving an adjective and noun or a determiner/possessive determiner and noun. In some cases, the problem is even more complex as illustrated by the following subtitle:

Subtitle A73:   to patterns in_PRP the_AT0 earliest_AJS   (27)

occupations_NN at Monte Alban.   (27)

Here, although it is possible to place the adjective on the lower line so that it is not separated from the modifying noun, doing so still leaves us with a bad line break, as *earliest occupations* is part of a NP containing a determiner (*the earliest occupations*) and this NP is in turn part of a PP (*in the earliest occupations*). This is clearly shown in the phrase-structure tree in Figure 17.

Figure 17. Phrase-structure tree for the PP in the earliest occupations
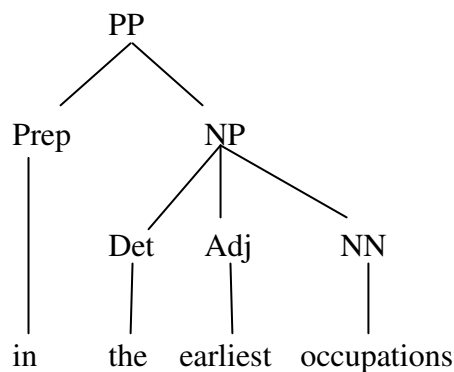
As a result, all line-break errors were classified again with results revealing a completely different picture (see Figure 18).

As shown in Figure 18, after the reclassification of line-break errors, almost 30% of the recorded instances (25 out of 90) now involve a PP, separating either the preposition from the NP or breaking the internal structure of the NP itself, i.e.

separating an article (AT0), possessive determiner (DPS) or genitive structure (POS) from the main noun. The second most common category again involves separating a determiner from the noun (although there are now 13 instances compared to 10 in Figure 16); likewise, the distribution of the remainder of the categories remains the same as before.

Figure 18. Distribution of line-break errors
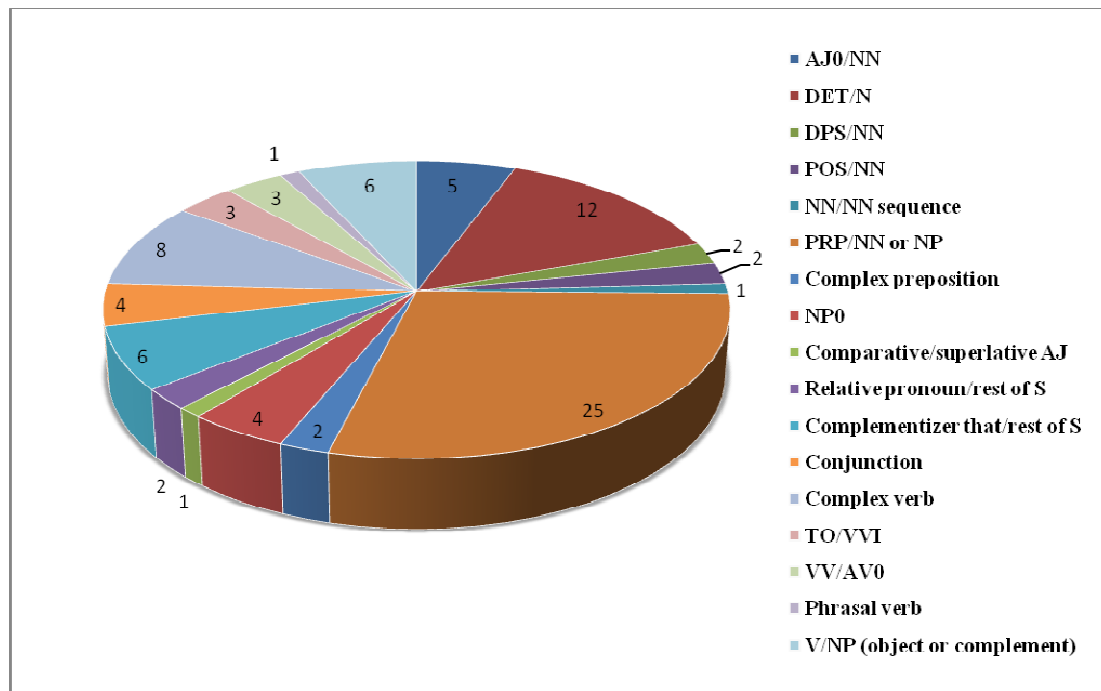
Among the 17 categories included in Figure 18, the last one (V/NP object or complement) is of particular interest, as it has only been rarely mentioned in the relevant literature and guidelines (the only examples are Díaz Cintas and Remael, 2007 and Neves, 2005b, see also section 3.3.4). Errors in this category arise when separating a verb from its direct object or in the case of copular verbs from its complement, as illustrated by the following examples:

Subtitle B21:   We're only allowed to_TO0 use_VVI   (25)

                chimpanzees_NN-obj to study diseases   (29)

Subtitle C75: So, the decision was made to leave_VVI   (34)

him_NN-obj there? Yes, for the time being. (35)

Subtitle A90: The first thing they grew wasn't_VBD   (32)

food__NN-comp. They grew bottle gourds, (30)

which they used to carry water.   (31)

What all the categories in Figure 18 have in common is that line-break errors falling under each category can be accounted for in terms of Karamitroglou's (1998) general guideline that segmentation should occur at the highest syntactic node possible (see section 3.3.4). However, what this guideline does not state is that sometimes segmentation even at the highest node will result in a poor line-break. Let us consider the following subtitle:

Subtitle C66: In the Islamic world, doors   (27)

were closing against Bin Laden.   (31)

In this example, both the NP (*doors*) and the VP (*were closing against Bin Laden*) have remained intact and segmentation has occurred at the highest syntactic node possible (as shown in Figure 19).

Figure 19. Phrase-structure tree for the sentence *doors were closing against Bin Laden*
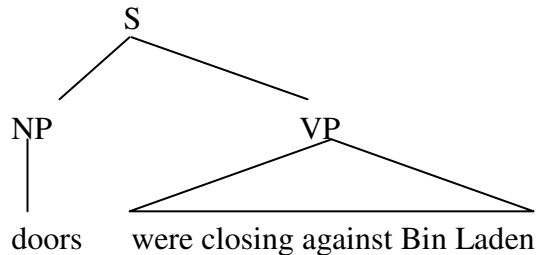
However, breaking the line at *doors* still results in poor segmentation. It would seem therefore that Karamitroglou's guideline is not applicable in the particular case, as the sentence should not be segmented at all. Therefore, in addition to the categories in Figure 18, we need to introduce a further category to accommodate cases where

the line-break error is due to separating the subject from the predicate. A rule specifying that the subject should not be separated from the predicate, however, is not applicable to all sentences. For instance, let us consider the following example:

| Subtitle C125: | Most of the bulk of the followers | (33) |
| | is outside Afghanistan. | (23) |

In this case as in subtitle C66, it is impossible to have the subject (*most of the bulk of the followers*) and predicate (*is outside Afghanistan*) on the same line so they have to be separated from each other. However, while subtitle C66 is an example of poor segmentation, subtitle C125 seems to contain no line-break errors. In fact, most of the relevant subtitles in the corpus are like subtitle C125. There are only five instances in which separating the subject from the predicate results in a poor line break. Apart from subtitle C66, there are only four more examples (subtitle A92 contains two instances):

Subtitle A92:	It was some 6,000 years after they	(34)
	began planting gourds that people	(33)
	here chose to settle in villages.	(33)

| Subtitle B58: | to compensate for the animals they | (34) |
| | thought would soon be dying of AIDS. | (33) |

| Subtitle B62: | For five or six years, people just | (34) |
| | thought we hadn't had enough time. | (34) |

What all these examples seem to have in common compared to cases where separating the subject from the predicate does not result in poor segmentation is that the subject NP is extremely short consisting of either a personal pronoun or a one-word noun or proper name. In other words, it would seem that when the subject is limited to only one word, breaking the sentence even at the highest node still creates a bad line break. When on the other hand, the subject is longer and so is the

predicate, separating them from one another does not seem to cause a problem. As a result, the relevant rule would be as follows:

Rule 1

"Do not break the sentence even at the highest level, i.e. do not separate the subject from the predicate, *when the subject NP consists of a personal pronoun or another one-word noun phrase.* If both the subject and predicate are long, then you can separate them without creating a poor line break".[42]

What is interesting about Rule 1 is what happens when the subject is short but the predicate too long to place them both on the same line. In that case, rule 1 comes in direct contradiction to the rule behind the last category in Figure 18, which specifies that a verb should not be separated from its object or complement. This results in partly flouting Rule 1 and completely flouting the object/complement rule. Consider for instance the following examples, which seem, nevertheless, to have been properly segmented:

| Subtitle A69: | Archaeologists have revealed | (28) |
| | the birth of a new civilisation - | (33) |

Subtitle B2:	Now scientists are facing	(25)
	the consequences of experimenting	(33)
	on our closest living relatives.	(32)

In both cases, the subject is rather short but the predicate is too long to place them both on the same line. What both subtitlers have opted to do is to break the predicate and place the subject and verb on the same line, while placing the object on the second line.

Thus, Rule 1 has to be modified as follows:

---

[42] It is important to note here that what is long and what is short in the particular rule seems to be a matter of contention (see relevant discussion in section 6.4).

Rule 1

a) Do not separate the subject from the predicate (i.e. do not separate the sentence) if the subject is short, that is if it only consists of a pronoun or any other one-word NP.

b) Do not separate the subject from the verb, but separate the verb from the complement/object if the subject is short but the predicate is long and both do not fit on the same line.

c) Break the sentence at the highest node and place the subject and predicate on separate lines, if both are long and do not fit on the same line.

This rule is of extreme importance, as this aspect has not been previously mentioned either in the literature or in relevant guidelines. In addition, while it does not to any extent invalidate the general recommendation by Karamitroglou to segment at the highest syntactic node possible, it does show (parts a and b) that this guideline also has its limitations. Figure 20 is a revision of Figure 18 modified to accommodate for instances where Rule 1 has been flouted.

Figure 19. Final distribution of line-break errors this time in percentages
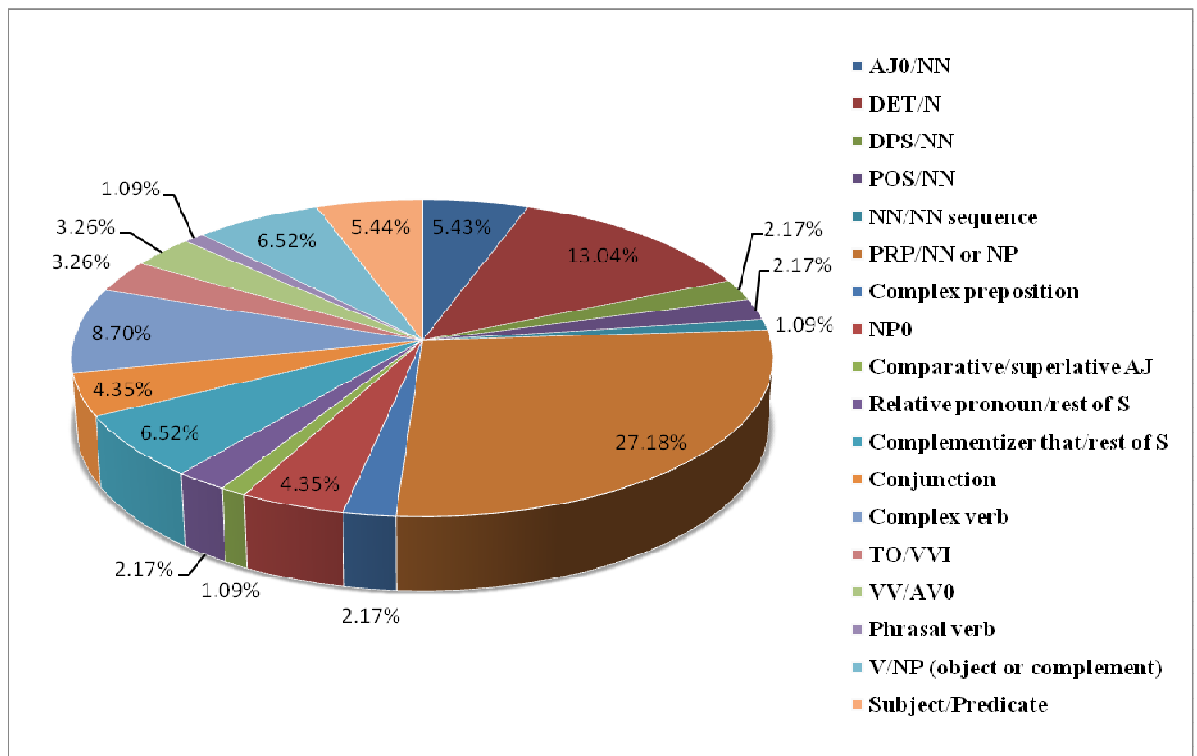
In all, there are 92 line-break errors in 89 subtitles out of the 360 subtitles analysed. This entails that almost a quarter of all subtitles analysed contain at least one line-break error. The distribution of the errors per programme is shown in Figure 21.

Figure 20. Distribution of line-break errors across the corpus
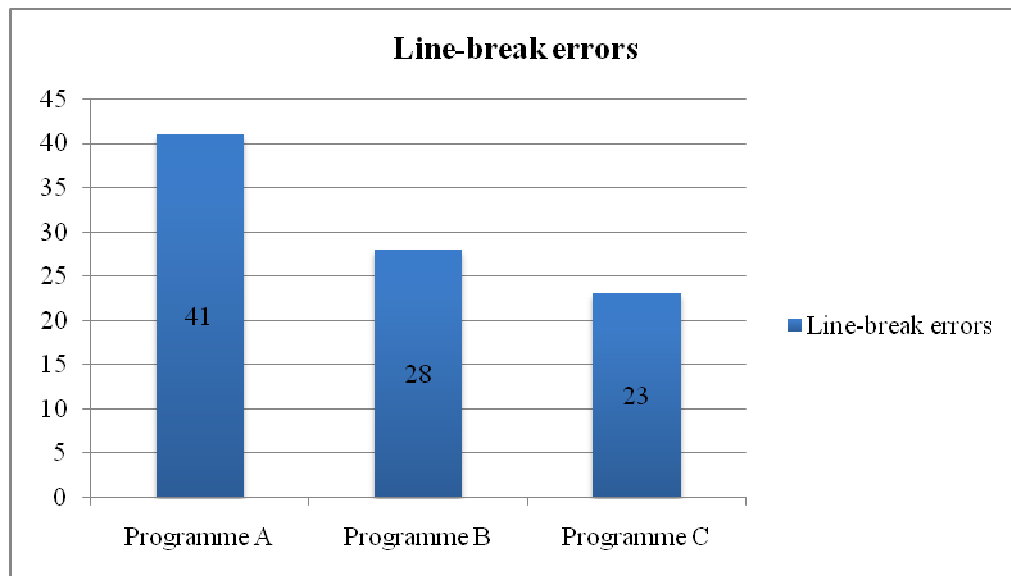
According to Figure 21, programme A is the most error-prone followed by programme B and then programme C. However, a rather different picture emerges if we measure the percentage of errors in the total number of actual line breaks per programme.[43] Thus, in a total of 130 line breaks for *Atlantis Uncovered*, 145 for *Chimps on Death Row* and 82 for *The World's Most Wanted*, 32%, 19% and 28% respectively have been poorly segmented (see Figure 22). This entails that while programme A is still the one with the highest number of errors, it is now followed by programme C rather than programme B. Interestingly, it also entails that i) the programme with the fewest line breaks (programme C) displays a high percentage of errors very close to the highest recorded, and ii) the programme with the highest number of line breaks (programme B) is actually the programme with the fewest errors. In all cases, however, it could be maintained that the three percentages

---

[43] Measuring the percentage of errors in the total number of actual line breaks is more objective than measuring it in the total number of subtitles, as on the one hand, it takes into account three-line subtitles (where there are two line breaks) and on the other, it excludes one-line subtitles which naturally do not contain any line breaks at all.

recorded, particularly those for programmes A and C, are rather high (see also Kyle, 1992 mentioned in sections 3.3.4 and 6.3).

Figure 21. Percentage of poor line breaks in the total number of actual line breaks
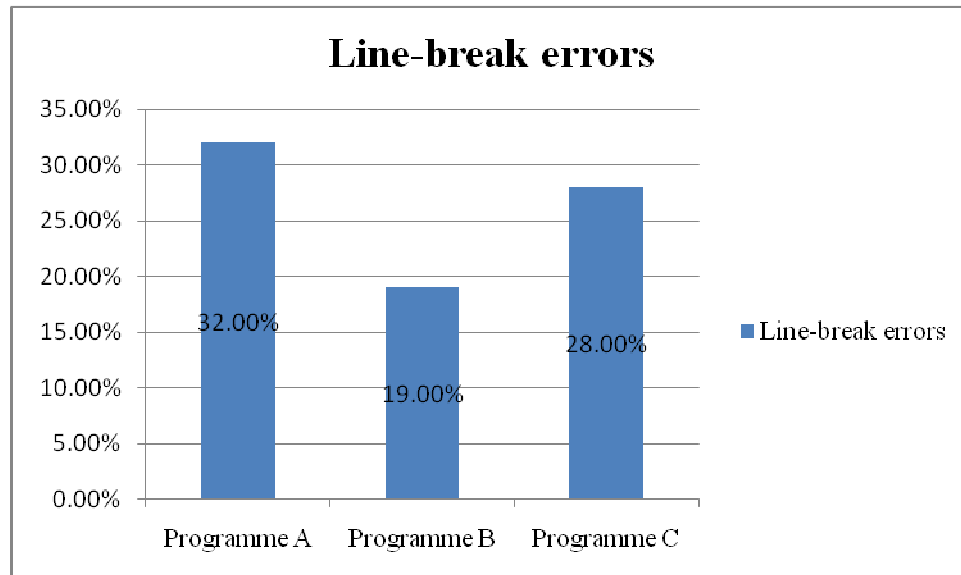
In order to further investigate the reasons behind this, all line-break errors were listed and an alternative subtitle containing no segmentation errors was suggested for each instance (see Appendices 8, 9 and 10). It is important to note here that since there was no access to the video of each programme, it was not possible to check whether any of the errors recorded were the by-product of the subtitler's effort to avoid covering important picture information by subtitle text. Neither was it possible to confirm whether any important on-screen information would be covered by lines in the alternative subtitles suggested. Despite these limitations, however, this turned out to be an informative exercise as will be discussed below.

All alternative subtitles in Appendices 8, 9 and 10 have been suggested exclusively on linguistic grounds and taking into account the maximum number of characters per subtitle in each programme and the display time available for the original subtitle. Matters of geometry (see section 3.3.4) have only been taken into account when not in conflict with linguistic considerations. Original subtitles along with their alternative versions come under four categories:

i) Original subtitles where all that is required to eradicate the error is to break the subtitle at a different point, as in the following examples:

Subtitle C89:	they had a concrete indication as_PRP21	(33)
	to_PRP22 who was ultimately responsible	(33)

Alternative:	they had a concrete indication	(30)
	as to who was ultimately responsible	(36)

Subtitle A103:	Science continues to be ignored	(31)
	by a public yearning for_PRP the_AT0	(28)
	romance_NN of a more mysterious past.	(34)

Alternative:	Science continues to be ignored	(31)
	by a public yearning for the romance	(36)
	of a more mysterious past.	(26)

ii) Original subtitles that only required minor editing, such as simplifying a complex verb tense, deleting a more or less redundant item or replacing a word with a shorter synonym[44] as seen in the examples below:

Subtitle B58:	to compensate for the animals they_Subj	(34)
	thought would soon be dying of AIDS.	(36)

Alternative:	to compensate for the animals	(29)
	they thought would soon die of AIDS.	(36)

---

[44] Replacing a word with a shorter synonym, however, might not be acceptable to viewers favouring verbatim subtitling.

Subtitle C83[45]:may be even rich or prosperous_AJ0   (30)

        people_NN and maybe highly educated.   (33)

Alternative: may be even rich or prosperous   (30)

        and maybe highly educated.   (33)

Subtitle C10:   If he thinks he can hide from_PRP   (29)

        the_AT0 USA_NP0 or allies, he is mistaken.(34)

Alternative: If he thinks he can hide   (24)

        from the USA or allies, he is wrong.   (36)

iii) Original subtitles where line-break errors can only be corrected after major editing, e.g.

Subtitle B59:   When we first put material   (26)

        in the animal, we were continuously   (35)

        looking for disease.   (20)

Alternative: When we first put   (17)

        the material in the animal,   (27)

        we kept looking for disease.   (28)

In some cases, even more radical changes are required, although this might mean that the solution might be less acceptable to viewers, particularly those in favour of verbatim subtitling (see also sections 2.4.6 and 3.3.4). For instance, let us consider the following instance which contains two line-break errors:

---

[45] The word *people* is here redundant as the previous subtitle (C82) reads: *Those people who believe in this*

Subtitle A91:	It was the_AT0 Americas' first-known	(32)
	plant domestication_NN - prompted by_PRP	(33)
	practicality_NN rather than hunger.	(32)

| Alternative: | It was the Americas' first-known | (32) |
| | plant domestication | (19) |

| | prompted not by hunger | (22) |
| | but by practicality. | (20)[46] |

This category also includes cases where errors can be corrected using deletion. Unlike what happens in category ii), however, the word deleted is not redundant. For instance:

| Subtitle C87: | As dawn broke over the_AT0 shattered_AJ0 | (32) |
| | skyline_NN of New York on 12 September, | (36) |

| Alternative: | As dawn broke over the skyline | (30) |
| | of New York on 12 September, | (28) |

iv) Original subtitles where an alternative error-free version does not seem to be available, e.g.

| Subtitle A77: | There's a simple reason why so many | (35) |
| | ancient peoples built pyramids. | (31) |

| Alternative: | None seems to be available |

---

[46] The alternative versions suggested including *the material* for subtitle B59 and *prompted not by hunger but by practicality* for subtitle A91 are both closer to the transcript than the subtitles broadcast (see Appendices 8 and 9).

On the basis of these four categories, Figure 23 presents the distribution of subtitles containing line-break errors in the three programmes. For the purposes of our discussion, we will treat simply changing lines and minor editing together and also consider major editing as equivalent to cases where no alternative seems available.

Figure 22. Breakdown of subtitles that contain line-break errors according to alternative versions
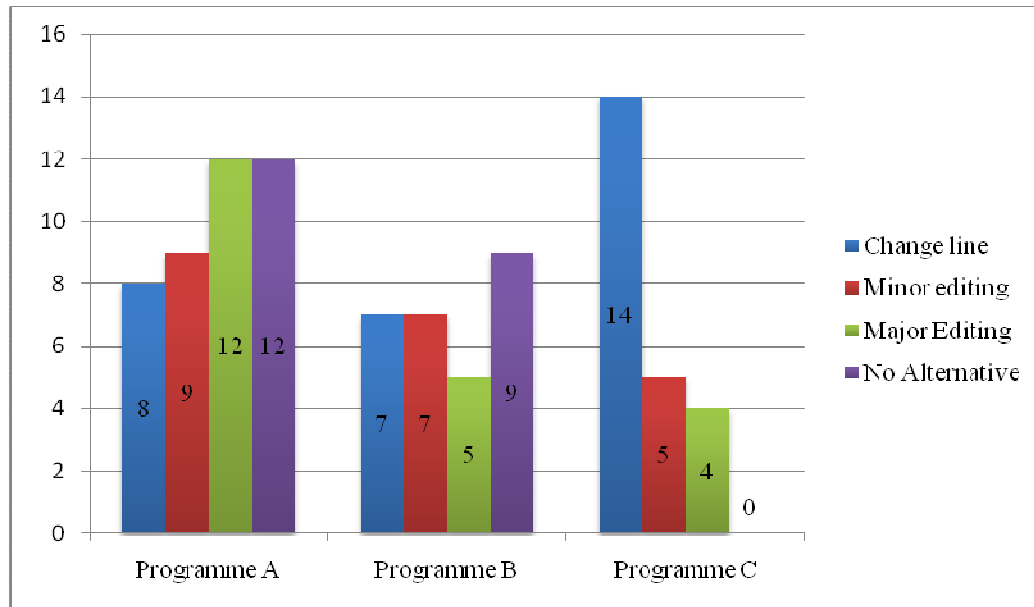
The data provided in Figures 21 and 22 reveals some quite interesting facts. First of all, in programme A, the majority of subtitles with poor segmentation (59%) require either major editing or cannot be edited at all. This might explain why the particular programme is by far the one with the highest number of line-break errors. Programme B, on the other hand, displays a more balanced image: 50% of subtitles with line-break errors can be edited by simply changing lines or making minor changes, while the other 50% do not have an error-free version or would require major changes. The most interesting figures are those related to programme C. Here, the overwhelming majority of subtitles (83%) may become error-free relatively easily and 61% merely involve changing lines. Only 17% of the total subtitles require some form of major editing and significantly, there are no subtitles for which no alternative can be suggested.

The above creates the impression that while subtitler C could have quite easily avoided line-break errors in the majority of cases, the subtitlers of programmes A and B could only have done so in relatively fewer cases, as to a great extent they were bound to create subtitles with errors. This raises the interesting question why the subtitlers, in particular subtitler C, did not choose an alternative error-free version in the many cases when this was relatively easy to achieve and also consistent with the BBC Subtitling Guide recommendations (see section 3.3.4). Examining the relevant instances more closely suggests the following relevant reasons:

• **Geometry**: As mentioned in section 3.3.4, segmentation is not only a linguistic matter, but also a matter of geometry, which is the need to have all lines of a subtitle as equal in length as possible (Karamitroglou, 1998). Although according to available guidelines (Baker et al., 1982; Karamitroglou, 1998; Neves, 2004), segmentation should be a compromise between linguistics and geometry with linguistic considerations taking precedence when the result is a bad line-break, no such recommendation is included in the BBC Subtitling Guide. As a result, in 17 of the total 50 line-break errors that might have been corrected easily, poor segmentation appears to be the result of the subtitlers' attempt to create symmetrical subtitles, as in the following examples[47]:

Subtitle B87: To take these valuable animals and   (34)
               throw them <u>out</u> in a field in Texas?!   (36)

Alternative: To take these valuable animals   (30)
               and throw them in a field in Texas?!   (36)

---

[47] Examples in the following sections are provided without tags and using Arial (which seems to be the most recommended font, see e.g. Karamitroglou, 1998 and Díaz Cintas and Remael, 2007) in order to show the geometry more accurately.

Subtitle C76:   Bin Laden continued to taunt the   (32)

                 American President, Bill Clinton   (33)

Alternative: Bin Laden continued to taunt   (28)

                 the American President, Bill Clinton   (37)

Geometry is also a factor in a further 14 of the total line-break errors along with the third factor that will be presented below.

• **Shorter first line**: This aspect is essentially the opposite of geometry, as it refers to cases where it is not possible or desirable to create a subtitle with lines of equal or similar length. According to relevant recommendations, in such cases and when feasible, the top line should be shorter than lower lines in order to keep free as much of the image as possible (see for instance Ivarsson and Carroll, 1998b and Díaz Cintas and Remael, 2007). As with geometry, however, such aesthetic concerns should give way to linguistic considerations whenever necessary in order "to secure line-breaks which are appropriate syntactically, respecting and reflecting the logic of the sentence" (Díaz Cintas and Remael, 2007, p. 87). Again, this recommendation is not included in the BBC Subtitling Guide. Thus, 10 of the 50 errors, that might have been corrected easily either by changing lines or with minor editing, seem to have been due to the subtitlers' attempt to keep the first line shorter, as appropriate segmentation would have created the opposite result. Let us consider the following examples:

Subtitle A28:   Archaeology was conducted   (25)

                 largely by adventurers and rogues.   (34)

Alternative: Archaeology was conducted largely   (33)

                 by adventurers and rogues.   (26)

Subtitle C21:   It's ironic that in his personal   (32)

                life he is very soft, humble, smooth.   (37)

Alternative: It's ironic that in his personal life   (37)

                he is very soft, humble, smooth.   (32)

We need to point out that as mentioned above, since the present corpus does not include the video of the three programmes, we cannot confirm whether keeping the first line shorter in the particular subtitles was actually required in order to avoid covering important picture information by subtitle text. Although this is a possibility which makes the present discussion tentative, in seven instances (including all six in programme C), it does not seem to be a very likely reason, as the first line in each of these cases is already quite long with more than 30 characters.

In addition, it is important to mention that when this aspect clashes with geometry considerations, it appears that geometry is the factor that takes precedence, even in cases where a shorter first line would ensure proper segmentation. This is suggested by examples, such as the following:

Subtitle A53:   ..The second part underneath, but   (33)

                that also appears to be broken.   (31)

Alternative: ..The second part underneath,   (29)

                but that also appears to be broken.   (35)

No doubt, we would need to study other subtitles not containing any line-break errors to confirm whether this is really a tendency.

• **Preference for verbatim subtitling**: This factor refers specifically to subtitles including line-break errors that could have been corrected with minor editing (see category ii) and seems to apply to all recorded cases (21) either independently (7

cases) or in conjunction with considerations of geometry (14 cases), as illustrated by the following subtitles:

Subtitle B14:	They show characteristics like	(30)
	altruism, compassion and so forth.	(34)

Alternative:	They show characteristics	(25)
	like altruism and compassion.	(29)

Subtitle C117:	Now if they intend to destroy Bin	(33)
	Laden, they will have done nothing.	(34)

Alternative:	Now if they plan to destroy Bin Laden	(37)
	they will have done nothing.	(28)

• **Unknown**: Among the 50 instances that could be corrected easily there are also two subtitles where it is not clear why subtitlers chose this segmentation.

Subtitle B106:	His life at the Waystation depends	(34)
	on donations from the public.	(29)

Alternative:	His life at the Waystation	(26)
	depends on donations from the public.	(28)

Apart from separating the prepositional verb *depends on*, the original subtitle also ignores geometry and the recommendation to have a shorter first line, although in the particular case, all these considerations would result in a more readable, linguistically motivated line-break. The same is true for the following example:

Subtitle C99:   Further implications of Bin Laden's   (33)

                guilt had been found.   (21)

Alternative: Further implications   (20)

                of Bin Laden's guilt had been found.   (36)

All factors are shown in Figure 24.

Figure 23. Distribution of some potential reasons for line-break errors
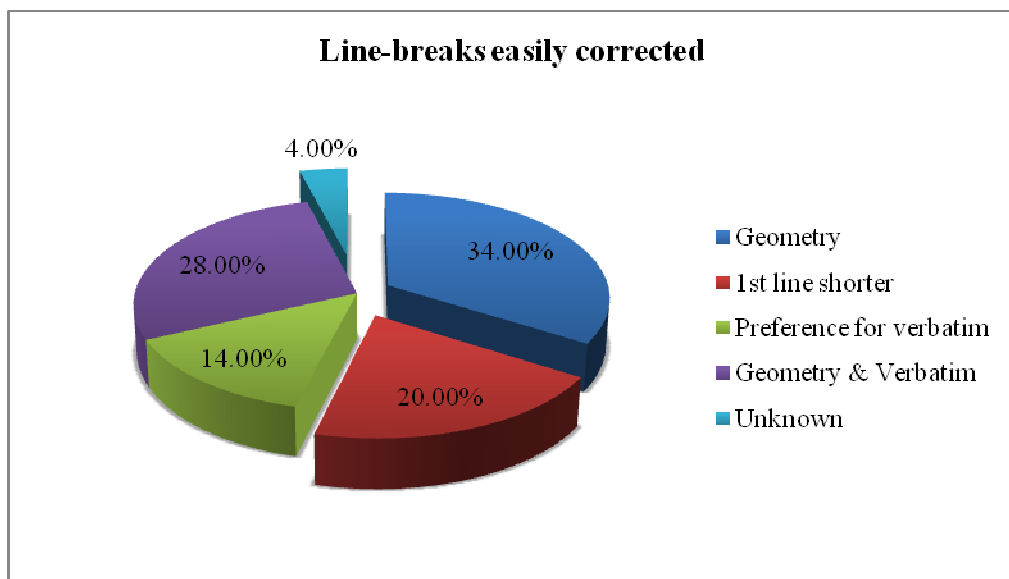
As can be seen in the figure, geometry is by far the most influential factor either independently or in association with the preference for verbatim transcription. In all, it would appear that the three factors identified all compete with linguistic considerations in the following order:

(1) Linguistics vs. Geometry

(2) Linguistics vs. Verbatim transcription

(3) Linguistics vs. Preference for a shorter first line

This, however, only partially answers the original question of why the subtitlers of the three programmes, in particular programme C, did not choose the alternative error-free versions even though this was possible for over half of the line-break

errors recorded. In fact, the original question becomes even more interesting given both the recommendations relevant to segmentation in the BBC Subtitling Guide (see section 3.3.4) and the current recommendations in other guidelines to favour linguistic considerations over both geometry and the preference for a shorter first line (as long as no important picture information is concealed). Possible answers include the following:

a) The subtitlers were not aware of the relevant recommendations in the BBC Subtitling Guide or the general recommendations available

b) The subtitlers were aware of the BBC Subtitling Guide recommendations and for them good editing also equals choosing verbatim subtitling over proper segmentation

c) The subtitlers were aware of the general recommendations regarding proper segmentation, but chose to ignore them because they are not part of the BBC Subtitling Guide or because they feel that geometry and the preference for a shorter first line should take precedence

d) The subtitlers were aware of the recommendations, but did not pay particular attention because they consider proper segmentation a mere detail particularly when they are under time pressure as usually happens (see also section 3.3.4)

Apart from the first reason, which it would seem reasonable to reject considering that line segmentation is not only one of the subtitling aspects always discussed in relevant guidelines, but is also mentioned in the BBC Subtitling Guide, the remaining reasons all appear plausible. However, they are only assumptions and as such would need to be confirmed by research involving professional subtitlers. Such research is all the more important, if we consider that reason c) suggests that even guidelines which do not seem controversial are not universal or always heeded. If that is true and given that the purpose of guidelines is to promote best practices and ensure high-quality subtitling, ways of promoting guidelines and their application to instances where they could be easily and with reason implemented might also be sought.

In addition, research should be extended to include viewers, in particular D/deaf and HoH people in the case of SDH. What do viewers value more and under what circumstances? Do they feel that verbatim transcription should always take precedence over sound segmentation even when minor editing is involved? What about geometry and the need to have a short first line even when no important picture information is obscured? More importantly, objective methods should be used to determine how and under what conditions these priorities contribute to viewers' understanding and enjoyment of the subtitles and the programme. Undoubtedly, such research is urgently required and projects that focus exclusively on aspects of segmentation such as that by Perego and Del Missier (2008)[48] are essential within the context of SDH.

6.2 Line-break errors and lack of consensus

As mentioned above, one of the possible reasons why the three subtitlers did not change at least some of the subtitles that might have been corrected either by simply changing lines or with minor editing might be that their priorities were different than those put forward by relevant guidelines. Naturally then, different subtitlers might be expected to have different priorities and make different decisions even when dealing with the same example. This impression seems to be in line with the results of an informal questionnaire that was designed for the purposes of the present research and distributed to participants at the conference *Audiovisual Translation: Multidisciplinary Approaches* held in Montpellier in June 2008. The questionnaire included 20 subtitles and any alternative versions available and asked participants to specify which version they preferred or in the case no alternatives were listed to

---

[48] In this paper, Perego and Del Missier presented preliminary results of an ongoing project, which aims at determining the effect of poor segmentation on viewers. The project has focused on hearing subjects watching foreign film-clips with Italian subtitles. Initial results suggest that unlike long-held beliefs, segmentation does not make a difference. The main criticisms of conference participants when the paper was presented were that i) only specific features of segmentation were examined and ii) results will most likely be totally different when studying D/deaf and HoH people. The present author in particular mentioned her assumption (that needs to be confirmed) that while advanced readers might just find poor segmentation annoying, if they notice it at all, its effects on less able readers, including many D/deaf and HoH people, are actually detrimental (as suggested by existing guidelines and previous studies, e.g. Neves, 2005).

indicate whether they were happy with the original subtitle as it was; in both cases they were also invited to suggest their own solutions if they wanted to. Only six questionnaires were returned, but results suggest the following:

1) Subtitlers have different priorities when judging segmentation and choosing between a poor line-break and any of the other factors. As a matter of fact, the six subtitlers that returned their questionnaires all chose the same error-free (in terms of linguistics) version in just 2 cases. The distribution of answers is shown in Figure 25.

Figure 24. Responses to segmentation questionnaire
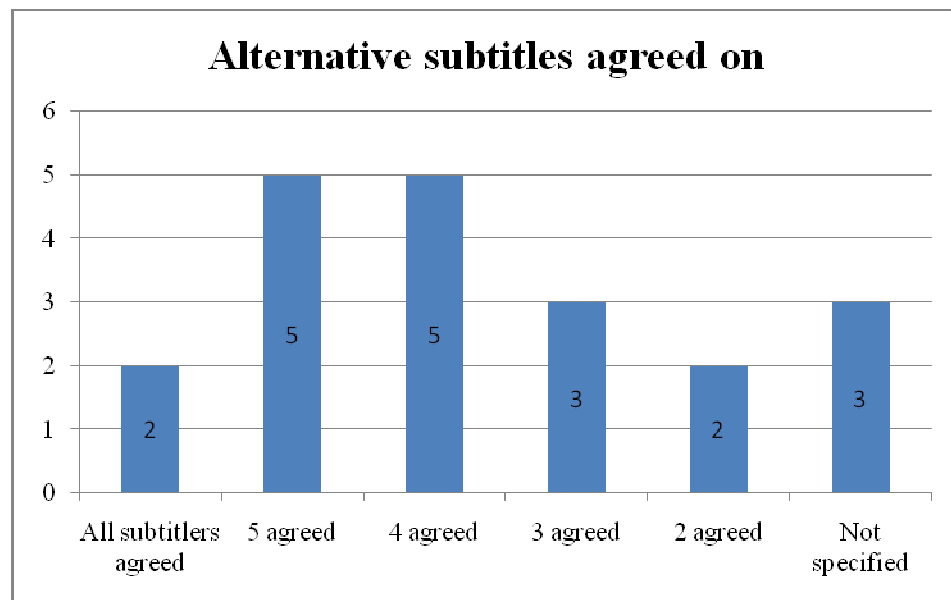
As seen in Figure 25 and excluding the three subtitles where at least one of the respondents did not provide an answer, the alternative error-free subtitles were chosen by at least 4 subtitlers in 12 cases out of 17. However, agreement was not always reached to the same extent. In addition, there were another three subtitles where respondents were divided between the original and the alternative version and two more where only two subtitlers chose the alternative, while one the original and the remainder three specified they were happy with neither.

2) It would seem that sometimes choosing a subtitle with poor segmentation might not be a matter of priorities, but might be a matter of not actually perceiving there is

anything wrong with the subtitle in the first place. This is partly suggested by replies to the first question in the questionnaire. Respondents were given the following subtitle

Subtitle A79: a sloping pile was the only
high structure you could build.

and were requested to specify if they were happy with it or not and why. Only three of the subtitlers provided an answer and of these, two mentioned they were happy and only one that they were not happy with the subtitle, but that it seemed to be the only solution. Although there could be other reasons, this might also be an indication that what is a poor line break for one person is not necessarily always for another. This assumption should nevertheless be tested by further research.

Further research would also be instrumental in testing one more hypothesis, namely that not all segmentation errors appear equally unacceptable. For instance, putting aside, to the extent it is possible, other considerations, it would seem that separating a conjunction from the phrase/sentence it introduces or an article from the noun in a simple NP consisting of just those two elements appears more serious an error than breaking an adjective from the noun it modifies in a complex NP such as that in subtitle A79 above. Unfortunately, although one of the reasons for undertaking the questionnaire was to test this particular hypothesis, no useful responses were obtained. However, if that assumption is found to be true to some extent, it would more than likely entail that at least in some cases, considerations such as geometry and the need to keep the first line short for merely aesthetic reasons might actually be given priority even if the result is a poor line break.

6.3 Relation to other research

As mentioned in section 3.3.4, not much empirical research has been devoted to segmentation. The only study that has so far identified the types of line-break errors

in programmes is the one by Kyle (1992). Analysing two two-minute extracts from two programmes including 22 and 20 line breaks respectively, he reported that only 77% and 45% of line breaks respectively were correct. Of the errors recorded, the majority (10% and 30% respectively) involved splitting articles or adjectives from their nouns, another 4% and 10% splitting prepositions from their appropriate phrase and another 4% and 10% splitting verbs or adverbs.[49] The main finding was consequently that "there is a great deal of improvement to be achieved" as far as line breaks are concerned (ibid, p. 90). This finding seems to be supported by the present research. Indeed, only 68% of the line breaks analysed in *Atlantis Uncovered*, 81% in *Chimps on Death Row* and 72% in *The World's Most Wanted* were found to have been segmented correctly. Regarding the most common categories of errors, the present data matches that of Kyle. The only difference regards the distribution of the categories, since as shown in Figure 20 the most common category in the present corpus involves separating a preposition from the rest of the PP. However, it is very likely that the figures provided by Kyle were calculated in the same way as the initial figures for the present corpus in Figure 16, that is without taking into account whether the actual NP is part of a larger PP.

Regarding the on-going research by Perego (2008), many of the categories identified in her research corpus (see section 3.3.4) have also been identified in the present corpus. The only categories Perego mentions for which no examples were recorded in the corpus presented in this thesis are nominals, titles preceding proper names and negative constructions. In addition, the only subordinating conjunction separated from the rest of the clause that has been recorded here as seen in Figure 20 is the complementizer *that*. On the other hand, the present corpus also contains some categories not encountered by Perego, namely separating complex prepositions (such as *in terms of*), verbs from their adverbs, relative pronouns from the rest of the clause, object/complement NPs from their verbs and subjects from their predicates. As mentioned in section 3.3.4, Perego's treatment of the various types of line-break errors that can occur is qualitative and no quantitative information is provided. Thus, findings cannot be compared any further. However, one thing that we can comment

---

[49] No information is provided on the rest of the errors.

on is Perego's belief that prepositional phrases are thought to be the easiest of the three intra-clausal categories to be properly segmented, since prepositions are short enough to be moved around conveniently (2008, p. 50). Although this might true for examples such as the ones provided by Perego (*della / barca*, *from / France*), it is not equally true for more complex PPs, such as the following:

Subtitle B91: that have value throughout_PRP their_DPS (32)

life_NN for the medical understanding (34)

of diseases of old age. (23)

Subtitle A73: to patterns in_PRP the_AT0 earliest_AJS (30)

occupations_NN at Monte Alban. (37)

In both cases, moving around the preposition so that it is on the same line as the head noun is not possible, as both NPs contain more than just a noun. In fact, out of the 25 PPs recorded in the corpus, 17 are of this type and at least for six there seems to be no alternative error-free version.

6.4 Linguistic Rules

This section presents the linguistic rules whose flouting led to the 92 errors recorded in the present corpus and shown in Figure 20. It also incorporates some other rules mentioned in the relevant literature and guidelines, but not encountered in the corpus. All rules have also been expressed formally following the pattern of Generalised Phrase Structure Grammar rules[50] in the belief that such a formal approach would be particularly helpful if the rules were to be incorporated in a relevant subtitling tool (see section 6.5, all relevant abbreviations are included in Appendices 4 and 7). Regarding notation, the plus sign is here used to indicate that the relevant constituents should not be separated, whereas the slash (/) is used to indicate that the constituent that follows it *should* be separated from the rest of the

---

[50] A comprehensive introduction to this syntactic theory is provided by Bennett, 1995.

phrase/sentence. Moreover, curly brackets ({}) indicate that we have to choose one of the enclosed categories, parentheses show an optional element and the Kleene star shows iteration.

- **Subject and Predicate**

1a) If the subject is short, that is if it only consists of a pronoun or any other one-word NP and the predicate is long → Do not separate the subject from the predicate, i.e. do not separate the sentence at all.

S→ NP + VP

when NP → PNP or any other case where the NP consists of one word

1b) If the subject is short but the predicate is long and both do not fit on the same line → Do not separate the subject from the verb, but separate the verb from the complement/object. If necessary, edit the subtitle.

S → NP-subj + Vtrans / NP-obj
S → NP-subj + Vcop / {NP, AP)

when NP → PNP or any other case where the NP consists of one word

Exception: If this is not desirable (e.g. in case of a collocation such as *commit murder*, the verb and object should not be separated), then edit the subtitle more and apply rule 1a.

1c) If both the subject and predicate are long and do not fit on the same line → Break the sentence at the highest syntactic node and place the subject and predicate on separate lines.

S → NP / VP

It is important to note here that what is short and what is long in the particular rule seems to be a matter of contention. Restricting ourselves to the subject, it seems that any subject consisting of at least up to 10 characters might be considered short. For instance, subtitle C66

In the Islamic World, doors
were closing against Bin Laden

was judged to have been poorly segmented by 5 of the 6 subtitlers that returned the questionnaire. On the other hand, subjects with more than 12 characters might be considered long; for instance, this might explain why five of the subtitlers considered the alternative suggested for subtitle A112 as acceptable

"These claims
to an ancestral heritage in Atlantis"

and four of the subtitlers rejected subtitle C38

"Osama Bin Laden was
one of thousands of young Saudis"

in favour of the alternative

"Osama Bin Laden
was one of thousands of young Saudis."

This seems to go beyond linguistic considerations alone and research including both subtitlers and viewers would be required to determine if that hypothesis is correct and the number of characters over which a subject/predicate might be considered to be long.

- **Verb and Object/Complement**

2) Do not separate a verb from its direct object or in the case of copular verbs from its complement.

VP → Vtrans + NP-obj

VP → Vcop + {NP, AP}

Exception: If the subject is short and does not fit on the same line as the VP, then use rule 1b.

- **NP rules**

3) Do not separate the determiner from the noun.

NP → AT0 (AP*) + NN

NP → DT0 (AP*) + NN

NP → DPS (AP*) + NN

4) Do not separate the adjective from the noun it modifies.

NP →{AT0, DT0, DPS} AJ0 + NN

5) Do not split proper names, i.e. do not separate the first name from the last name and do not separate the title from the name.

NP0 → NP012 + NP022

6) Do not separate nouns in a noun-noun sequence (including collocations).

NP → NN + NN

7) Do not separate a possessive/genitive form from the noun it modifies.

NP → NN POS + NN

- **AP rules**

8) Do not separate the adverb from the adjective it modifies.

AP → AV0 + AJ0

- **PP rules**

9) Do not separate the preposition from the rest of the prepositional phrase.

PP → PRP + NP
where NP → {NP0, PNP, NN}
PP → PRP + (XX0) + {VBG, VDG, VHG, VVG} (in case the preposition is followed by a verb and an optional negative)
PP → PRP + AV0

10) Do not split complex prepositions.

PRP → PRP12 + PRP22
PRP → PRP13 + PRPR23 + PRP33

- **VP rules**

11) Do not separate the adverb from the verb it modifies.

VP → V + AV0
VP → AV0 + V

12) Do not separate the infinitive marker from the infinitive.

VP → TO0 + V

13) Do not split complex verb forms.

VP → Vaux VP

Specific rules:

VP → {VBB, VBZ} + (XX0) + {VDG, VHG, VVG}   (present continuous)

VP → VBD + (XX0) + {VDG, VHG, VVG}   (past continuous)

VP → {VHB, VHZ} + (XX0) + {VBN, VDN, VHN, VVN} (present perfect)

VP → {VHB, VHZ} + (XX0) + VBN + {VDG, VHG, VVG} (present perfect continuous)

VP → VHD + (XX0) + {VBN, VDN, VHN, VVN} (past perfect)

VP → VHD + (XX0) + VBN + {VDG, VHG, VVG} (past perfect continuous)

VP → VM0 + (XX0) + {VBI, VDI, VHI, VVI) (future, future in the past and other modals)

VP → VM0 + (XX0) + VBI + {VDG, VHG, VVG} (future continuous)

VP→ VM0 + (XX0) + VHI + {VBN, VDN, VHN, VVN} (future perfect)

VP → VM0 + (XX0) + VHI + VBN + {VDG, VHG, VVG} (future perfect continuous)

VP → {VDD, VBD} + {VDI, VHI, VVI} (interrogative simple present and simple past)

14) Do not split phrasal verbs.

VP → V + AV
VP → V + PRP

Subordination

15) Do not separate the subordinating conjunction from the rest of the subordinate clause.[51]

S′ → CJT + S

16) Do not separate the relative pronoun/adverb from the rest of the relative clause.

S′ → PNQ + rest of clause
S′ → AVQ + S

17) If possible, in case there is one main clause and one subordinate clause and they do not both fit on the same line, use one line for each clause.

Subtitle → S / S′
or
Subtitle → S′ / S

- **Coordination**

18) Do not separate a coordinating conjunction from the phrase it introduces.[52]

NPplus → CJC + NP
AJplus → CJC + AP
VPplus → CJC + VP
Splus → CJC + S

---

[51] S′ denotes a dependent clause
[52] NPplus, APplus, VPplus, Splus are the categories that allow us to deal with coordination in the case of noun phrases, adjective phrases, verb phrases and entire sentences. Thus, the rule for "Mary and John" is NP → NP NPplus.

19) If possible, in case there are two main clauses and they do not both fit on the same line, use one line for each clause.

Subtitle →S / Splus

- **Other**

20) Do not split collocations.

21) Do not hyphenate words.

As can be seen, this list brings together most of the categories mentioned in the relevant literature and subtitling guidelines, in particular Neves 2005b and Díaz Cintas and Remael, 2007 (see section 3.3.4). In some cases, some of the existing rules have been rearranged; for instance rules (15) to (17) have all been suggested before, but have now been grouped together under the heading of *Subordination*. In other cases, existing categories have been expanded; for instance, the rule regarding objects (Rule 2) has been expanded to complements and has also been associated with Rule 1, while the rule regarding articles (Rule 3) has now been extended to cover all types of determiners. More importantly, three new rules have been introduced, namely Rule 7 regarding genitive and possessive forms, Rule 10 regarding complex prepositions such as *in terms of* and Rule 1, which details when and whether the subject should be separated from the predicate. Although tentative, as it seems to go beyond purely linguistic considerations, this rule is quite interesting as it specifies cases where the sentence should not be segmented at all, cases where editing is required to achieve this, as well as cases which go against the general recommendation to segment at the highest syntactic node possible.

It is important to note here that these rules are only rules in terms of linguistics. In terms of subtitling, they can be no more than recommendations promoting what are believed to be best practices according to guidelines available today. Even then, considering the multiplicity of factors influencing segmentation decisions, and as

suggested in section 6.2, these linguistic rules need to be tested by involving both subtitlers and viewers in order to determine whether they should always take precedence when subtitling (as suggested by current guidelines) or whether (and when) priority should be given to other factors. In all cases, however, the rules suggested in this thesis are believed to be a comprehensive starting point for further research involving subtitlers and viewers and more importantly, a useful point of reference for both students and professional subtitlers.

6.5 Segmentation and Language Technology

As mentioned in the introduction of Chapter 4, description is one of the three strands required to ensure high-quality subtitling. Corpus-based findings should inform empirical research involving viewers, as well as subtitlers themselves, and collective findings should then be used to inform existing guidelines and ultimately subtitling practice itself. In the case of segmentation, one way to ensure that this research process comes to full circle would involve designing a relevant tool to be used by subtitlers. Such tools already exist and have been integrated into subtitling software, yet their function is to either automatically break up imported text into individual subtitles in the case of off-line subtitling or to intelligently break live input text in the case of live subtitling (see for instance the text splitting function of WinCAPS). However, it would seem that although such automatic tools are particularly useful and in the case of live subtitling essential, they might not be the most appropriate solution in the case of off-line subtitling. Instead, a tool that would merely alert subtitlers by flagging potential segmentation problems appears to be more suitable for the following reasons:

1) Such a tool would only take into account linguistic aspects of segmentation and alert subtitlers to violations of linguistic rules. It would be up to subtitlers to decide whether to amend the subtitle according to the rule or whether to give precedence to other factors, such as geometry, the preference for a shorter first line or the preference for verbatim transcription.

2) In relation to 1, it would seem that building a fully-automatic segmentation tool that would take into consideration factors such as the need to have a short first line and the preference for verbatim subtitling is currently extremely complex if at all feasible. Yet, these factors are often taken into account by subtitlers.

3) In both types of tool, it would be up to the subtitler to decide whether to activate it or not. However, if both activated, a non-fully automatic tool would allow the subtitler to retain control of the process.

4) Given 1, 2 and 3 above and taking into consideration that automatic tools such as the text splitting function of WinCAPS only take into account linguistic rules and geometric considerations[53], one might doubt whether a fully automatic tool is the best solution within the context of off-line subtitling.

As mentioned above, the tool envisaged here would not break subtitles, but would alert subtitlers by flagging violations of linguistic rules pertaining to segmentation. This might be done for instance by colouring the words that have been poorly segmented and/or displaying a message *possible segmentation error* next to the information showing words per minute and characters per line (see Figure 26).

For such a tool to work, it would be necessary to either:

a) adapt the linguistic rules already incorporated in automatic tools such that of WinCAPS, and/or

b) exploit the rules suggested in section 6.4.

The second scenario would entail using natural language processing (NLP) tools, such as a tagger and parser and adapting them for the purposes of segmentation.[54] We could imagine two possible architectures.

---

[53] Punctuation is also taken into consideration when deciding where to break a line.
[54] For an introduction to taggers and parsers (and other NLP tools), see Jurafsky and Martin, 2000.

Figure 25. Screen shot of WinCAPS, segmentation error message could be displayed next to wpm information or be colour-coded
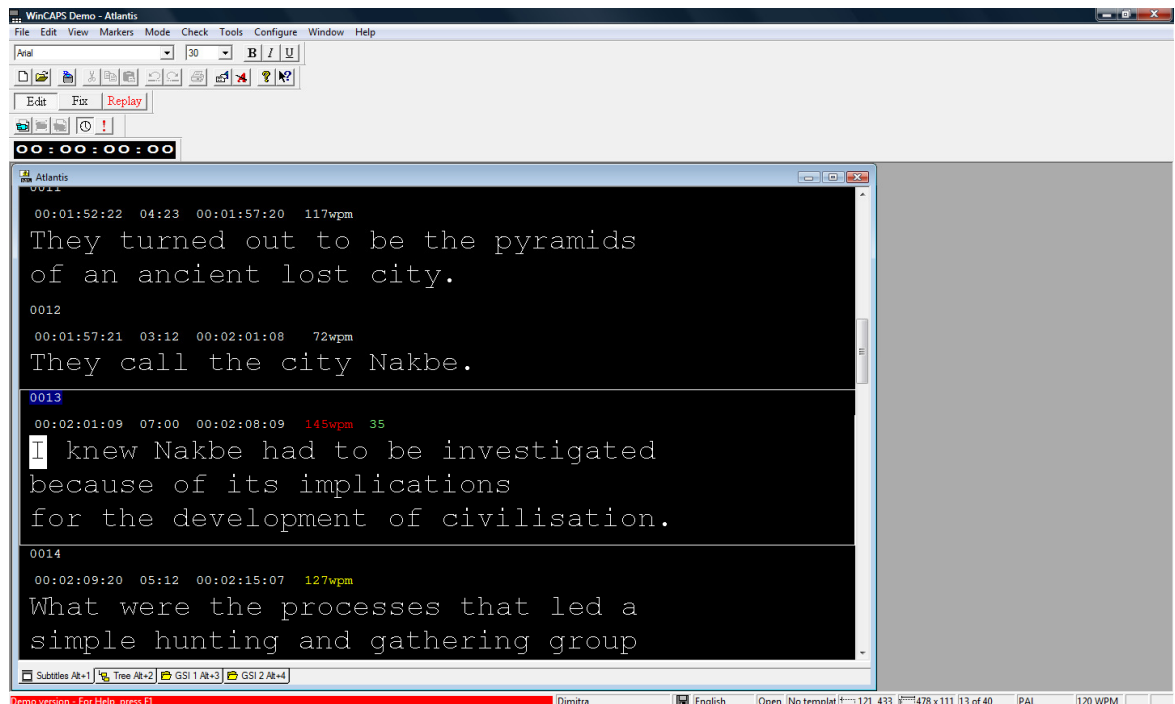
1) Tagger – Segmentation Module

The tagger assigns part-of-speech (POS) tags to the words of each two- and three-line subtitle. The proper segmentation module then compares the tagger's output to the right-hand side of available rules in order to determine whether any errors have occurred and flags them accordingly. For instance, when the segmentation module encounters a verb at the end of one line and an adverb at the beginning of the next, it consults the rules, determines which rule applies (here Rule 11) and decides that this is a poor line break and flags it accordingly. In addition, for some rules (namely 3, 7, 10, 12, 15, 16, 19), it suffices to simply tag the last word of the line. Thus, for instance, when the module encounters an article at the end of a line, it determines that the rule that applies is Rule 3 and that it has been flouted, since the article is the last word (and cannot be possibly followed by a noun on the same line). Consequently, it might be useful to just tag the last word first, check if any of the particular rules applies and if not, only then tag the entire subtitle and check again.

In this approach, the segmentation module only checks for co-occurrence of items on the basis of the rules without making proper use of grammar. As a result, the approach remains simple, but cannot help us identify errors under rules 1 and 2, while it might mistakenly combine items that do not belong together (since there is no syntactic analysis to identify constituents).

2) Tagger – Parser – Segmentation Module

In the second scenario, the tagger assigns POS tags to the words of each two- and three-line subtitle. The parser then assigns a syntactic structure to each sentence and the segmentation module finally checks against rules to determine whether any have been flouted. This way all rules are now covered and syntactic analysis makes it less likely to combine items that do not belong together and flag them as poorly segmented when they are not on the same line. Such a solution, however, would be quite resource-intensive and possibly not a welcome addition to subtitling software. A better solution might therefore involve combining the best features of the two architectures.

These two architectures are only suggestions that would need to be investigated further from an NLP aspect[55]. Implementing and testing them, as well as other possible solutions, would be the only reliable way to confirm what would be the most efficient way to build a tool that would help subtitlers better segment the subtitles they create. In any case, it is believed that the rules suggested in section 6.4 might be a fruitful reference point for any rule-based approach to segmentation.

6.6 Conclusion

This chapter has focused exclusively on segmentation within subtitles. First, results were presented on the basis of the present research corpus. Almost 25% of subtitles

---

[55] Chunking or shallow parsing, which provides more information than a simple tagger, but less than a full parser, might be a plausible alternative that would require less resources (for more on chunking, see Abney, 1991).

analysed were found to contain at least one line-break error. All errors were classified according to the linguistic categories involved. In addition, some of the potential factors influencing the subtitlers' choice not to create a proper (in linguistic terms) line break even when this was easily achievable were also identified; these included geometry, the preference for a shorter first line for purely aesthetic reasons[56] and the preference for verbatim subtitling. Of these factors, only geometry has been mentioned in existing guidelines with recommendations suggesting that linguistic considerations should take precedence when the result is a poor line break. There is little doubt that the other two factors should also be added to guidelines discussing segmentation. More importantly, given the different priorities between guidelines and some subtitlers on the one hand and other subtitlers on the other hand, further research is required to investigate whether linguistic considerations should always take precedence or whether (and under what circumstances) they could be overridden by any of the other factors. Crucially, such research should involve both subtitlers and viewers, in the particular case D/deaf and HoH people.

Moreover, this chapter also suggested a list of linguistic rules whose flouting leads to poor segmentation. This list brings together many of the categories mentioned in the relevant literature and guidelines, while some of the rules have been rearranged or expanded to cover more instances. In addition, three more rules have been included; of these, the most interesting is the rule which details when and whether the subject should be separated from the predicate. Although tentative, as it seems to go beyond purely linguistic considerations, this rule specifies cases where the sentence should not be segmented at all, cases where editing is required to achieve this, as well as cases which go against the general recommendation to segment at the highest syntactic node possible. It was also pointed out that the rules suggested here are only rules in linguistic terms; in terms of subtitling they can be no more than general guidelines given the multiplicity of factors influencing segmentation. In all cases, however, they are thought to be a comprehensive point of reference for both students and professional subtitlers, as well as for further research. In addition, it is believed that they could be further elaborated and used for the creation of a tool that would

---

[56] As opposed to instances where having a long first line actually conceals important picture information.

alert subtitlers working off-line by flagging potential segmentation problems in subtitles. Such a tool would only focus on linguistic aspects of segmentation and is considered a more appropriate solution in this context in comparison to current tools that automatically segment subtitles, as it does away with the need to take into consideration factors that cannot be automated (the preference for a shorter first line or the preference for verbatim transcription) and also allows the subtitler to decide whether to amend the subtitle according to the linguistic rules or whether to give precedence to other factors.

7. Conclusion & Suggestions for Further Research

There is no doubt that SDH offers an invaluable service to millions of D/deaf and HoH viewers providing them with access to audiovisual media broadcast on television and increasingly in cinemas, DVDs and on the Internet. What seems more doubtful, however, is whether SDH actually offers *full* accessibility to D/deaf and HoH people given that subtitles today still purport to be suitable for all, when by all accounts not only are the D/deaf and HoH two different groups, but what is more important within each group members differ not only in outlook, but also in language and communication needs. The time now seems opportune to provide answers to crucial questions, such as *who are D/deaf and HoH people*, *what is the intended audience as seen by broadcasters, subtitling companies and subtitlers*, *what is the actual audience* and *ultimately what should the intended audience be?* In order to provide an adequate reply to such questions, research would first need to:

1) provide a detailed description of the characteristics of current subtitle output,
2) provide an adequate account of how particular subtitle characteristics are related to D/deaf and HoH viewers' response and how this response varies among the D/deaf and HoH, and
3) compare both with existing guidelines and amend guidelines as necessary.

There is little doubt that such research should be conducted on a large scale covering a large number of genres, programmes, subtitles and characteristics, as well as a large number of D/deaf and HoH people, relevant associations, broadcasters, subtitling companies and subtitlers, researchers and universities.

This thesis has focused on the first and to a lesser extent the third step of this tripartite research plan. The following aims that were set in Chapter 1 have been addressed:

- Characteristics to measure and metrics to use were suggested in order to describe current subtitles (see Chapter 4) and were subsequently used to analyse a small sample corpus (see Chapters 5 and 6). The characteristics covered include subtitle rates, display time, whether time constraints are respected or not, interval times between subtitles, lines per subtitle, reduction rates, strategies used, distribution of editing strategies, the nature of deletions and line-break errors and their distribution. These characteristics are thought to be essential in providing as complete a description as possible. Regrettably, speech rates, another important characteristic, could not be recorded due to the fact that the research corpus did not contain the video for any of the programmes analysed.

- A new typology of deletions was suggested (see section 4.3.1.10). This typology is believed to be more intuitive than the one suggested by Kovačič (1992, 1994, 1998), which is based on linguistic terms that are not intuitively clear (*ideational*, *interpersonal* and *textual*). In addition, the present taxonomy draws a major distinction between oral elements such as false starts, hesitations, exclamations and overlapping speech from other deletions.

- Initial findings from implementing the typology to the present research corpus (see sections 5.6 and 5.8) seem to suggest that the deletion of oral features is not so prominent in documentaries and current affairs programmes. Although this is only a tentative finding that needs to be confirmed by further research, it comes in contrast to previous studies, which have reported a predominant deletion of oral features. This difference, however, is most likely to be due to the difference between interlingual and intralingual subtitling, and/or a difference in genres, as past studies, unlike the present research, come from the field of interlingual subtitling for hearing viewers and have mostly focused on the analysis of films and TV series.

- This thesis has also provided an updated list of recommendations regarding the linguistic aspect of segmentation (see section 6.4). This list has not only brought together recommendations currently scattered in existing guidelines, but has also

rearranged and expanded some of the recommendations to cover more instances. More importantly, three new recommendations have been added, of which the most interesting is the one which details when and whether the subject should be separated from the predicate. Although tentative, as it seems to go beyond purely linguistic considerations, this recommendation specifies cases where the sentence should not be segmented at all, cases where editing is required to achieve this, as well as cases which go against the general recommendation currently available to segment at the highest syntactic node possible. Overall, it might be hoped that this updated list of recommendations could become a point of reference for both students and subtitlers, as well as for further research.

- The aforementioned list of recommendations has also been presented in the form of linguistic rules in the belief that they could be further elaborated and used to either update existing automatic segmentation tools or to design a new tool that would alert subtitlers working off-line by flagging potential segmentation problems in subtitles. Such a tool is considered a more appropriate solution in this context in comparison to current automatic solutions, as it does away with the need to take into consideration factors that cannot be automated (the preference for a shorter first line or for verbatim transcription) and also allows the subtitler to decide whether to amend the subtitle according to the linguistic rules or whether to give precedence to other factors.

- The analysis of the present sample corpus has yielded a tentative account of the factors that appear to compete with linguistic considerations and may be responsible for linguistically poor line-breaks (see section 6.1). These factors include geometry, the preference for a shorter first line for purely aesthetic reasons[57] and the preference for verbatim subtitling. Of these factors, only geometry has been mentioned in existing guidelines (with relevant recommendations suggesting that linguistic considerations should take precedence

---

[57] As opposed to instances where having a long first line actually conceals important picture information.

when the result is a poor line break). There is little doubt that the other two factors should also be added to guidelines discussing segmentation.

In addition, we have argued for the need to adapt the current corpus typology used in translation studies to accommodate typical intralingual SDH corpora, i.e. corpora consisting of texts in one language and their subtitles in the same language. Such corpora are definitely not comparable, but cannot be considered parallel either, as according to the current definition of the term, parallel corpora are by default either bilingual or multilingual. Extending the term to distinguish between monolingual and bilingual/multilingual parallel corpora seems not only reasonable, but also essential.

Naturally, as with any project, this research is not without its limitations. Two are particularly important. The first one has to do with the fact that the research corpus only contained programme transcripts and subtitles and not the actual video of the programmes. This meant that there was no way to record essential information, such as speech rates, the complexity of visual information on screen, etc. The need for such information repeatedly emerged while analysing the three programmes, in particular when discussing subtitle rates, deletion and even segmentation. The second limitation concerns the fact that analysis was carried out on only two genres (documentaries and current affairs programmes), three programmes and 120 subtitles from each programme. Accordingly, any findings reported in this thesis can only apply to the particular programmes analysed and cannot be assumed to apply to documentaries and current affairs programmes in general. Further research involving a bigger sample would be required and this research would again need to be extended by including other genres, broadcasters and ultimately even other languages. Although this limitation is important, the aim of this research was to provide a blueprint for the large-scale research advocated here by suggesting characteristics to measure and metrics to use and offering some preliminary data for further research.

Among the data provided in the present thesis, further research would be indispensable in investigating and agreeing on a suitable metric for subtitle rate (see sections 4.3.1.5 and 5.9). It would also help us test some of the hypotheses

formulated and only partly tested in this research, namely the hypothesis that there is a relationship between the kind of text in the subtitles (narration, interview, conversation, extralinguistic information) and subtitle rate (see section 5.1) and the hypothesis that there is a relationship between the extent of deletion of oral features and the type of subtitling (intralingual vs. interlingual) and/or programme genre (see sections 5.6 and 5.8). In addition, regarding segmentation, further research involving both subtitlers and D/deaf and HoH people seems essential to determine not only the needs and preferences of viewers, but also to reconsider and if necessary update existing SDH guidelines which at least on the basis of the present findings appear to differ somewhat from subtitling practice. On a more general note, it is expected that combining corpus-based research describing current subtitles with empirical research into the needs and preferences of D/deaf and HoH people can only be instrumental in better mapping the world of SDH, in determining the actual and more importantly intended audience and in ultimately providing consistently high-quality subtitling for the millions of people relying on this invaluable service.

Appendix 1: Research Corpus – *Atlantis Uncovered*

Transcript File

NARRATOR (DILLY BARLOW): There is a theory that the origins of civilisation lie at the bottom of the ocean.   That there was once a great kingdom called Atlantis, home to an advanced and sophisticated society.   The story goes that it perished in a flood around 12,000 years ago, but a few survivors escaped in ships and brought civilisation to the primitive peoples around the world.   Atlantis is one of the most popular, yet controversial, ideas of the age.   Could it possible be true?   Beneath the thick canopy of the Guatemalan jungle lies a place of tantalising mystery.   When scientists first flew over these peaks 60 years ago they thought they were the tops of volcanoes.   They turned out to be the pyramids of an ancient, lost city.   They called the city Nakbe.

DR. RICHARD HANSEN (University of California, Los Angeles): I could see Nakbe on the horizon and I knew that it needed to be investigated because of the implications that it had for the development of civilisation.   What were the processes that led a simple hunting and gathering group to a great and magnificent ancient civilisation and I felt like it was necessary to get to Nakbe and understand that.

NARRATOR: This is a place that might solve the greatest mystery of our past: how did humans rise from barbarity to civilisation?   It's the question that drives archaeology.

RICHARD HANSEN: This is a fabulous site.   It's a wonderful opportunity to view the, the human process in action through an ancient window.

NARRATOR: Most archaeologists believe civilisation did not come from Atlantis, but if they dig deep enough at places like Nakbe perhaps they'll find they're wrong. Archaeologists have been discovering lost cities like Nakbe since the first Europeans went in search of colonies abroad.   All over the world in places that were totally unconnected they stumbled upon the spectacular remains of ancient societies.   The locals often knew nothing about the people who'd built these monuments.

DR. KEN FEDER (Central Connecticut State University): Archaeology was conducted largely by adventurers and rogues who went out to, to some horrible place in the middle of nowhere and discovered some ancient kingdom or lost civilisation and after all, if you have people finding forgotten tombs in Ancient Egypt and people finding lost civilisations in the middle of the tropical rainforest of Mexico certainly must have peeked people's curiosity at the time.

NARRATOR: The ancient world they discovered was full of impossible coincidences.   Of all the structures they could have built ancient civilisations on both sides of the Atlantic

chose to build pyramids.   Of all the ways of disposing of their dead they had mummified them.   Instead of alphabets they had written in strange pictures - hieroglyphs. Archaeologists had no way of dating most of these wonders.   To many it seemed inconceivable that these similar societies could have evolved separately.   It was logical to think they'd all started as one unique culture that spread, but others disagreed.

[PART 2]

RICHARD HANSEN: This building is, is radically different from Egyptian pyramids. These structures served as religious temples.   There were stairways going to the summit, there were rituals performed at the top. This is all part of a long tradition of architectural development and a long tradition of intense religious belief.

NARRATOR: It took centuries more for the Maya to perfect their technique.   But the Maya are only part of the Mezzo-American story.   Standing alone in the Oaxaca Valley in southern Mexico is a place even more revealing than Nakbe: Monte Alban.   The first archaeologists who came here couldn't tell who had built it.   When the site was carbon dated the result was a shock.   These pyramids were the oldest in the Americas, older even than the first Mayan monuments of this scale.   What mysterious people had built them?   The only way to find the answer was to dig throughout the valley for clues.   This work is still going on.

GARY FEINMAN: ...use anything that's in the corner of this wall...

NARRATOR: Gary and Linda Feinman are uncovering a new site - El Palmillo.

GARY FEINMAN: ...there's a second pot underneath, but that also appears to be broken.

LINDA FEINMAN (Field Museum, Chicago): We've had many faces of occupation up here where we have a wall and then we have another house foundation that was placed over that wall and then a plaster floor that was put over that later and in several areas we've got 3 or 4 levels of houses.

NARRATOR: Layer after painstaking layer throughout the valley archaeologists have sought the answer to the mystery of Monte Alban.

LINDA FEINMAN: It's been fascinating, it's been a surprise every day.   As soon as we started it was always oh look at this, here's another wall, oh, oh look at this, I've got plaster, oh look at this, that we've got, I've got a, a burial.

MAN: Gary, we've got another offering here. It's like peat.   It looks like carnivore teeth.

LINDA FEINMAN: That's real.

MAN: Oh boy.

LINDA: What's that down there, maybe that is a jaw.

MAN: It might be, he might have been buried with a dog.

LINDA: I think there's a dog in here... Oh my. Yeah, yeah that's a dog.

MAN: That's a dog.

MAN: That's a dog.

NARRATOR: Little by little, archaeologists throughout the valley have revealed no less than the birth of a new civilisation: the Zapotec, who evolved as gradually and as independently as did the Maya.

DR. GARY FEINMAN (Field Museum, Chicago): Well what we see here is a lot of signs of continuity in terms of how people built their houses, how people buried their dead and these things bear continuities to patterns that we see back to the earliest occupations at Monte Alban and sometimes even before.

NARRATOR: Archaeology has revealed the startling fact that in one small corner of the world the long road to a similar monument building society was taken by at least two different groups of people independently, proof that it could have happened all over the world again and again. There's a simple reason why so many ancient peoples built pyramids. Before engineers had invented the dome, the spire or structural steel a sloping pile was the only high structure you could build. If you wanted a high showy monument when all there was to build with was heavy stone it had to be a pyramid. It didn't take a master race from Atlantis to work it out.

[PART 3]

GARY FEINMAN: We knew that agricultural villages were established in the valley by 1500BC but how did that agriculture come about? That was a big question.

NARRATOR: The obvious place to look for clues was in the mountains where hunter/gatherers had lived. The people here had faced a perennial problem. A life of foraging high in the mountains where water was a long way off. When archaeologists studied the plant remains in this cave they found evidence of an ingenious solution. The hunter/gatherers had learnt to plant, but the first thing they grew wasn't food. They grew bottle gourds which they used to carry water. This is the first known plant domestication in the Americas, prompted not by hunger but by practicality. It was some 6,000 years after they began planting gourds that people here finally chose to settle in villages. By that time they were growing very different foods to those of the near eastern farmers: beans, chilli and

maize. Had the Atlantians brought farming to the Americas there is no reason why they wouldn't have brought the same crops as already existed in the near east.

GARY FEINMAN: ~~You cannot say that the plants that were domesticated in the near east would not have grown or done well here in Mezzo-America. In fact, we know that when the Spanish came over in the 16th-century what they brought with them were many of those same plants - wheat, barley and then the animals - goat, sheep, cattle - and all of these things thrived in Mezzo-America and you still see them today.~~
NARRATOR: And there were many essential tools of farming the super-race from Atlantis failed to bring. The plough and the wheel were unknown in the Americas until the Spanish brought them, just 400 years ago.

DAVID HARRIS: ~~Well I think that it would be a brave person indeed who, who would try to sustain an argument now against the evidence that agriculture was either introduced from some mysterious source, Atlantis or outer space or whatever, or indeed was just invented once and was then spread round the world by one or more groups of people because the evidence just does not support that kind of interpretation.~~

NARRATOR: But in spite of all the evidence, the allure of a lost civilisation is more powerful now than ever. Every year crowds flock to ancient sites in search of lost wisdom. Science continues to be ignored by a public yearning for the romance of a more mysterious past. Should this be dismissed as harmless fantasy? History has shown that fantasies about the past can lead to disaster.

COLIN RENFREW: It is dangerous when people have myths about their own past which have no foundation in reality. We've seen myths of that kind in our own time have tragic consequences. The National Socialists in Germany, the Nazis, had the notion of Aryan supremacy and the Holocaust was built on pernicious myths of that kind.

NARRATOR: The Nazi idea of an Aryan elite is well documented. What is less well known is that prominent Nazis believed that the master race originated in Atlantis. ~~One of the most passionate believers was Heinrich Himmler, Head of the SS. Himmler directed Germany scientists to seek the descendants of the Atlantian super-race in places from the Andes to Tibet. They scrutinised the physical features of the natives in search of any shred of evidence to support Himmler's notion that his Aryan ancestors, the Atlantians, had lived there.~~ These claims to an ancestral heritage in Atlantis fed the Nazis belief in the supremacy of the Aryan master race.

KEN FEDER: When we come to something like the lost continent of Atlantis we are better off knowing that civilisations developed more or less independently just so nobody can say some people are better than others, some are smarter than others because we know what happens down the line when we believe that, so I'm not going to tell you that belief in Atlantis is necessarily the first step towards genocide, or Holocaust, but what I'm telling you is we are on a very slippery slope if we believe in fantasies and that those fantasies lead us down to places we really don't want to go.

VOICE: Next week in the second of this two-part special Horizon examines the controversial theory of best-selling author Graham Hancock which is challenging mainstream archaeology.

KEN FEDER: Do you believe in Atlantis?

HORIZON: Atlantis Uncovered

Subtitle File

*** Timecode: In: 10:0:17:8 Out: 10:0:22:15
*** Framecount: In: 900433 Out: 900565
There is a theory
that the origins of civilisation

*** Timecode: In: 10:0:22:16 Out: 10:0:25:18
*** Framecount: In: 900566 Out: 900643
lie at the bottom of the ocean.

*** Timecode: In: 10:0:30:18 Out: 10:0:35:3
*** Framecount: In: 900768 Out: 900878
That there was once
a great kingdom called Atlantis -

*** Timecode: In: 10:0:35:4 Out: 10:0:39:6
*** Framecount: In: 900879 Out: 900981
home to an advanced
and sophisticated society.

*** Timecode: In: 10:0:41:10 Out: 10:0:47:8
*** Framecount: In: 901035 Out: 901183
The story goes that it perished
in a flood around 12,000 years ago.

*** Timecode: In: 10:0:51:13 Out: 10:0:59:8
*** Framecount: In: 901288 Out: 901483
But some survivors escaped in ships
and brought civilisation to the
primitive peoples around the world.

*** Timecode: In: 10:1:1:0 Out: 10:1:6:3
*** Framecount: In: 901525 Out: 901653
Atlantis is one of the most popular,
yet controversial, ideas of the age.

*** Timecode: In: 10:1:6:4 Out: 10:1:9:9
*** Framecount: In: 901654 Out: 901734
Could it possibly be true?

*** Timecode: In: 10:1:33:24 Out: 10:1:40:10
*** Framecount: In: 902349 Out: 902510
Beneath the thick canopy
of the Guatemalan jungle,
lies a place of tantalising mystery.

*** Timecode: In: 10:1:43:19 Out: 10:1:50:23
*** Framecount: In: 902594 Out: 902773
When scientists flew over here
60 years ago, they thought these
peaks were the tops of volcanoes.

*** Timecode: In: 10:1:52:22 Out: 10:1:57:20
*** Framecount: In: 902822 Out: 902945
They turned out to be the pyramids
of an ancient lost city.

*** Timecode: In: 10:1:57:21 Out: 10:2:1:8
*** Framecount: In: 902946 Out: 903033
They call the city Nakbe.

*** Timecode: In: 10:2:1:9 Out: 10:2:8:9
*** Framecount: In: 903034 Out: 903209
I knew Nakbe had to be investigated
because of its implications
for the development of civilisation.

*** Timecode: In: 10:2:9:20 Out: 10:2:15:7
*** Framecount: In: 903245 Out: 903382
What were the processes that led a
simple hunting and gathering group

*** Timecode: In: 10:2:15:8 Out: 10:2:18:10
*** Framecount: In: 903383 Out: 903460
to a magnificent
ancient civilisation?

*** Timecode: In: 10:2:18:11 Out: 10:2:23:23
*** Framecount: In: 903461 Out: 903598
I felt it was necessary
to get to Nakbe and understand it.

*** Timecode: In: 10:2:25:20 Out: 10:2:30:21
*** Framecount: In: 903645 Out: 903771
This is a place that might solve
the greatest mystery of our past.

*** Timecode: In: 10:2:30:22 Out: 10:2:35:16
*** Framecount: In: 903772 Out: 903891
How did humans rise
from barbarity to civilisation?

*** Timecode: In: 10:2:35:17 Out: 10:2:39:22
*** Framecount: In: 903892 Out: 903997
It's the question
that drives archaeology.

*** Timecode: In: 10:2:43:9 Out: 10:2:45:24
*** Framecount: In: 904084 Out: 904149
This is a fabulous sight.

*** Timecode: In: 10:2:46:0 Out: 10:2:52:20
*** Framecount: In: 904150 Out: 904320
It's a wonderful opportunity to view
the human process in action
through an ancient window.

*** Timecode: In: 10:2:52:21 Out: 10:2:58:2
*** Framecount: In: 904321 Out: 904452
Most archaeologists don't believe
civilisation came from Atlantis.

*** Timecode: In: 10:2:58:3 Out: 10:3:2:16
*** Framecount: In: 904453 Out: 904566
But if they dig deep enough,
they may find they're wrong.

*** Timecode: In: 10:3:10:24 Out: 10:3:17:24
*** Framecount: In: 904774 Out: 904949
Archaeologists have discovered
lost cities since Europeans first
went in search of colonies abroad.

*** Timecode: In: 10:3:29:2 Out: 10:3:34:7
*** Framecount: In: 905227 Out: 905357
All over the world, in places
that were totally unconnected,

*** Timecode: In: 10:3:34:8 Out: 10:3:40:5
*** Framecount: In: 905358 Out: 905505
they stumbled upon the spectacular
remains of ancient societies.

*** Timecode: In: 10:3:42:23 Out: 10:3:48:11
*** Framecount: In: 905573 Out: 905711
The locals often knew nothing about
those who had built these monuments.

*** Timecode: In: 10:3:49:24 Out: 10:3:55:8
*** Framecount: In: 905749 Out: 905883
Archaeology was conducted
largely by adventurers and rogues.

*** Timecode: In: 10:4:0:6 Out: 10:4:5:12
*** Framecount: In: 906006 Out: 906137
He went out to some horrible place
and discovered some ancient kingdom.

*** Timecode: In: 10:4:5:13 Out: 10:4:11:3
*** Framecount: In: 906138 Out: 906278
And, if you have people finding
forgotten tombs in ancient Egypt

*** Timecode: In: 10:4:11:4 Out: 10:4:16:5
*** Framecount: In: 906279 Out: 906405
and lost civilisations in the middle
of a tropical rainforest in Mexico,

*** Timecode: In: 10:4:16:6 Out: 10:4:20:6
*** Framecount: In: 906406 Out: 906506
it must have piqued
people's curiosity at the time.

*** Timecode: In: 10:4:24:16 Out: 10:4:29:20
*** Framecount: In: 906616 Out: 906745
The ancient world they discovered
was full of impossible coincidences.

*** Timecode: In: 10:4:29:21 Out: 10:4:33:12
*** Framecount: In: 906746 Out: 906837
Of all the structures
they could have built,

*** Timecode: In: 10:4:33:13 Out: 10:4:39:10
*** Framecount: In: 906838 Out: 906985
ancient civilisations on both sides
of the Atlantic chose pyramids.

*** Timecode: In: 10:4:39:11 Out: 10:4:44:8
*** Framecount: In: 906986 Out: 907108
Of all the ways of disposing of
their dead, they had mummified them.

*** Timecode: In: 10:4:47:5 Out: 10:4:53:23
*** Framecount: In: 907180 Out: 907348
Instead of alphabets, they had
written in strange pictures -
hieroglyphs.

*** Timecode: In: 10:4:56:12 Out: 10:5:0:3
*** Framecount: In: 907412 Out: 907503
Archaeologists
had no way of dating these wonders.

*** Timecode: In: 10:5:0:4 Out: 10:5:7:4
*** Framecount: In: 907504 Out: 907679
To many, it seemed inconceivable
that these similar societies
could have evolved separately.

*** Timecode: In: 10:5:7:5 Out: 10:5:14:5

*** Framecount: In: 907680 Out: 907855
It was logical to think they'd
all started as one unique culture
that spread.

[PART 2]

*** Timecode: In: 10:24:49:4 Out: 10:24:57:4
*** Framecount: In: 937229 Out: 937429
This was part of a long tradition
of architectural development
and intense religious belief.

*** Timecode: In: 10:24:59:20 Out: 10:25:5:2
*** Framecount: In: 937495 Out: 937627
It took centuries more for the Maya
to perfect their technique.

*** Timecode: In: 10:25:7:14 Out: 10:25:12:17
*** Framecount: In: 937689 Out: 937817
But the Maya are only part
of the Mesoamerican story.

*** Timecode: In: 10:25:15:14 Out: 10:25:23:6
*** Framecount: In: 937889 Out: 938081
Standing alone in the Oaxaca Valley
in Mexico is a place even more
revealing than Nakbe - Monte Alban.

*** Timecode: In: 10:25:23:7 Out: 10:25:28:9
*** Framecount: In: 938082 Out: 938209
The first archaeologists who came
here couldn't tell who had built it.

*** Timecode: In: 10:25:28:10 Out: 10:25:32:13
*** Framecount: In: 938210 Out: 938313
When the site was carbon dated,

*** Timecode: In: 10:25:32:14 Out: 10:25:35:7
*** Framecount: In: 938314 Out: 938382
the result was a shock.

*** Timecode: In: 10:25:35:8 Out: 10:25:39:22
*** Framecount: In: 938383 Out: 938497
These pyramids
were the oldest in the Americas.

*** Timecode: In: 10:25:39:23 Out: 10:25:45:0
*** Framecount: In: 938498 Out: 938625
Older, even, than the first
Mayan monuments of this scale.

*** Timecode: In: 10:25:45:1 Out: 10:25:47:20
*** Framecount: In: 938626 Out: 938695
What mysterious people built them?

*** Timecode: In: 10:25:56:0 Out: 10:26:3:12
*** Framecount: In: 938900 Out: 939087
The only way to find out was to dig
throughout the valley for clues.
This work is still going on.

*** Timecode: In: 10:26:5:10 Out: 10:26:10:18
*** Framecount: In: 939135 Out: 939268
Gary and Linda Feinman are
uncovering a new site - El Palmillo.

*** Timecode: In: 10:26:10:19 Out: 10:26:16:18
*** Framecount: In: 939269 Out: 939418
..The second part underneath, but
that also appears to be broken.

*** Timecode: In: 10:26:16:19 Out: 10:26:21:3
*** Framecount: In: 939419 Out: 939528
We've had many phases of occupation
up here -

*** Timecode: In: 10:26:21:4 Out: 10:26:26:13
*** Framecount: In: 939529 Out: 939663
a wall, another house foundation
placed over it,

*** Timecode: In: 10:26:26:14 Out: 10:26:29:7
*** Framecount: In: 939664 Out: 939732
and a plaster floor over that.

*** Timecode: In: 10:26:29:8 Out: 10:26:32:2
*** Framecount: In: 939733 Out: 939802
Sometimes, four levels of houses.

*** Timecode: In: 10:26:35:8 Out: 10:26:38:9
*** Framecount: In: 939883 Out: 939959
Layer after painstaking layer,

*** Timecode: In: 10:26:38:10 Out: 10:26:44:9
*** Framecount: In: 939960 Out: 940109
archaeologists have sought answers
to the mystery of Monte Alban.

*** Timecode: In: 10:26:44:10 Out: 10:26:48:4
*** Framecount: In: 940110 Out: 940204
It's been a surprise every day.

*** Timecode: In: 10:26:48:5 Out: 10:26:53:9
*** Framecount: In: 940205 Out: 940334
As soon as we began, it was "Here's
a wall!" "I've got plaster!"

*** Timecode: In: 10:26:53:10 Out: 10:26:56:19
*** Framecount: In: 940335 Out: 940419
"Look at this - I've got a burial."

*** Timecode: In: 10:26:56:20 Out: 10:27:0:4
*** Framecount: In: 940420 Out: 940504
Gary? I've got another offering. >

*** Timecode: In: 10:27:0:5 Out: 10:27:3:1
*** Framecount: In: 940505 Out: 940576
It's like teeth. >

*** Timecode: In: 10:27:3:2 Out: 10:27:8:6
*** Framecount: In: 940577 Out: 940706
It looks like carnivore teeth.
It looks real.

*** Timecode: In: 10:27:10:17 Out: 10:27:15:19
*** Framecount: In: 940767 Out: 940894
What's that down there?
Maybe that is a jaw.

*** Timecode: In: 10:27:15:20 Out: 10:27:19:6
*** Framecount: In: 940895 Out: 940981
He might have been buried
with a dog.

*** Timecode: In: 10:27:19:7 Out: 10:27:22:4
*** Framecount: In: 940982 Out: 941054
Yeah, that's a dog.

*** Timecode: In: 10:27:25:14 Out: 10:27:31:14
*** Framecount: In: 941139 Out: 941289
Archaeologists have revealed
the birth of a new civilisation -

*** Timecode: In: 10:27:31:15 Out: 10:27:39:7
*** Framecount: In: 941290 Out: 941482
the Zapotec -
who evolved as gradually and
as independently as did the Maya.

*** Timecode: In: 10:27:39:8 Out: 10:27:44:20
*** Framecount: In: 941483 Out: 941620
We see signs of continuity in terms
of how people built their houses,

*** Timecode: In: 10:27:44:21 Out: 10:27:49:21
*** Framecount: In: 941621 Out: 941746
how people buried their dead...
These things bear continuities

*** Timecode: In: 10:27:49:22 Out: 10:27:54:12
*** Framecount: In: 941747 Out: 941862
to patterns in the earliest
occupations at Monte Alban.

*** Timecode: In: 10:27:54:13 Out: 10:28:2:3
*** Framecount: In: 941863 Out: 942053
Archaeology has revealed
the startling fact
that in one corner of the world

*** Timecode: In: 10:28:2:4 Out: 10:28:9:22
*** Framecount: In: 942054 Out: 942247
the road to a monument-building
society was taken by at least two
groups of people independently.

*** Timecode: In: 10:28:9:23 Out: 10:28:15:14
*** Framecount: In: 942248 Out: 942389
Proof that it could have happened
all over the world again and again.

*** Timecode: In: 10:28:25:11 Out: 10:28:30:11
*** Framecount: In: 942636 Out: 942761
There's a simple reason why so many
ancient peoples built pyramids.

*** Timecode: In: 10:28:30:12 Out: 10:28:36:13
*** Framecount: In: 942762 Out: 942913
Before engineers had invented the
dome, the spire or structural steel,

*** Timecode: In: 10:28:36:14 Out: 10:28:42:1
*** Framecount: In: 942914 Out: 943051
a sloping pile was the only
high structure you could build.

*** Timecode: In: 10:28:42:2 Out: 10:28:49:5
*** Framecount: In: 943052 Out: 943230
If you wanted a high, showy
monument, when all there was
to build with was heavy stone,

[PART 3]

*** Timecode: In: 10:39:47:10 Out: 10:39:54:4
*** Framecount: In: 959685 Out: 959854
We knew that agricultural villages
were established in the valley
by 1500 BC.

*** Timecode: In: 10:39:54:5 Out: 10:39:58:6
*** Framecount: In: 959855 Out: 959956
But how did that agriculture
come about?

*** Timecode: In: 10:40:10:7 Out: 10:40:13:11
*** Framecount: In: 960257 Out: 960336
The obvious place to look for clues

*** Timecode: In: 10:40:13:12 Out: 10:40:19:4
*** Framecount: In: 960337 Out: 960479
was in the mountains
where hunter-gatherers had lived.

*** Timecode: In: 10:40:24:18 Out: 10:40:29:1
*** Framecount: In: 960618 Out: 960726
The people here
had faced a perennial problem -

*** Timecode: In: 10:40:29:2 Out: 10:40:34:2
*** Framecount: In: 960727 Out: 960852
a life of foraging in the mountains,
where water was a long way off.

*** Timecode: In: 10:40:34:3 Out: 10:40:39:8
*** Framecount: In: 960853 Out: 960983
When archaeologists studied
the plant remains in this cave,

*** Timecode: In: 10:40:39:9 Out: 10:40:43:18
*** Framecount: In: 960984 Out: 961093
they found evidence
of an ingenious solution.

*** Timecode: In: 10:40:43:19 Out: 10:40:47:14
*** Framecount: In: 961094 Out: 961189
The hunter-gatherers
had learnt to plant.

*** Timecode: In: 10:40:47:15 Out: 10:40:54:18
*** Framecount: In: 961190 Out: 961368
The first thing they grew wasn't
food. They grew bottle gourds,
which they used to carry water.

*** Timecode: In: 10:40:54:19 Out: 10:41:1:18
*** Framecount: In: 961369 Out: 961543
It was the Americas' first-known
plant domestication - prompted by
practicality rather than hunger.

*** Timecode: In: 10:41:9:2 Out: 10:41:16:24
*** Framecount: In: 961727 Out: 961924
It was some 6,000 years after they
began planting gourds that people
here chose to settle in villages.

*** Timecode: In: 10:41:17:0 Out: 10:41:22:23
*** Framecount: In: 961925 Out: 962073
Then, they were growing different
foods to the Near Eastern farmers -

*** Timecode: In: 10:41:22:24 Out: 10:41:25:13
*** Framecount: In: 962074 Out: 962138
beans, chilli and maize.

*** Timecode: In: 10:41:25:14 Out: 10:41:32:10
*** Framecount: In: 962139 Out: 962310
Had the Atlanteans brought farming
to the Americas, why wouldn't
they have brought the same crops

*** Timecode: In: 10:41:32:11 Out: 10:41:35:23
*** Framecount: In: 962311 Out: 962398
that already existed
in the Near East?

*** Timecode: In: 10:41:35:24 Out: 10:41:42:4
*** Framecount: In: 962399 Out: 962554
And there were many essential tools
of farming the super-race
from Atlantis failed to bring.

*** Timecode: In: 10:41:42:5 Out: 10:41:49:0
*** Framecount: In: 962555 Out: 962725
The plough and the wheel
were unknown in the Americas
until the Spanish brought them,

*** Timecode: In: 10:41:49:1 Out: 10:41:51:14
*** Framecount: In: 962726 Out: 962789
just 400 years ago.

*** Timecode: In: 10:41:54:22 Out: 10:42:0:1
*** Framecount: In: 962872 Out: 963001
But in spite of all the evidence,
the allure of a lost civilisation

*** Timecode: In: 10:42:0:2 Out: 10:42:2:20
*** Framecount: In: 963002 Out: 963070
is more powerful now than ever.

*** Timecode: In: 10:42:11:6 Out: 10:42:17:8
*** Framecount: In: 963281 Out: 963433
Every year, crowds flock to ancient
sites in search of lost wisdom.

*** Timecode: In: 10:42:22:17 Out: 10:42:30:4
*** Framecount: In: 963567 Out: 963754
Science continues to be ignored
by a public yearning for the
romance of a more mysterious past.

*** Timecode: In: 10:42:30:5 Out: 10:42:35:1
*** Framecount: In: 963755 Out: 963876
Should this be dismissed
as harmless fantasy?

*** Timecode: In: 10:42:35:2 Out: 10:42:41:1
*** Framecount: In: 963877 Out: 964026
History has shown that fantasies
about the past can lead to disaster.

*** Timecode: In: 10:42:41:2 Out: 10:42:48:2
*** Framecount: In: 964027 Out: 964202
It is dangerous when people
have myths about their own past
which have no foundation in reality.

*** Timecode: In: 10:42:48:3 Out: 10:42:53:8
*** Framecount: In: 964203 Out: 964333
We've seen myths of that kind in our
own time have tragic consequences -

*** Timecode: In: 10:42:53:9 Out: 10:42:57:7
*** Framecount: In: 964334 Out: 964432
the Nazis
had the notion of Aryan supremacy

*** Timecode: In: 10:42:57:8 Out: 10:43:2:7
*** Framecount: In: 964433 Out: 964557
and the holocaust was built on
pernicious myths of that kind.

*** Timecode: In: 10:43:2:8 Out: 10:43:7:24
*** Framecount: In: 964558 Out: 964699
The Nazi idea of an Aryan elite
is well documented.

*** Timecode: In: 10:43:8:0 Out: 10:43:15:8
*** Framecount: In: 964700 Out: 964883
What is less well-known is that
prominent Nazis believe that the
master race originated in Atlantis.

*** Timecode: In: 10:43:19:17 Out: 10:43:24:10
*** Framecount: In: 964992 Out: 965110
These claims to
an ancestral heritage in Atlantis

*** Timecode: In: 10:43:24:11 Out: 10:43:30:9
*** Framecount: In: 965111 Out: 965259
fed the Nazis' belief in the
supremacy of the Aryan master race.

*** Timecode: In: 10:43:36:1 Out: 10:43:41:1
*** Framecount: In: 965401 Out: 965526
When we come to something like
the lost continent of Atlantis,

*** Timecode: In: 10:43:41:2 Out: 10:43:46:24
*** Framecount: In: 965527 Out: 965674
we're better off knowing
that civilisations
developed independently

*** Timecode: In: 10:43:47:0 Out: 10:43:53:24
*** Framecount: In: 965675 Out: 965849
just so nobody can say some people
are better than others...
Some are smarter than others...

*** Timecode: In: 10:43:54:0 Out: 10:43:58:13
*** Framecount: In: 965850 Out: 965963
We know what happens down the line
when we believe that.

*** Timecode: In: 10:43:58:14 Out: 10:44:6:3
*** Framecount: In: 965964 Out: 966153
I'm not going to say that belief
in Atlantis is necessarily the first
step towards genocide or holocaust.

*** Timecode: In: 10:44:6:4 Out: 10:44:11:16
*** Framecount: In: 966154 Out: 966291
But we are on a very slippery
slope if we believe in fantasies.

*** Timecode: In: 10:44:11:17 Out: 10:44:16:14
*** Framecount: In: 966292 Out: 966414
Fantasies lead us to places
we don't want to go.

Merge with annotation

NARRATOR ~~(DILLY BARLOW):~~
<sub1>
There is a theory
that the origins of civilisation
</sub1><strat VERBATIM><duration 5.07><TC respecting>

<sub2>
lie at the bottom of the ocean.
</sub2><strat VERBATIM><duration 3.02><TC respecting>

<sub3>
That there was once
a great kingdom called Atlantis - ~~,~~
</sub3><strat VERBATIM><duration 4.10><TC respecting>

<sub4>
home to an advanced
and sophisticated society.
</sub4><strat VERBATIM><duration 4.02><TC respecting>

<sub5>
The story goes that it perished
in a flood around 12,000 years ago . ~~,~~
</sub5><strat VERBATIM><duration 5.23><TC respecting>

<sub6>
But ~~but~~ <CONDENSATION> some ~~a few~~ </CONDENSATION> survivors escaped in ships
and brought civilisation to <w PRP> the <w AT0> <LINEBREAKERR>
primitive peoples <w NN> around the world.
</sub6><strat CONDENSATION><duration 7.20><TC not respecting><3-liner>

<sub7>
Atlantis is one of the most popular,
yet controversial, ideas of the age.
</sub7><strat VERBATIM><duration 5.03><TC not respecting>

<sub8>
Could it <ERRCORRECTION> possibly~~e~~ </ERRCORRECTION> be true?
</sub8><strat VERBATIM+ERRCORRECTION><duration 3.05><TC respecting>

Beneath the thick canopy
of the Guatemalan jungle ,
lies a place of tantalising mystery.
</sub9><strat VERBATIM><duration 6.10><TC not respecting><3-liner>

<sub10>
When scientists <DELETION> ~~first~~ </DELETION> flew over <CONDENSATION> here
~~these~~ ~~peaks~~ <CONDENSATION>
60 years ago , they thought <SPECIFICATION> these <w DT0> <LINEBREAKERR>
peaks <w NN> ~~they~~ </SPECIFICATION> were the tops of volcanoes.
</sub10><strat DELETION, CONDENSATION, SPECIFICATION><duration 7.04><TC
not respecting><3-liner>

<sub11>
They turned out to be the pyramids
of an ancient , lost city.
</sub11><strat VERBATIM><duration 4.23><TC respecting>

<sub12>
They <CONDENSATION> call ~~called~~ </CONDENSATION> the city Nakbe.
</sub12><strat CONDENSATION><duration 4.12><TC respecting but too long>

DR. RICHARD HANSEN (University of California, Los Angeles):
<sub13>
<DELETION> ~~I could see Nakbe on the horizon~~ </DELETION> <DELETION> ~~and~~
</DELETION> I knew <DELETION> ~~that~~ </DELETION> <SPECIFICATION> Nakbe
~~it~~</SPECIFICATION> <CONDENSATION> had ~~needed~~ </CONDENSATION> to be
investigated
because of <CONDENSATION> its ~~the~~ implications ~~that it had~~ </CONDENSATION>

for the development of civilisation.
</sub13><strat DELETION, DELETION, DELETION, SPECIFICATION,
CONDENSATION, CONDENSATION><duration 7.00><TC not respecting><3-liner

<sub14>
What were the processes that led a <w AT0> <LINEBREAKERR>
simple hunting and gathering group <w NN>
</sub14><strat VERBATIM><duration 5.12><TC not respecting>

<sub15>
to <w PRP> a <w AT0> <DELETION> ~~great~~ </DELETION> <DELETION> ~~and~~
</DELETION> magnificent <w AJ0> <LINEBREAKERR>
ancient <w AJ0> civilisation <w NN> ?
</sub15><strat DELETION, DELETION><duration 3.02><TC not respecting>

<sub16>
</DELETION> ~~and~~ </DELETION> I felt </DELETION> ~~like~~ </DELETION> it was necessary
to get to Nakbe and understand <CONDENSATION> it ~~that~~ </CONDENSATION> .
</sub16><strat DELETION, DELETION, CONDENSATION><duration 5.12><TC respecting>

NARRATOR:
<sub17>
This is a place that might solve
the greatest mystery of our past . ~~:~~
</sub17><strat VERBATIM><duration 5.01><TC not respecting>

<sub18>
How ~~how~~ did humans rise
from barbarity to civilisation?
</sub18><strat VERBATIM><duration 4.19><TC respecting>

<sub19>
It's the question
that drives archaeology.
</sub19><strat VERBATIM><duration 4.05><TC respecting>

RICHARD HANSEN:
<sub20>
This is a fabulous <ERRCORRECTION> sight ~~site~~ </ERRCORRECTION> .
</sub20><strat VERBATIM+ERRCORRECTION><duration 2.15><TC respecting>

<sub21>
It's a wonderful opportunity to view
<DELETION> the ~~,the~~ <DELETION> human process in action
through an ancient window.
</sub21><strat DELETION><duration 6.20><TC not respecting><3-liner>

NARRATOR:
<sub22>
Most archaeologists <PARAPHRASE> don't believe civilisation
came ~~did not come~~ </PARAPHRASE> from Atlantis. ~~,~~
</sub22><strat PARAPHRASE><duration 5.06><TC not respecting>

<sub23>
But ~~but~~ if they dig deep enough <DELETION> ~~at places like Nakbe~~ </DELETION> .

<CONDENSATION> ~~perhaps~~ they may ~~'ll~~ </CONDENSATION> find they're wrong.
</sub23><strat DELETION, CONDENSATION><duration 4.13><TC not respecting>

<sub24>
Archaeologists <CONDENSATION> have <u>discovered</u> ~~been discovering~~
</CONDENSATION>
lost cities <DELETION> ~~like Nakbe~~ </DELETION> since <PARAPHRASE> ~~the first~~
Europeans <u>first</u> <w AV0> </PARAPHRASE> <LINEBREAKERR>
went <w VVD> in search of colonies abroad.
</sub24><strat CONDENSATION, DELETION, PARAPHRASE><duration 7.00><TC not
respecting><3-liner>

<sub25>
All over the world <u>,</u> in places
that were totally unconnected <u>,</u>
</sub25><strat VERBATIM><duration 5.05><TC respecting>

<sub26>
they stumbled upon <w PRP> the <w AT0> spectacular <w AJ0> <LINEBREAKERR>
remains <w NN> of ancient societies.
</sub26><strat VERBATIM><duration 5.22><TC respecting>

<sub27>
The locals often knew nothing about <w PRP> <LINEBREAKERR>
</CONDENSATION> those <w DT0> ~~the people~~ </CONDENSATION> who
<SPECIFICATION> <u>had</u> ~~'d~~ </SPECIFICATION> built these monuments.
</sub27><strat CONDENSATION, SPECIFICATION><duration 5.13><TC not
respecting>

DR. KEN FEDER (Central Connecticut State University):
<sub28>
Archaeology was conducted <w VVD> <LINEBREAKERR>
largely <w AV0> by adventurers and rogues<u>.</u>
</sub28><strat VERBATIM><duration 5.09><TC respecting>

<sub 29>
<PARAPHRASE> <u>He</u> ~~who~~ </PARAPHRASE> went out <DELETION> to ~~, to~~
<DELETION> some horrible place <DELETION> ~~in the middle of nowhere~~
</DELETION>
and discovered some ancient kingdom <DELETION> ~~or~~</DELETION> <DELETION>
~~lost civilisation~~ </DELETION> ~~-~~ <u>.</u>
</sub29><strat PARAPHRASE, DELETION, DELETION, DELETION,
DELETION><duration 5.06><TC not respecting>

<sub30>
And ~~and~~ <DELETION> ~~after~~ ~~all~~ </DELETION>, if you have people finding
forgotten tombs in <u>ancient</u> ~~Ancient~~ Egypt
</sub30><strat DELETION><duration 5.15><TC respecting>

<sub31>
And <DELETION> people finding </DELETION> lost civilisations in the middle
of <CONDENSATION> a the </CONDENSATION> tropical rainforest <PARAPHRASE>
in of </PARAPHRASE> Mexico .
</sub31><strat DELETION, CONDENSATION, PARAPHRASE><duration 5.01><TC not
respecting>

<sub32>
<ADDITION> It </ADDITION> <DELETION> certainly </DELETION> must have
<ERRCORRECTION> piqued peeked </ERRCORRECTION>
people's curiosity at the time.
</sub32><strat ADDITION, DELETION><duration 4.00><TC not respecting>

NARRATOR:
<sub33>
The ancient world they discovered
was full of impossible coincidences.
</sub33><strat VERBATIM><duration 5.04><TC not respecting>

<sub34>
Of all the structures
they could have built .
</sub34><strat VERBATIM><duration 3.16><TC not respecting>

<sub35>
ancient civilisations on both sides
of the Atlantic chose <DELETION> to build </DELETION> pyramids.
</sub35><strat DELETION><duration 5.22><TC respecting>

<sub36>
Of all the ways of disposing of <w PRP> <LINEBREAKERR>
their <w DPS> dead <w NN> , they had mummified them.
</sub36><strat VERBATIM><duration 4.22><TC not respecting>

<sub37>
Instead of alphabets , they had <w VHD> <LINEBREAKERR>
written <w VVN> in strange pictures –
hieroglyphs.
</sub37><strat VERBATIM><duration 6.18><TC respecting><3-liner>

<sub38>
Archaeologists
had no way of dating <DELETION> most of </DELETION> these wonders.
</sub38><strat DELETION><duration 3.16><TC not respecting>

To many , it seemed inconceivable
that these similar societies
could have evolved separately.
</sub39><strat VERBATIM><duration 7.00><TC not respecting>

<sub40>
It was logical to think they'd <w VHD> <LINEBREAKERR>
all started <w VVN> as one unique culture
that spread . ,
</sub40><strat VERBATIM><duration 7.00><TC respecting><3-liner>

<sub41>
This <PARAPHRASE> was is </PARAPHRASE> <DELETION> all </DELETION> part
of a long tradition
of architectural development
and <DELETION> a long tradition of </DELETION> intense religious belief.
</sub41><strat PARAPHRASE, DELETION, DELETION<duration 8.00><TC
respecting><3-liner>

NARRATOR:
<sub42>
It took centuries more for the Maya
to perfect their technique.
</sub42><strat VERBATIM><duration 5.07><TC respecting>

<sub43>
But the Maya are only part
of the <ALTSP> Mesoamerican Mezzo-American </ALTSP> story.
</sub43><strat VERBATIM+ALTSP><duration 5.03><TC respecting>

<sub44>
Standing alone in the Oaxaca Valley
in <DELETION> southern </DELETION> Mexico is a place even more <w AV0>
<LINEBREAKERR>
revealing <w AJ0> than Nakbe - : Monte Alban.
</sub44><strat DELETION><duration 7.17><TC not respecting><3-liner>

<sub45>
The first archaeologists who came <w VVD> <LINEBREAKERR>
here <w AV0> couldn't tell who had built it.
</sub45><strat VERBATIM><duration 5.02><TC not respecting>

<sub46>
When the site was carbon dated .
</sub46><strat VERBATIM><duration 4.03><TC respecting but too long>

<sub47>
the result was a shock.
</sub47><strat VERBATIM><duration 2.18><TC respecting>

<sub48>
These pyramids
were the oldest in the Americas . ,
</sub48><strat VERBATIM><duration 4.14><TC respecting>

<sub49>
Older , even , than <w PRP> the <w AT0> first <w AJ0> <LINEBREAKERR>
Mayan <w AJ0> monuments <w NN> of this scale.
</sub49><strat VERBATIM><duration 5.02><TC respecting>

<sub50>
What mysterious people <CONDENSATION> had built </CONDENSATION> them?
</sub50><strat CONDENSATION><duration 2.19><TC not respecting>

<sub51>
The only way to <CONDENSATION> find out the answer </CONDENSATION> was to
dig
throughout the valley for clues.
This work is still going on.
</sub51><strat CONDENSATION><duration 7.12><TC not respecting><3-liner>

GARY FEINMAN:
<DELETION> ...use anything that's in the corner of this wall... </DELETION>
<no corresponding subtitle><strat DELETION>

NARRATOR:
<sub52>
Gary and Linda Feinman are <w VVB> <LINEBREAKERR>
uncovering <w VVG> a new site - El Palmillo.
</sub52><strat VERBATIM><duration 5.08><TC respecting>

GARY FEINMAN:
<sub53>
.. ... <CONDENSATION> The there's a <CONDENSATION> second </ERROR>
part pot </ERROR> underneath, but <w CJC> <LINEBREAKERR>
that also appears to be broken.
</sub53><strat CONDENSATION><duration 5.24><TC respecting>
LINDA FEINMAN (Field Museum, Chicago):

<sub54>
We've had many <ERRCORRECTION> phases faces </ERRCORRECTION> of occupation
up here –
</sub54><strat VERBATIM+ERRCORRECTION><duration 4.09><TC respecting>

<sub55>
<DELETION> where we have </DELETION> a wall . <DELETION> and </DELETION>
<DELETION> then </DELETION> <DELETION> we have </DELETION> another
house foundation
<DELETION> that was </DELETION> placed over <CONDENSATION> it that well
</CONDENSATION> .
</sub55><strat DELETION, DELETION, DELETION, DELETION, DELETION,
CONDENSATION><duration 5.09><TC respecting>

<sub56>
and <DELETION> then </DELETION> a plaster floor <DELETION> that was put
</DELETION> over that <DELETION> later </DELETION> .
</sub56><strat DELETION, DELETION, DELETION><duration 2.18><TC respecting>

<sub57>
<DELETION> and </DELETION> <CONDENSATION> Sometimes in several areas
</CONDENSATION> . <DELETION> we've got </DELETION> <DELETION> 3
</DELETION> <DELETION> or </DELETION> four 4 levels of houses.
</sub57><strat DELETION, CONDENSATION, DELETION, DELETION,
DELETION><duration 2.19><TC not respecting>

NARRATOR:
<sub58>
Layer after painstaking layer <DELETION> throughout the valley </DELETION> .
</sub58><strat DELETION><duration 3.01><TC respecting>

<sub59>
archaeologists have sought <CONDENSATION> answers the answer
</CONDENSATION>
to the mystery of Monte Alban.
</sub59><strat CONDENSATION><duration 5.24><TC respecting>

~~LINDA FEINMAN:~~
<sub60>
it's been fascinating, </DELETION> It's ~~it's~~ been a surprise every day.
</sub60><strat DELETION><duration 3.19><TC respecting>

<sub61>
As soon as we <CONDENSATION> began ~~started~~ <CONDENSATION> , it was
<DELETION> ~~always~~ </DELETION> <DELETION>~~oh~~</DELETION>
<DELETION>~~look at this,~~ </DELETION> " Here's ~~here's~~
<CONDENSATION> ~~another~~ a </CONDENSATION> wall ! , " <DELETION> ~~oh, oh~~
~~look at this~~ </DELETION> , " I've got plaster ! , "
</sub61><strat CONDENSATION, DELETION, DELETION, DELETION,
CONDENSATION, DELETION><duration 5.04><TC respecting>

<sub62>
" <DELETION> ~~oh~~ </DELETION> Look ~~look~~ at this - , <DELETION> ~~that we've got,~~
</DELETION> I've got <DELETION> ~~a,~~ a </DELETION> burial. "
</sub62><strat DELETION, DELETION, DELETION><duration 3.09><TC respecting>

MAN:
<sub63>
Gary ? , <CONDENSATION> I' ~~we'~~ <CONDENSATION2> ve got another offering
<DELETION> ~~here~~ </DELETION> . >
</sub63><strat CONDENSATION, DELETION><duration 3.09><TC respecting>

<sub64>
It's like <ERRCORRECTION> teeth ~~peat~~ </ERRCORRECTION> . >
</sub64><strat VERBATIM+ERRCORRECTION ><duration 2.21><TC respecting>

<sub65>
It looks like carnivore teeth.
LINDA FEINMAN: <PARAPHRASE> It looks ~~That's~~ </PARAPHRASE> real.
</sub65><strat PARAPHRASE><duration 5.04><TC respecting>

MAN: Oh boy.
<No corresponding subtitle><strat DELETION>

LINDA:
<sub66>
What's that down there ? ,
Maybe ~~maybe~~ that is a jaw.
</sub66><strat VERBATIM><duration 5.02><TC respecting>

~~MAN:~~
<sub67>
<DELETION> ~~It might be,~~ </DELETION> He ~~he~~ might have been buried
with a dog.
</sub67><strat DELETION><duration 3.11><TC respecting>

LINDA:
<sub68>
<DELETION> ~~I think there's a dog in here...~~</DELETION> <DELETION> ~~Oh my.~~
</DELETION>
<DELETION> Yeah, ~~yeah~~ </DELETION> that's a dog.
</sub68><strat DELETION, DELETION, DELETION><duration 2.22><TC respecting>

MAN: That's a dog.
<No corresponding subtitle><strat DELETION>

MAN: That's a dog.
<No corresponding subtitle><strat DELETION>

NARRATOR:
<sub69>
<DELETION> ~~Little by little,~~ </DELETION> Archaeologists ~~archaeologists~~ <DELETION>
~~throughout the valley~~ </DELETION> have revealed
<DELETION> ~~no less than~~ </DELETION> the birth of a new civilisation ~~-~~.
</sub69><strat DELETION, DELETION, DELETION><duration 6.00><TC respecting>

<sub70>
 the Zapotec - ~~,~~
who evolved as gradually and <w CJC> <LINEBREAKERR>
as independently as did the Maya.
</sub70><strat VERBATIM><duration 7.17><TC respecting><3-liner>

DR. GARY FEINMAN (Field Museum, Chicago):
<sub71>
<DELETION> ~~Well~~ </DELETION> <CONDENSATION> ~~what~~ We ~~we~~ see
<DELETION> ~~here~~ </DELETION> ~~is~~ </CONDENSATION> <DELETION> ~~a lot of~~
</DELETION> signs of continuity in <w PRP31> terms <w PRP32> <LINEBREAKERR>

of <w PRP33> how people built their houses,
</sub71><strat DELETION, CONDENSATION, DELETION, DELETION><duration
5.12><TC not respecting>

<sub72>
how people buried their dead ...
<DELETION> ~~and~~ </DELETION> These ~~these~~ things bear continuities
</sub72><strat DELETION><duration 5><TC not respecting>

<sub73>
to patterns <CONDENSATION> in <w PRP> ~~that we see back to~~ <CONDENSATION> the <w AT0> earliest <w AJS> <LINEBREAKERR>
occupations <w NN> at Monte Alban <DELETION> ~~and~~ </DELETION> <DELETION>~~sometimes~~ </DELETION> <DELETION>~~even before~~ </DELETION> .
</sub73><strat CONDENSATION, DELETION, DELETION, DELETION><duration 4.15><TC not respecting>

NARRATOR:
<sub74>
Archaeology has revealed
the startling fact
that in one <DELETION> ~~small~~ </DELETION> corner of the world
</sub74><strat DELETION><duration 7.15><TC respecting><3-liner>

<sub75>
the <DELETION> ~~long~~ </DELETION> road to <w PRP> a <w AT0> <DELETION> ~~similar~~ </DELETION> monument–building <w AJ0> <LINEBREAKERR>
society <w NN> was taken by at least two <DELETION> ~~different~~ </DELETION> <LINEBREAKERR>
groups of people independently . ~~;~~
</sub75><strat DELETION, DELETION, DELETION><duration 7.18><TC not respecting><3-liner>

<sub76>
Proof ~~proof~~ that it could have happened
all over the world again and again.
</sub76><strat VERBATIM><duration 5.16><TC not respecting>

<sub77>
There's a simple reason why <w AVQ> so<w AV0> many <w DT0> <LINEBREAKERR>
ancient <w AJ0> peoples <w NN> built pyramids.
</sub77><strat VERBATIM><duration 5.00><TC not respecting>

<sub78>
Before engineers had invented the <w AT0> <LINEBREAKERR>
dome <w NN>, the spire or structural steel .
</sub78><strat VERBATIM><duration 6.01><TC respecting>

<sub79>
a sloping pile was the <w AT0> only <w AJ0> <LINEBREAKERR>
high <w AJ0> structure <w NN> you could build.
</sub79><strat VERBATIM><duration 5.12><TC respecting>

<sub80>
If you wanted a <w AT0> high <w AJ0> , showy <w AJ0> <LINEBREAKERR>
monument <w NN> , when all there was
to build with was heavy stone .
</sub80><strat VERBATIM><duration 7.03><TC respecting><3-liner>

GARY FEINMAN:
<sub 81>
We knew that agricultural villages
were established in the valley
by 1500BC .
</sub81><strat VERBATIM><duration 6.19><TC respecting><3-liner>

<sub 82>
But but how did that agriculture
come about? <DELETION> That was a big question. </DELETION>
</sub82><strat DELETION><duration 4.01><TC respecting>

NARRATOR:
<sub 83>
The obvious place to look for clues
</sub83><strat VERBATIM><duration 3.04><TC respecting>

<sub 84>
was in the mountains
where <ALTSP> hunter-gatherers hunter/gatherers </ALTSP> had lived.
</sub84><strat VERBATIM+ALTSP><duration 5.17><TC respecting>

<sub 85>
The people here
had faced a perennial problem - .
</sub85><strat VERBATIM><duration 4.08><TC respecting>

<sub 86>
a A life of foraging <DELETION> high </DELETION> in the mountains,
where water was a long way off.
</sub86><strat DELETION><duration 5><TC not respecting>

<sub 87>
When archaeologists studied
the plant remains in this cave .
</sub87><strat VERBATIM><duration 5.05><TC respecting>

<sub 88>
they found evidence
of an ingenious solution.
</sub88><strat VERBATIM><duration 4.09><TC respecting>

<sub 89>
The <ALTSP> hunter-gatherers hunter/gatherers </ALTSP>
had learnt to plant. ,
</sub89><strat VERBATIM+ALTSP><duration 3.20><TC respecting>

<sub 90>
<DELETION> but </DELETION> The the first thing they grew wasn't_Verb
<LINEBREAKERR>
food_Complement. They grew bottle gourds,
which they used to carry water.
</sub90><strat DELETION><duration 7.03><TC not respecting><3-liner>

<sub 91>
<CONDENSATION> It This </CONDENSATION> <PARAPHRASE> was is
</PARAPHRASE> the <w AT0> <CONDENSATION> Americas' first –known <w AJ0>
<LINEBREAKERR>
plant domestication <w NN> in the Americas <CONDENSATION> - , prompted by <w
PREP> <LINEBREAKERR>
<PARAPHRASE> not by hunger but practicality <w NN> rather than hunger
</PARAPHRASE>.

</sub91><strat CONDENSATION, PARAPHRASE, CONDENSATION,
PARAPHRASE><duration 6.24><TC not respecting><3-liner>

<sub 92>
It was some 6,000 years after they_Subj <LINEBREAKERR>
began planting gourds that people_Subj <LINEBREAKERR>
{here <DELETION> finally </DELETION> chose to settle in villages.}_predicate
</sub92><strat DELETION><duration 7.22><TC not respecting><3-liner>

<sub 93>
<CONDENSATION> Then By that time </CONDENSATION> , they were growing
<DELETION> very </DELETION> different <w AJ0> <LINEBREAKERR>
foods <w NN> to <DELETION> those of </DELETION> the Near Eastern near eastern
farmers - .
</sub93><strat CONDENSATION, DELETION, DELETION><duration 5.23><TC
respecting>

<sub 94>
beans, chilli and maize.
</sub94><strat VERBATIM><duration 2.14><TC respecting>

<sub 95>
Had the <ALTSP> Atlanteans ~~Atlantians~~ </ALTSP> brought farming
to the Americas, <CONDENSATIONa> ~~there is no reason~~ why ~~they~~ wouldn't <w VM0>
<LINEBREAKERR>
they have <w VHI> brought <w VVN> the same crops
</sub95><strat CONDENSATION6><duration 6.21><TC not respecting><3-liner>

<sub 96>
<ERRCORRECTION> that ~~as~~ </ERRCORRECTION> already existed
in the Near East ~~near east~~ ? ~~.~~ </CONDENSATIONa>
</sub96><strat CONDENSATION><duration 3.12><TC respecting>

~~GARY FEINMAN: You cannot say that the plants that were domesticated in the near east would not have grown or done well here in Mezzo-America. In fact, we know that when the Spanish came over in the 16th-century what they brought with them were many of those same plants - wheat, barley and then the animals - goat, sheep, cattle - and all of these things thrived in Mezzo-America and you still see them today.~~

NARRATOR:
<sub 97>
And there were many essential tools
of farming the super-race
from Atlantis failed to bring.
</sub97><strat VERBATIM><duration 6.05><TC not respecting><3-liner>

<sub 98>
The plough and the wheel
were unknown in the Americas
until the Spanish brought them,
</sub98><strat VERBATIM><duration 6.20><TC not respecting><3-liner>

<sub 99>
just 400 years ago.
</sub99><strat VERBATIM><duration 2.13><TC respecting>

~~DAVID HARRIS: Well I think that it would be a brave person indeed who, who would try to sustain an argument now against the evidence that agriculture was either introduced from some mysterious source, Atlantis or outer space or whatever, or indeed was just invented once and was then spread round the world by one or more groups of people because the evidence just does not support that kind of interpretation.~~

NARRATOR:
<sub 100>
But in spite of all the evidence,
the allure of a lost civilisation
</sub100><strat VERBATIM><duration 5.04><TC not respecting>
<sub 101>
is more powerful now than ever.
</sub101><strat VERBATIM><duration 2.18><TC respecting>

<sub 102>
Every year, crowds flock to <w PRP> ancient <w AJ0> <LINEBREAKERR>
sites <w NN> in search of lost wisdom.
</sub102><strat VERBATIM><duration 6.02><TC respecting>

<sub 103>
Science continues to be ignored
by a public yearning for <w PRP> the <w AT0> <LINEBREAKERR>
romance <w NN> of a more mysterious past.
</sub103><strat VERBATIM><duration 7.12><TC not respecting><3-liner>

<sub 104>
Should this be dismissed
as harmless fantasy?
</sub104><strat VERBATIM><duration 4.21><TC respecting>

<sub 105>
History has shown that fantasies
about the past can lead to disaster.
</sub105><strat VERBATIM><duration 5.24><TC respecting>

COLIN RENFREW:
<sub 106>
It is dangerous when people <LINEBREAKERR>
have myths about their own past
which have no foundation in reality.
</sub106><strat VERBATIM><duration 7.00><TC not respecting><3-liner>

<sub 107>
We've seen myths of that kind in <w PRP> our <w DPS> <LINEBREAKERR>
own <w DT0> time <w NN> have tragic consequences -.
</sub107><strat VERBATIM><duration 5.05><TC not respecting>

<sub 108>
<DELETION>The National Socialists in Germany,</DELETION> the Nazis ,
had the notion of Aryan supremacy
</sub108><strat DELETION><duration 3.23><TC respecting>

<sub 109>
and the holocaust ~~Holocaust~~ was built on <LINEBREAKERR>
pernicious myths of that kind.
</sub109><strat VERBATIM><duration 4.24><TC not respecting>

NARRATOR:
<sub 110>
The Nazi idea of an Aryan elite
is well documented.
</sub110><strat VERBATIM><duration 5.16><TC respecting>

<sub 111>
What is less well -known is that <w CJT> <LINEBREAKERR>
prominent Nazis <CONDENSATION> believe ~~believed~~ </CONDENSATION> that <w
CJT> the <w AT0> <LINEBREAKERR>
master race <w NN> originated in Atlantis.
</sub111><strat CONDENSATION><duration 7.08><TC not respecting><3-liner>

~~One of the most passionate believers was Heinrich Himmler, Head of the SS. Himmler directed Germany scientists to seek the descendants of the Atlantian super-race in places from the Andes to Tibet. They scrutinised the physical features of the natives in search of any shred of evidence to support Himmler's notion that his Aryan ancestors, the Atlantians, had lived there.~~

<sub 112>
These claims to <w PRP> <LINEBREAKERR>
an <w AT0> ancestral heritage <w NN> in Atlantis
</sub112><strat VERBATIM><duration 4.18><TC respecting>

<sub 113>
fed the <ERRCORRECTION> Nazis' ~~Nazis~~ </ERRORCORRECTION> belief in <w PRP>
the <w AT0> <LINEBREAKERR>
supremacy <w NN> of the Aryan master race.
</sub113><strat VERBATIM+ERRCORRECTION><duration 5.23><TC respecting>

KEN FEDER:
<sub 114>
When we come to something like <w PRP> <LINEBREAKERR>
the <w AT0> lost continent <w NN> of Atlantis,
</sub114><strat VERBATIM><duration 5.00><TC not respecting>

<sub 115>
we <CONDENSATION> 're ~~are~~ </CONDENSATION> better off knowing
that civilisations
developed <DELETION> ~~more or less~~ </DELETION> independently
</sub115><strat CONDENSATION, DELETION><duration 5.22><TC not respecting><3-liner>

<sub 116>
just so nobody can say some people <LINEBREAKERR>
are better than others ... ~~,~~
Some ~~some~~ are smarter than others ...
</sub116><strat VERBATIM><duration 6.24><TC not respecting><3-liner>

<sub 117>
<DELETION> ~~because~~ </DELETION> We ~~we~~ know what happens down the line
when we believe that . ~~,~~
</sub117><strat DELETION><duration 4.13><TC not respecting>

<sub 118>
<DELETION> ~~so~~ </DELETION> I'm not going to <CONDENSATION> say ~~tell you~~ </CONDENSATION> that belief
in Atlantis is necessarily the <w AT0> first <w AJ0> <LINEBREAKERR>
step <w NN> towards genocide ~~,~~ or holocaust ~~Holocaust~~ . ~~,~~
</sub118><strat DELETION, CONDENSATION><duration 7.14><TC not respecting><3-liner>

<sub 119>
But ~~but~~ <DELETION> ~~what I'm telling you is~~ </DELETION> we are on <w PRP> a <w AT0> very <w AV0> slippery <w AJ0> <LINEBREAKERR>
slope <w NN> if we believe in fantasies.
</sub119><strat DELETION><duration 5.12><TC respecting>

<sub 120>
<DELETION> ~~and~~ </DELETION> <DELETION> ~~that~~ </DELETION> Fantasies ~~fantasies~~
lead us <DELETION> ~~down~~ </DELETION> to places
we <DELETION> ~~really~~ </DELETION> don't want to go.
</sub120><strat DELETION, DELETION, DELETION, DELETION, DELETION><duration 4.22><TC respecting>

~~VOICE: Next week in the second of this two-part special Horizon examines the controversial theory of best-selling author Graham Hancock which is challenging mainstream archaeology.~~

~~KEN FEDER: Do you believe in Atlantis?~~

HORIZON: Atlantis Uncovered

Appendix 2: Research Corpus – *Chimps on Death Row*

Transcript File

NARRATOR (JACK FORTUNE): The chimpanzees at the Holloman Air Force Base have been used in research for 40 years. Now they're forcing scientists to face the consequences of experimenting on our closest living relatives.

DR. TOM INSEL (YERKES PRIMATE RESEARCH CENTRE): There's no question that the use of chimpanzees presents a greater moral dilemma than the use of any other non-human animal for research.

DR JIM MAHONEY (LABORATORY FOR EXPERIMENTAL MEDICINE & SURGERY IN PRIMATES 1977-98): I now believe that we have no actual right to use animals, but I still can't deny that we have a need to use them because I can't see any alternative.

PROF. ROGER FOUTS CHIMPANZEE & HUMAN COMMUNICATION INSTITUTE): I hope some day that, that when we look back on this time we're, we're keeping chimpanzees in captivity against their will and with the same shame that I look back on my great-grandfather who was a slaver holder and I hope that some day my great-grand-children will look back on us with shame.

NARRATOR: Chimpanzees are special animals. As apes, they share features with humans which set them apart from monkeys and the rest of the animal kingdom.

DR. JANE GOODALL: These beings are so like us they share 98.6% of our DNA, our genetic material. They have brains and central nervous systems more like ours than that of any other living creature.   They show characteristics like altruism, compassion and so forth and in addition, without any question, they have emotions like joy and sorrow, fear and no, mental as well as physical suffering.

NARRATOR: The chimpanzees' similarity to humans has made them useful to science, but it has also condemned some of them to an awful fate. At this primate centre in Holland, 130 chimps are held for use by researchers from around Europe, including Britain. They're used to study a few deadly human diseases which infect no other animal.

DR RONALD BONTROP (BIOMEDICAL PRIMATE RESEARCH CENTRE): You're only allowed to use chimpanzees to study diseases which are a plague or a pest of the human kind so we use chimpanzees mainly for research into fields of HIV, AIDS, malaria and hepatitis research of course.

NARRATOR: For the chimps these human diseases are rarely fatal and unlike any other animal used in research, nowadays chimpanzees are never deliberately killed.

RONALD BONTROP: In this Institute we have the policy not to engage in experiments whereon beforehand it is known that the chimpanzee may die within the experiments.   We simply say no to these type of experiments.   A chimpanzee will reach the honourable age of 50 years in its life and we also have the policy not to euthanise these chimpanzees.

NARRATOR: Because they're not euthanised the old animals are trapped behind bars, retired from research until they die of old age.   This is the story of what chimpanzees have endured in the name of science and what future awaits the thousands of survivors.   The Holloman Air Force chimp colony in New Mexico was one of the first in the world.   In the early 50s a zoo of dogs, pigs and bears was assembled to test the effects of high altitude flight on a mammal's body.   As the best physical model for humans the Air Force chose chimpanzees.

DR JERRY FINEG (HOLLOMAN AIR FORCE BASE 1957-1962): My responsibilities

[PART 2]

DR MIKE KEELING (UNIVERSITY OF TEXAS CANCER CENTRE): ... We'd used them in just about everything that, that you can use a chimp for.

JIM MAHONEY: There was a surplus developing of animals that had been exposed to hepatitis B so they were sitting in the colonies by increasing numbers eating the colonies away with no supporting money.

NARRATOR: The surplus raised an awful question: was it acceptable to kill the chimps they didn't need?

MIKE KEELING: I guess at that point in time there was consideration of euthanasia as a population control method.

NARRATOR: But the moral debate was cut short by the arrival of a deadly new disease.

JIM MAHONEY: AIDS saved the chimp.

NARRATOR: Nothing like AIDS had ever been seen before.   HIV killed by directly attacking the body's white blood cells. No-one knew how it was spread or how to stop it. The researchers were desperate for an animal model they could use to study HIV.

TOM INSEL: We were looking for any animal model that might work.   As it turns out the only animal that's known that gets infected with HIV 1, which is the form of the virus that causes AIDS in humans, was the chimpanzee.

NARRATOR: Scientists were soon queuing up to inject chimps with HIV. Far from having a surplus, primate centres around the world stepped up their breeding programmes to compensate for all the animals they thought would soon be dying of AIDS.

MIKE KEELING: When we first started putting the material in the animal we were continuously looking for disease. The disease didn't come, you know, for 2 and 3 and 4 years, but as we know it's a very latent infection and what we had to consider was the fact that it may lay dormant for years before it actually caused infection in the chimp, so probably for 5 or 6 years people were just thinking well, we just haven't had enough time, but once we reached that time there was the opinion that more than likely this isn't going to be a disease model.

NARRATOR: There are now 200 HIV-positive chimps around the world and only one has ever developed full-blown AIDS. They are still a health hazard to humans and have to be housed separately from other, clean chimps. The Biomedical Primate Research Centre in Holland was one of the first to begin AIDS work with chimps.

RONALD BONTROP: You see one of our HIV-infected chi, chimpanzees. His named is called Koo. He was infected in 1982, one of the first chimps that received the serum sample from an AIDS patient from America. This chimp is still completely healthy, although he was infected about 16 years ago.

NARRATOR: It's possible that Koo and the other HIV chimps here will develop AIDS and die. If they don't, they could be in this bio containment housing for another 20 years. When it's warm enough and not raining the roof folds back to give them a glimpse of the outside world. AIDS may have meant a stay of execution, but now the chimps face a life sentence.

TOM INSEL: ~~Well I think we blew it. I think as a, as a scientific community we made a bad decision some years ago, but we made it under duress at the time when the AIDS epidemic was really raging in the United States and there was a bit of a scientific panic about how we were going to control this.~~

NARRATOR: ~~Just in time for the arrival of AIDS Booee was on the move again. By the early 80s the initial excitement about teaching chimps sign language had faded and the American government withdrew its funding. For 13-year-old Booee it meant going back to the harsh world of biomedical research, this time to Jim Mahoney's lab.~~

JIM MAHONEY: ~~Booee arrived along with almost 40 other chimps. He seemed to settle in to LEMSIP. Obviously there was an enormous change in his lifestyle, and he became just one of the chimps who was going to be used in hepatitis research and eventually, of course, he was supposed to have been used in AIDS.~~

NARRATOR: ~~At LEMSIP Booee was put in a cage on his own, but this was one of the first centres to address the boredom that chimps faced in solitary confinement.~~

JIM MAHONEY: ~~It was truly quite spectacular to see the way they would pick up a little plastic mirror and a comb or a, a tooth-brush and use the mirror to examine parts of their body and you'd find them cleaning their teeth and doing their eyes and just amazing, even down to playing a, a mouth organ and I remember there was one chimp in particular was, was incredible on the harmonica and he'd even hit it every now and again to knock the spit out, you know. Sometimes these chimps will take to human activities as though they'd always done them. Every now and again what we called a circus and that's when there were enough people that were free that we could put on a little party for the animals that would last maybe an hour and we'd blow up balloons and bring in all sorts of games and we'd go in in jump-suits and we'd paint the white jump-suits different colours and patterns and we'd wear light-coloured wigs and we'd just play with the chimps and let them play with the balloons.~~

NARRATOR: ~~After 16 years at LEMSIP an American TV company heard about the sad story of Booee the signing chimp, now in biomedical research. They suggested he be reunited with the man who'd taught him sign language.~~

ROGER FOUTS: ~~And they said wouldn't this be a great reunion to, to get you out. Do you think he will remember you, and I didn't think he would and, and it had been 16 years. I'd changed, he'd changed. I went in, I pulled the mask down, he recognised me immediately and we started playing games that we hadn't played in 16 years.~~

~~Booee, how you doing Booee, how you doing? How you doing? Do you remember? What do you want, what do you want, what do you want? Booee, yes, Booee.~~

~~And not only did he recognise me but he remembered something I'd forgotten in that Booee had a nickname for him, for me and that he was such a fast signer, he was always very, very, I'd guess he'd be comparable of a fast talker, somebody who talks really, really fast, a jabber mouth or a jabber hands if you will, and his name sign was Booee and my name sign is Roger and what he would do is when he'd sign Booee or when he'd want us to do something rather than signing Booee/ Roger, he would take his name sign and just flick the top of his ear, it was like saying Rog or something, and I'd forgotten that. He was the only chimp that's ever called me that.~~

~~Do you remember Roger? Yes...~~

~~And he started doing this and suddenly I realised my gosh, that's my name, he remembered my name, and it was great fun. I mean it was like no time had passed and all, he was the same little boy even though he's probably 140/150lbs now. What gave me nightmares after that was when I told him I had to leave now. I said I have to leave now, and rather than screaming or baying to come or to go he accepted it, but he accepted it with such weight and sorrow and that you could physically see him in this shot just, it was like somebody put a huge weight on his back and he was crushed and he moved to the back of his cage.~~

NARRATOR: In the 40 years they've been used in research the lives of captive chimps have been proved.

[PART 3]

NARRATOR: ... Bids to support the chimps came from groups wanting to retire them, and also from the Coulston Research Foundation.   Their founder is totally against retiring these chimpanzees.
DR FRED COULSTON (THE COULSTON FOUNDATION): They are like human beings that are kept in a hospital, in a clinic as they get older.   We will keep them in the same way, we will treat them with, for their diseases, we will learn from them how to treat human diseases of comparable age groups.   That's the true sanctuary. To just take these valuable animals and throw 'em out in a field in, in South Texas.

NARRATOR: Dr. Coulston wants to use the ageing survivors of the space programme and their descendants to study prostate cancer, diabetes and any other human diseases that chimps might develop in their old age.

FRED COULSTON: The point is these are animals that have value throughout their life for the medical understanding of diseases of old age.

NARRATOR: ~~Last month the Air Force announced the majority of the chimps would be handed over to Dr. Coulston.~~   Funds couldn't be raised to build a retirement sanctuary. Neither the drugs companies who'd used them nor the government who founded the colony, nor the public were willing to provide the money.   A lucky few have escaped from research. At the wildlife way station outside Los Angeles one of the latest arrivals is Booee, now 32 years old.

JIM MAHONEY   This is a typical chimpanzee response to strangers coming into his territory and I hope you aren't too offended by this, but that is a normal thing for a chimp to want to do.   It doesn't make him unnecessarily nasty or...   Look at him, he's laughing now, you see, you see.   There he was displaying just a moment ago and now he's laughing and now he's laughing with his head up and he wants to play tickle and chase.

NARRATOR: Jim Mahoney visits Booee regularly to check his health.   Booee escaped AIDS research, but he was infected with hepatitis so he has to be housed separately from the other chimps here.   His life at the way station depends on donations from the public.   After all he's been through, Booee's fate is still uncertain. He could live for another 20 years.   The arguments over the morality of experimenting on chimps continue, but both sides now agree. Those we have used deserve better treatment.

MIKE KEELING: I simply prioritise the advancement of human health and welfare at the cost of the chimp having to donate some experimental effort in their life, but what I think everyone agrees is if we're going to ask the chimp to do that we ought to be willing to back it up by when they've done their service we provide 'em a high quality life at a stabilised level of funding.

JANE GOODALL: Since chimpanzees are our closest living relatives, since in their own right they're wondrous and amazing beings, we should feel such a weight of guilt that we've

brought them from their native forests, we've subjected them to horrendous experimentation and we're not prepared to get together and sort out a decent future for them.

Subtitle File

*** Timecode: In: 10:1:0:1 Out: 10:1:7:9
*** Framecount: In: 901501 Out: 901684
The chimpanzees of the Holloman
Air Force Base have been used
in research for 40 years.

*** Timecode: In: 10:1:7:10 Out: 10:1:14:13
*** Framecount: In: 901685 Out: 901863
Now scientists are facing
the consequences of experimenting
on our closest living relatives.

*** Timecode: In: 10:1:14:14 Out: 10:1:19:12
*** Framecount: In: 901864 Out: 901987
The use of chimpanzees presents
a greater moral dilemma

*** Timecode: In: 10:1:19:13 Out: 10:1:23:16
*** Framecount: In: 901988 Out: 902091
than the use of any other animal
for research.

*** Timecode: In: 10:1:23:17 Out: 10:1:28:17
*** Framecount: In: 902092 Out: 902217
I now believe that we have
no actual right to use animals.

*** Timecode: In: 10:1:28:18 Out: 10:1:36:7
*** Framecount: In: 902218 Out: 902407
But I still can't deny
that we have a NEED to use them
because I can't see any alternative.

*** Timecode: In: 10:1:36:8 Out: 10:1:41:19
*** Framecount: In: 902408 Out: 902544
I hope we look back on this time
where we keep chimpanzees captive

*** Timecode: In: 10:1:41:20 Out: 10:1:46:20
*** Framecount: In: 902545 Out: 902670
with the shame that I feel for my
great-grandfather as a slaveholder.

*** Timecode: In: 10:1:46:21 Out: 10:1:51:9
*** Framecount: In: 902671 Out: 902784
My great-grandchildren
should look at us with shame.

*** Timecode: In: 10:2:8:21 Out: 10:2:15:6
*** Framecount: In: 903221 Out: 903381
Chimpanzees are special animals.
As apes, they share features
with humans

*** Timecode: In: 10:2:15:7 Out: 10:2:20:6
*** Framecount: In: 903382 Out: 903506
which set them apart from the rest
of the animal kingdom.

*** Timecode: In: 10:2:22:23 Out: 10:2:30:5
*** Framecount: In: 903573 Out: 903755
These beings are so like us
that they share 98.6% of our DNA,
our genetic material.

*** Timecode: In: 10:2:30:6 Out: 10:2:35:17
*** Framecount: In: 903756 Out: 903892
They have brains and central
nervous systems similar to ours.

*** Timecode: In: 10:2:35:18 Out: 10:2:41:23
*** Framecount: In: 903893 Out: 904048
They show characteristics like
altruism, compassion and so forth.

*** Timecode: In: 10:2:42:24 Out: 10:2:50:16
*** Framecount: In: 904074 Out: 904266
In addition, they have emotions
like joy, sorrow and fear, and know
mental AND physical suffering.

*** Timecode: In: 10:2:53:17 Out: 10:3:0:5
*** Framecount: In: 904342 Out: 904505
The chimpanzee's similarity
to humans has made them useful
to science,

*** Timecode: In: 10:3:0:6 Out: 10:3:5:3
*** Framecount: In: 904506 Out: 904628
but it has also condemned
some of them to an awful fate.

*** Timecode: In: 10:3:5:4 Out: 10:3:10:0
*** Framecount: In: 904629 Out: 904750
At this primate centre in Holland,
130 chimps are held for use

*** Timecode: In: 10:3:10:1 Out: 10:3:14:13
*** Framecount: In: 904751 Out: 904863
by researchers from around Europe,
including Britain.

*** Timecode: In: 10:3:16:2 Out: 10:3:22:19
*** Framecount: In: 904902 Out: 905069
They are used to study
a few deadly human diseases
which infect no other animal.

*** Timecode: In: 10:3:22:20 Out: 10:3:28:9
*** Framecount: In: 905070 Out: 905209
We're only allowed to use
chimpanzees to study diseases

*** Timecode: In: 10:3:28:10 Out: 10:3:32:10
*** Framecount: In: 905210 Out: 905310
which are a plague or a pest
to humankind.

*** Timecode: In: 10:3:32:11 Out: 10:3:36:23
*** Framecount: In: 905311 Out: 905423
So we use chimpanzees
mainly for research in the fields

*** Timecode: In: 10:3:36:24 Out: 10:3:41:8
*** Framecount: In: 905424 Out: 905533
of HIV, AIDS, malaria and hepatitis,
of course.

*** Timecode: In: 10:3:44:9 Out: 10:3:49:2
*** Framecount: In: 905609 Out: 905727
For the chimps, these human
diseases are rarely fatal,

*** Timecode: In: 10:3:49:3 Out: 10:3:56:19
*** Framecount: In: 905728 Out: 905919
and unlike any other animal used
in research, nowadays chimpanzees
are never deliberately killed.

*** Timecode: In: 10:4:1:15 Out: 10:4:6:15
*** Framecount: In: 906040 Out: 906165
In this institute, we have the
policy not to engage in experiments

*** Timecode: In: 10:4:6:16 Out: 10:4:11:16
*** Framecount: In: 906166 Out: 906291
where it is known that the chimp
may die within the experiment.

*** Timecode: In: 10:4:11:17 Out: 10:4:15:23
*** Framecount: In: 906292 Out: 906398
We simply say no
to these types of experiments.

*** Timecode: In: 10:4:15:24 Out: 10:4:20:24
*** Framecount: In: 906399 Out: 906524
The chimpanzee will reach
the honourable age of 50 in its life

*** Timecode: In: 10:4:21:0 Out: 10:4:24:12
*** Framecount: In: 906525 Out: 906612
and we do not euthanise
these chimpanzees.

*** Timecode: In: 10:4:26:1 Out: 10:4:33:1
*** Framecount: In: 906651 Out: 906826
Because they're not euthanised,
the old animals are trapped
behind bars, retired from research

*** Timecode: In: 10:4:33:2 Out: 10:4:35:15
*** Framecount: In: 906827 Out: 906890
until they die of old age.

*** Timecode: In: 10:4:36:21 Out: 10:4:43:5
*** Framecount: In: 906921 Out: 907080
This is the story
of what chimpanzees have endured
in the name of science

*** Timecode: In: 10:4:43:6 Out: 10:4:46:19
*** Framecount: In: 907081 Out: 907169
and what future awaits
the survivors.

*** Timecode: In: 10:4:52:0 Out: 10:4:58:13
*** Framecount: In: 907300 Out: 907463
The Holloman Air Force chimp colony
in New Mexico
was one of the first in the world.

*** Timecode: In: 10:5:2:23 Out: 10:5:5:21
*** Framecount: In: 907573 Out: 907646
In the early '50s,

*** Timecode: In: 10:5:5:22 Out: 10:5:13:6
*** Framecount: In: 907647 Out: 907831
a zoo of dogs, pigs and bears
was assembled to test the effects
of high-altitude flight.

*** Timecode: In: 10:5:13:7 Out: 10:5:18:20
*** Framecount: In: 907832 Out: 907970
As the best physical model
for humans, they chose chimpanzees.

*** Timecode: In: 10:5:20:11 Out: 10:5:23:3
*** Framecount: In: 908011 Out: 908078
My responsibilities,

[PART 2]

*** Timecode: In: 10:22:29:19 Out: 10:22:34:4
*** Framecount: In: 933744 Out: 933854
We used them in everything
you can use a chimp for.

*** Timecode: In: 10:22:34:5 Out: 10:22:39:4
*** Framecount: In: 933855 Out: 933979
There was a surplus of animals that
had been exposed to hepatitis B,

*** Timecode: In: 10:22:39:5 Out: 10:22:43:17
*** Framecount: In: 933980 Out: 934092
so they were sitting in the colonies
in increasing numbers,

*** Timecode: In: 10:22:43:18 Out: 10:22:47:19
*** Framecount: In: 934093 Out: 934194
eating the colonies away
with no supporting money.

*** Timecode: In: 10:22:47:20 Out: 10:22:54:24
*** Framecount: In: 934195 Out: 934374
The surplus raised an awful
question - was it acceptable to
kill the chimps they didn't need?

*** Timecode: In: 10:22:55:0 Out: 10:23:0:8
*** Framecount: In: 934375 Out: 934508
Euthanasia was considered
as a population control method.

*** Timecode: In: 10:23:0:9 Out: 10:23:6:22
*** Framecount: In: 934509 Out: 934672
But the moral debate was cut short
by the arrival
of a deadly new disease.

*** Timecode: In: 10:23:8:6 Out: 10:23:11:6
*** Framecount: In: 934706 Out: 934781
AIDS saved the chimp.

*** Timecode: In: 10:23:11:7 Out: 10:23:15:7
*** Framecount: In: 934782 Out: 934882
Nothing like AIDS
had ever been seen before.

*** Timecode: In: 10:23:15:8 Out: 10:23:20:8
*** Framecount: In: 934883 Out: 935008
HIV killed by directly attacking
the body's white blood cells.

*** Timecode: In: 10:23:20:9 Out: 10:23:24:5
*** Framecount: In: 935009 Out: 935105
No-one knew how it was spread
or how to stop it.

*** Timecode: In: 10:23:24:6 Out: 10:23:30:19
*** Framecount: In: 935106 Out: 935269
The researchers were desperate
for an animal model they could use
to study HIV.

*** Timecode: In: 10:23:31:20 Out: 10:23:36:7
*** Framecount: In: 935295 Out: 935407
We were looking for
any animal model that might work.

*** Timecode: In: 10:23:36:8 Out: 10:23:41:8
*** Framecount: In: 935408 Out: 935533
The only animal that's known
that gets infected with HIV 1,

*** Timecode: In: 10:23:41:9 Out: 10:23:45:8
*** Framecount: In: 935534 Out: 935633
which causes AIDS in humans,
was the chimpanzee.

*** Timecode: In: 10:23:46:9 Out: 10:23:51:9
*** Framecount: In: 935659 Out: 935784
Scientists were soon queueing up
to inject chimps with HIV.

*** Timecode: In: 10:23:51:10 Out: 10:23:57:17
*** Framecount: In: 935785 Out: 935942
Far from having a surplus,
primate centres stepped up
their breeding programmes

*** Timecode: In: 10:23:57:18 Out: 10:24:2:18
*** Framecount: In: 935943 Out: 936068
to compensate for the animals they
thought would soon be dying of AIDS.

*** Timecode: In: 10:24:2:19 Out: 10:24:9:7
*** Framecount: In: 936069 Out: 936232
When we first put material
in the animal, we were continuously
looking for disease.

*** Timecode: In: 10:24:9:8 Out: 10:24:15:21
*** Framecount: In: 936233 Out: 936396
The disease didn't come
for two, three, four years,
but it's a very latent infection.

*** Timecode: In: 10:24:15:22 Out: 10:24:22:22
*** Framecount: In: 936397 Out: 936572
We had to consider that it may
lay dormant for years before
causing infection in the chimp.

*** Timecode: In: 10:24:22:23 Out: 10:24:27:23
*** Framecount: In: 936573 Out: 936698
For five or six years, people just
thought we hadn't had enough time.

*** Timecode: In: 10:24:27:24 Out: 10:24:34:14
*** Framecount: In: 936699 Out: 936864
Once we reached that time,
the opinion was that
this wouldn't be a disease model.

*** Timecode: In: 10:24:37:15 Out: 10:24:42:15
*** Framecount: In: 936940 Out: 937065
There are now 200 HIV-positive
chimps around the world

*** Timecode: In: 10:24:42:16 Out: 10:24:46:16
*** Framecount: In: 937066 Out: 937166
and only one has ever developed
full-blown AIDS.

*** Timecode: In: 10:24:46:17 Out: 10:24:51:23
*** Framecount: In: 937167 Out: 937298
They are a health hazard and are
housed separately from other chimps.

*** Timecode: In: 10:24:51:24 Out: 10:24:56:7
*** Framecount: In: 937299 Out: 937407
The Biomedical Primate
Research Centre in Holland

*** Timecode: In: 10:24:56:8 Out: 10:25:0:21
*** Framecount: In: 937408 Out: 937521
was one of the first
to begin AIDS work with chimps.

*** Timecode: In: 10:25:2:21 Out: 10:25:7:10
*** Framecount: In: 937571 Out: 937685
You see one
of our HIV-infected chimpanzees.

*** Timecode: In: 10:25:7:11 Out: 10:25:12:8
*** Framecount: In: 937686 Out: 937808
He is called Koon.
He was infected in 1982,

*** Timecode: In: 10:25:12:9 Out: 10:25:17:1
*** Framecount: In: 937809 Out: 937926
one of the first chimps
that received a serum sample

*** Timecode: In: 10:25:17:2 Out: 10:25:19:20
*** Framecount: In: 937927 Out: 937995
from an AIDS patient from America.

*** Timecode: In: 10:25:19:21 Out: 10:25:23:16
*** Framecount: In: 937996 Out: 938091
This chimp is still
completely healthy,

*** Timecode: In: 10:25:23:17 Out: 10:25:27:17
*** Framecount: In: 938092 Out: 938192
although he was infected
about 16 years ago.

*** Timecode: In: 10:25:27:18 Out: 10:25:34:5
*** Framecount: In: 938193 Out: 938355
It is possible that Koon
and the other HIV chimps here
will develop AIDS and die.

*** Timecode: In: 10:25:34:6 Out: 10:25:36:16
*** Framecount: In: 938356 Out: 938416
If they don't,

*** Timecode: In: 10:25:36:17 Out: 10:25:42:13
*** Framecount: In: 938417 Out: 938563
they could be
in this bio-containment housing
for another 20 years.

*** Timecode: In: 10:25:42:14 Out: 10:25:48:6
*** Framecount: In: 938564 Out: 938706
If it's warm, the roof folds back
and they can see the outside world.

*** Timecode: In: 10:25:48:7 Out: 10:25:53:7
*** Framecount: In: 938707 Out: 938832
AIDS has meant a stay of execution,
but the chimps face a life sentence.

*** Timecode: In: 10:28:13:4 Out: 10:28:20:4
*** Framecount: In: 942329 Out: 942504
In the 40 years they've been used
in research, the lives
of captive chimps have improved.

[PART 3]

*** Timecode: In: 10:41:8:14 Out: 10:41:16:14
*** Framecount: In: 961714 Out: 961914
Bids to help the chimps came from
groups wanting to retire them and
the Coulston Research Foundation.

*** Timecode: In: 10:41:16:15 Out: 10:41:21:19
*** Framecount: In: 961915 Out: 962044
Their founder is totally against
retiring these chimpanzees.

*** Timecode: In: 10:41:21:20 Out: 10:41:26:20
*** Framecount: In: 962045 Out: 962170
They are like human beings kept
in a hospital as they get older.

*** Timecode: In: 10:41:26:21 Out: 10:41:31:21
*** Framecount: In: 962171 Out: 962296
We will keep them in the same way.
We'll treat them for their diseases.

*** Timecode: In: 10:41:31:22 Out: 10:41:38:10
*** Framecount: In: 962297 Out: 962460
We will learn from them
how to treat human diseases
of comparable age groups.

*** Timecode: In: 10:41:38:11 Out: 10:41:40:23
*** Framecount: In: 962461 Out: 962523
That's the true sanctuary.

*** Timecode: In: 10:41:40:24 Out: 10:41:45:23
*** Framecount: In: 962524 Out: 962648
To take these valuable animals and
throw them out in a field in Texas?!

*** Timecode: In: 10:41:45:24 Out: 10:41:53:6
*** Framecount: In: 962649 Out: 962831
Dr Coulston wants to use the ageing
survivors of the space programme
and their descendants

*** Timecode: In: 10:41:53:7 Out: 10:42:0:20
*** Framecount: In: 962832 Out: 963020
to study prostate cancer, diabetes
and any other human diseases that
chimps may develop in their old age.

*** Timecode: In: 10:42:0:21 Out: 10:42:4:24
*** Framecount: In: 963021 Out: 963124
The point is...
these are animals

*** Timecode: In: 10:42:5:0 Out: 10:42:12:20
*** Framecount: In: 963125 Out: 963320
that have value throughout their
life for the medical understanding
of diseases of old age.

*** Timecode: In: 10:42:12:21 Out: 10:42:17:9
*** Framecount: In: 963321 Out: 963434
Funds couldn't be raised
to build a retirement sanctuary.

*** Timecode: In: 10:42:17:10 Out: 10:42:24:2
*** Framecount: In: 963435 Out: 963602
Neither the drugs companies,
the government nor the public
would provide the money.

*** Timecode: In: 10:42:26:9 Out: 10:42:31:12
*** Framecount: In: 963659 Out: 963787
A lucky few have escaped
from research.

*** Timecode: In: 10:42:31:13 Out: 10:42:36:9
*** Framecount: In: 963788 Out: 963909
At the Wildlife Waystation
outside Los Angeles,

*** Timecode: In: 10:42:36:10 Out: 10:42:40:22
*** Framecount: In: 963910 Out: 964022
one of the latest arrivals is Booee,
now 32 years old.

*** Timecode: In: 10:42:55:17 Out: 10:43:2:5
*** Framecount: In: 964392 Out: 964555
This is a typical male chimpanzee
response to strangers
coming into his territory.

*** Timecode: In: 10:43:2:6 Out: 10:43:8:19
*** Framecount: In: 964556 Out: 964719
I hope you are not offended,
but that is a normal thing
for a chimp to want to do.

*** Timecode: In: 10:43:8:20 Out: 10:43:12:20
*** Framecount: In: 964720 Out: 964820
It doesn't make him
unnecessarily nasty.

*** Timecode: In: 10:43:12:21 Out: 10:43:16:23
*** Framecount: In: 964821 Out: 964923
Look at him, he's laughing now.
You see?

*** Timecode: In: 10:43:16:24 Out: 10:43:21:24
*** Framecount: In: 964924 Out: 965049
He was displaying just a moment ago.
Now he's laughing with his head up.

*** Timecode: In: 10:43:22:0 Out: 10:43:24:18
*** Framecount: In: 965050 Out: 965118
He wants to play tickle and chase.

*** Timecode: In: 10:43:24:19 Out: 10:43:29:7
*** Framecount: In: 965119 Out: 965232
Jim Mahoney visits Booee regularly
to check his health.

*** Timecode: In: 10:43:29:8 Out: 10:43:34:6
*** Framecount: In: 965233 Out: 965356
Booee escaped AIDS research,
but was infected with hepatitis,

*** Timecode: In: 10:43:34:7 Out: 10:43:37:20
*** Framecount: In: 965357 Out: 965445
so is housed separately
from other chimps.

*** Timecode: In: 10:43:39:21 Out: 10:43:44:22
*** Framecount: In: 965496 Out: 965622
His life at the Waystation depends
on donations from the public.

*** Timecode: In: 10:43:47:16 Out: 10:43:52:16
*** Framecount: In: 965691 Out: 965816
After all he's been through,
Booee's fate is still uncertain.

*** Timecode: In: 10:43:52:17 Out: 10:43:55:5
*** Framecount: In: 965817 Out: 965880
He could live for another 20 years.

*** Timecode: In: 10:43:55:6 Out: 10:44:0:15
*** Framecount: In: 965881 Out: 966015
The arguments over the morality
of experimenting on chimps continue,

*** Timecode: In: 10:44:0:16 Out: 10:44:3:4
*** Framecount: In: 966016 Out: 966079
but both sides now agree

*** Timecode: In: 10:44:3:5 Out: 10:44:7:19
*** Framecount: In: 966080 Out: 966194
that those we have used
deserve better treatment.

*** Timecode: In: 10:44:7:20 Out: 10:44:12:20
*** Framecount: In: 966195 Out: 966320
I prioritise the advancement
of human health and welfare

*** Timecode: In: 10:44:12:21 Out: 10:44:19:9
*** Framecount: In: 966321 Out: 966484
at the cost of the chimp having
to donate some experimental effort
in their life.

*** Timecode: In: 10:44:19:10 Out: 10:44:23:17
*** Framecount: In: 966485 Out: 966592
But everyone agrees
that if we ask the chimp to do that,

*** Timecode: In: 10:44:23:18 Out: 10:44:27:18
*** Framecount: In: 966593 Out: 966693
we must back it up
when they've done their service

*** Timecode: In: 10:44:27:19 Out: 10:44:32:19
*** Framecount: In: 966694 Out: 966819
by giving them a high quality life
at a stable level of funding.

*** Timecode: In: 10:44:32:20 Out: 10:44:40:8
*** Framecount: In: 966820 Out: 967008
Since chimpanzees are our closest
living relatives, since they're
wondrous and amazing beings,

*** Timecode: In: 10:44:40:9 Out: 10:44:47:4
*** Framecount: In: 967009 Out: 967179
we should feel such a weight
of guilt that we've brought them
from their native forests,

*** Timecode: In: 10:44:47:5 Out: 10:44:51:5
*** Framecount: In: 967180 Out: 967280
we've subjected them
to horrendous experimentation

*** Timecode: In: 10:44:51:6 Out: 10:44:55:21
*** Framecount: In: 967281 Out: 967396
and we are not prepared to sort out
a decent future for them.

Merge with annotation

NARRATOR (JACK FORTUNE):
<sub1>
The chimpanzees <PARAPHRASE> of at </PARAPHRASE> the Holloman <w NP041>
<LINEBREAKERR>
Air <w NP042> Force <w NP043> Base <w NP044> have been used
in research for 40 years.
</sub1><strat PARAPHRASE><duration 7.08><TC respecting><3-liner>

<sub2>
Now <CONDENSATION> they're forcing scientists are facing to face
</CONDENSATION>
the consequences of experimenting
on our closest living relatives.
</sub2><strat CONDENSATION><duration 7.03><TC not respecting><3-liner>

DR. TOM INSEL (YERKES PRIMATE RESEARCH CENTRE):
<sub3>
<DELETION> There's no question </DELETION> <DELETION> that </DELETION>
The the use of chimpanzees presents <w VVZ> <LINEBREAKERR>
a <w DT0> greater moral dilemma <w NN-obj>
</sub3><strat DELETION, DELETION><duration 4.14><TC respecting>

<sub4>
than the use of any other <DELETION> non-human </DELETION> animal
for research.
</sub4><strat DELETION><duration 4.03><TC respecting>

DR JIM MAHONEY (LABORATORY FOR EXPERIMENTAL MEDICINE & SURGERY
IN PRIMATES 1977-98):
<sub5>
I now believe that we have
no actual right to use animals. ,
</sub5><strat VERBATIM><duration 5><TC respecting>

<sub6>
But but I still can't deny
that we have a NEED need to use them
because I can't see any alternative.
</sub6><strat VERBATIM><duration 7.14><TC respecting><3-liner>

PROF. ROGER FOUTS CHIMPANZEE & HUMAN COMMUNICATION INSTITUTE):
</sub7>
I hope <DELETION> some day </DELETION> <DELETION> that <DELETION>,
<DELETION>that </DELETION> <DELETION> when </DELETION> we look back on
this time
<ADDITION> where </ADDITION> we <CONDENSATION>keep 're, <DELETION>
we're<DELETION> keeping </CONDENSATION> chimpanzees <CONDENSATION>
captive in captivity </CONDENSATION> <DELETION> against their will </DELETION>
</sub7><strat DELETION, DELETION, DELETION, DELETION, ADDITION,
CONDENSATION, DELETION, CONDENSATION, DELETION><duration 5.11><TC
respecting>

</sub8>
<DELETION> and </DELETION> with the <DELETION> same </DELETION> shame
that I <CONDENSATION> feel for <w PRP> look back on </CONDENSATION> my <w
DPS> <LINEBREAKERR>
great-grandfather <w NN> <CONDENSATION> as who was </CONDENSATION> a
<ALTSP+CORRECTION> slaveholder slaver holder </ALTSP+CORRECTION>.
</sub8><strat DELETION, DELETION, CONDENSATION, CONDENSATION><duration
5><TC not respecting>

</sub9>
<DELETION> and </DELETION> <CONDENSATION>I hope <DELETION> that
</DELETION> <DELETION> some day </DELETION> My my <ALTSP> great-
grandchildren great-grand-children </ALTSP>
should will look </CONDENSATION> <CONDENSATION> at back on
</CONDENSATION> us with shame.
</sub9><strat DELETION, CONDENSATION, DELETION, DELETION,
CONDENSATION><duration 4.13><TC respecting>

NARRATOR:
</sub10>
Chimpanzees are special animals.
As apes, they share features
with humans
</sub10><strat VERBATIM><duration 6.10><TC respecting><3-liner>

<sub11>
which set them apart from <DELETION> monkeys </DELETION> <DELETION> and
</DELETION> the rest
of the animal kingdom.
</sub11><strat DELETION, DELETION><duration 4.24><TC respecting>

DR. JANE GOODALL:
<sub12>
These beings are so like us
<ADDITION> that </ADDITION> they share 98.6% of our DNA,
our genetic material.
</sub12><strat ADDITION><duration 7.07><TC respecting>

<sub13>
They have brains and central <w AJ0> <LINEBREAKERR>
nervous <w AJ0> systems <w NN> <DELETIONa> more </DELETIONa>
<PARAPHRASE> similar to like </PARAPHRASE> ours <DELETIONa> than that of any
other living creature </DELETIONa>.
</sub13><strat DELETION, PARAPHRASE><duration 5.11><TC
 respecting>

<sub14>
They show characteristics like <w PRP> <LINEBREAKERR>
altruism <w NN>, compassion and so forth.
</sub14><strat VERBATIM><duration 6.05><TC respecting>

<sub15>
<DELETION> and </DELETION> In in addition, <DELETION> without any question,
</DELETION> they have emotions
like joy. <TRANSPOSITION> and sorrow and </TRANSPOSITION>, fear, and
<ERRCOR> know no </ERRCOR> ,
mental <CONDENSATION> AND as well as </CONDENSATION>physical suffering.
</sub15><strat DELETION, DELETION, TRANSPOSITION,
CONDENSATION><duration 7.17><TC not respecting>

NARRATOR:
<sub16>
The <ERROR> chimpanzee's chimpanzees' </ERROR> similarity
to humans has made them useful
to science,
</sub16><strat VERBATIM+ERROR><duration 6.13><TC respecting><3-liner>

<sub17>
but it has also condemned
some of them to an awful fate.
</sub17><strat VERBATIM><duration 4.22><TC respecting>

<sub18>
At this primate centre in Holland,
130 chimps are held for use
</sub18><strat VERBATIM><duration 4.21><TC not respecting>

\<sub19>
by researchers from around Europe,
including Britain.
\</sub19>\<strat VERBATIM>\<duration 4.12>\<TC not respecting>

\<sub20>
They \<SPECIFICATION> are 're \</SPECIFICATION> used to study \
a few deadly human diseases \
which infect no other animal.
\</sub20>\<strat SPECIFICATION>\<duration 6.17>\<TC respecting>\<3-liner>

DR RONALD BONTROP (BIOMEDICAL PRIMATE RESEARCH CENTRE):
\<sub21>
\<CONDENSATION> We You \</CONDENSATION> 're only allowed to \<w TO0> use \<w VVI> \<LINEBREAKERR> \
chimpanzees \<w NN-obj> to study diseases
\</sub21>\<strat CONDENSATION>\<duration 5.14>\<TC respecting>

\<sub22>
which are a plague or a pest \
\<PARAPHRASE> to of \</PARAPHRASE> \<DELETION> the \</DELETION> \<ALTSP> humankind human kind \</ALTSP>.
\</sub22>\<strat PARAPHRASE, DELETION>\<duration 4>\<TC respecting>

\<sub23>
So so we use chimpanzees \
mainly for research \<CONDENSATION> in into \</CONDENSATION> \<ADDITION> the \</ADDITION> fields \<LINEBREAKERR>
\</sub23>\<strat CONDENSATION, ADDITION>\<duration 4.12>\<TC not respecting>

\<sub24>
of HIV, AIDS, malaria and hepatitis \<DELETION> research \</DELETION>, \
of course.
\</sub24>\<strat DELETION>\<duration 4.09>\<TC respecting>

NARRATOR:
\<sub25>
For the chimps, these \<w DT0> human \<w AJ0> \<LINEBREAKERR>
diseases \<w NN> are rarely fatal ,
\</sub25>\<strat VERBATIM>\<duration 4.18>\<TC respecting>

\</sub26>
and unlike any other animal used
in research, nowadays chimpanzees
are never deliberately killed.
\</sub26>\<strat VERBATIM>\<duration 7.16>\<TC not respecting>\<3-liner>

RONALD BONTROP:
<sub27>
In this institute Institute , we have the <w AT0> <LINEBREAKERR>
policy <w NN> not to engage in experiments
</sub27><strat VERBATIM><duration 5><TC not respecting>

<sub28>
<CONDENSATION> where whereon </CONDENSATION> <DELETION> beforehand
</DELETION> it is known that the <CONDENSATION> chimp chimpanzee
</CONDENSATION>
may die within the <CONDENSATION> experiment experiments </CONDENSATION>.
</sub28><strat CONDENSATION, DELETION, CONDENSATION,
CONDENSATION><duration 5><TC not respecting>

<sub29>
We simply say no
to these types of experiments.
</sub29><strat VERBATIM><duration 4.06><TC respecting>

<sub30>
<PARAPHRASE> The A </PARAPHRASE> chimpanzee will reach
the honourable age of 50 <DELETION> years </DELETION> in its life
</sub30><strat PARAPHRASE, DELETION><duration 5><TC not respecting>

<sub31>
and we <DELETION> also </DELETION> <CONDENSATION do have the policy not to
</CONDENSATION> euthanize
these chimpanzees.
</sub31><strat DELETION, CONDENSATION><duration 3.12><TC not respecting>

NARRATOR:
<sub32>
Because they're not euthanised ,
the old animals are trapped
behind bars, retired from research
</sub32><strat VERBATIM><duration 7><TC not respecting>

<sub33>
until they die of old age.
</sub33><strat VERBATIM><duration 2.13><TC respecting>

<sub34>
This is the story
of what chimpanzees have endured
in the name of science
</sub34><strat VERBATIM><duration 6.09><TC respecting><3-liner>

<sub35>
and what future awaits
the <DELETION> ~~thousands of~~ </DELETION> survivors.
</sub35><strat DELETION><duration 3.13><TC respecting>

<sub36>
The Holloman Air Force chimp colony
in New Mexico
was one of the first in the world.
</sub36><strat VERBATIM><duration 6.13><TC not respecting><3-liner>

<sub37>
In the early <ERRCORRECTION> <u>'50s</u> ~~50s~~ </ERRCORRECTION><u>,</u>
</sub37><strat VERBATIM+ERRCORRECTION><duration 2.23><TC respecting>

<sub38>
a zoo of dogs, pigs and bears
was assembled to test the effects
of <ALTSP> high <u>-</u>altitude </ALTSP> flight <DELETION> ~~on a mammal's body~~
</DELETION>.
</sub38><strat DELETION><duration 7.09><TC respecting>

<sub39>
As the best physical model
for humans<u>,</u> <CONDENSATION> <u>they</u> ~~the Air Force~~ </CONDENSATION> chose
chimpanzees.
</sub39><strat CONDENSATION<duration 5.13><TC respecting>

DR JERRY FINEG (HOLLOMAN AIR FORCE BASE 1957-1962):
<sub40>
My responsibilities<u>,</u>
</sub40><strat VERBATIM><duration 2.17><TC respecting>

<sub41>
We <CONDENSATION> <u>'d</u> </CONDENSATION> used them in <DELETION> ~~just about~~
</DELETION> everything
<DELETION>~~that~~ </DELETION> <u>,</u><DELETION> ~~that~~ </DELETION> you can use a
chimp for.
</sub41><strat CONDENSATION, DELETION, DELETION, DELETION><duration
4.10><TC respecting>

JIM MAHONEY:
<sub42>
There was a surplus <DELETION> developing </DELETION> of animals that <w CJT> <LINEBREAKERR>
had been exposed to hepatitis B .
</sub42><strat DELETION><duration 4.24><TC not respecting>

<sub43>
so they were sitting in the colonies
<PARAPHRASE> in by </PARAPHRASE> increasing numbers .
</sub43><strat PARAPHRASE><duration 4.12><TC not respecting>

<sub44>
eating the colonies away
with no supporting money.
</sub44><strat VERBATIM><duration 4.01><TC not respecting>

NARRATOR:
<sub45>
The surplus raised an <w AT0> awful <w AJ0> <LINEBREAKERR>
question <w NN> – was it acceptable to <w TO0> <LINEBREAKERR>
kill <w VVI> the chimps they didn't need?
</sub45><strat VERBATIM><duration 7.04><TC not respecting><3-liner>

MIKE KEELING:
<sub46>
<DELETION> I guess </DELETION> <DELETION> at that point in time </DELETION>
<CONDENSATION> Euthanasia was considered there was consideration of euthanasia
</CONDENSATION>
as a population control method.
</sub46><strat DELETION, DELETION, CONDENSATION><duration 5.08><TC
respecting>

NARRATOR:
<sub47>
But the moral debate was cut short
by the arrival
of a deadly new disease.
</sub47><strat VERBATIM><duration 6.13><TC respecting><3-liner>

JIM MAHONEY:
<sub48>
AIDS saved the chimp.
</sub48><strat VERBATIM><duration 3><TC respecting>

NARRATOR:
<sub49>
Nothing like AIDS
had ever been seen before.
</sub49><strat VERBATIM><duration 4><TC respecting>

<sub50>
HIV killed by directly attacking
the body's white blood cells.
</sub50><strat VERBATIM><duration 5><TC not respecting>

<sub51>
No-one knew how it was spread
or how to stop it.
</sub51><strat VERBATIM><duration 3.21><TC not respecting>

<sub52>
The researchers were desperate
for an animal model they could use
to study HIV.
</sub52><strat VERBATIM><duration 6.13><TC respecting><3-liner>

TOM INSEL:
<sub53>
We were looking for
any animal model that might work.
</sub53><strat VERBATIM><duration 4.12><TC respecting>

<sub54>
<DELETION> As it turns out </DELETION> The the only animal that's known
that gets infected with HIV 1,
</sub54><strat DELETION><duration 5><TC respecting>

<sub55>
which <DELETION> is the form of the virus that </DELETION> causes AIDS in humans,
was the chimpanzee.
</sub55><strat DELETION><duration 3.24><TC not respecting>

NARRATOR:
<sub56>
Scientists were soon <ERROR> queueing queuing </ERROR> up
to inject chimps with HIV.
</sub56><strat VERBATIM+ERROR><duration 5><TC respecting>

<sub57>
Far from having a surplus,
primate centres <DELETION>~~around the world~~</DELETION> stepped up
their breeding programmes
</sub57><strat DELETION><duration 6.07><TC not respecting><3-liner>

<sub58>
to compensate for <DELETION> ~~all~~ </DELETION> the animals they_Subj
<LINEBREAKERR>
{thought would soon be dying of AIDS}_predicate.
</sub58><strat DELETION><duration 5><TC_not respecting>

MIKE KEELING:
<sub59>
When we first <CONDENSATION> put ~~started putting~~ </CONDENSATION>
<DELETION> ~~the~~ </DELETION> material
in the animal, we were <w VBD> continuously <w AV0> <LINEBREAKERR>
looking <w VVG> for disease.
</sub59><strat CONDENSATION, DELETION><duration 6.13><TC not respecting><3-
liner>

<sub60>
The disease didn't come, <DELETION> ~~you know~~ </DELETION>,
for two ~~2~~ <DELETION> ~~and~~ </DELETION> , three ~~3~~ <DELETION> ~~and~~ </DELETION> ,
four ~~4~~ years,
but <DELETION> ~~as we know~~ </DELETION> it's a very latent infection .
</sub60><strat DELETION, DELETION, DELETION, DELETION><duration 6.13><TC
not respecting><3-liner>

<sub61>
<DELETION> ~~and~~ </DELETION> <CONDENSATION> ~~what~~ had to consider ~~was~~
</CONDENSATION> <DELETION> ~~the fact~~ </DELETION> that it may <w VM0>
<LINEBREAKERR>
lay <w VVI> dormant for years before <LINEBREAKERR>
<CONDENSATION> causing ~~it~~<DELETION> ~~actually~~</DELETION> ~~caused~~
</CONDENSATION> infection in the chimp . ~~;~~.
</sub61><strat DELETION, CONDENSATION, DELETION, DELETION,
CONDENSATION><duration 7><TC not respecting>

<sub62>
<DELETION> ~~so~~ </DELETION> <DELETION> ~~probably~~ </DELETION> For ~~for~~ five ~~5~~
or six ~~6~~ years, people_Subj <CONDENSATION> ~~were~~just <LINEBREAKERR>
{thought ~~thinking~~ </CONDENSATION> <DELETION> ~~well~~ </DELETION> , we
<DELETION> ~~just~~ </DELETION> <CONDENSATION> hadn't ~~haven't~~ had
</CONDENSATION> enough time}_predicate . ~~;~~
</sub62><strat DELETION, DELETION, CONDENSATION, DELETION, DELETION,
CONDENSATION><duration 5><TC not respecting>

<sub63>
<DELETION> but </DELETION> Once once we reached that time,
<CONDENSATION> there was the opinion was </CONDENSATION> that <w CJT>
<LINEBREAKERR>
<DELETION> more than likely </DELETION> this <CONDENSATION> wouldn't isn't going to be </CONDENSATION> a disease model.
</sub63><strat DELETION, CONDENSATION, DELETION, CONDENSATION><duration 6.15><TC not respecting>

NARRATOR:
<sub64>
There are now 200 HIV–positive <w AJ0> <LINEBREAKERR>
chimps <w NN> around the world
</sub64><strat VERBATIM><duration 5><TC respecting>

<sub65>
and only one has ever developed
full-blown AIDS.
</sub65><strat VERBATIM><duration 4><TC not respecting>

<sub66>
They are <DELETION> still </DELETION> a health hazard <DELETION> to humans </DELETION> and <CONDENSATION> are <w VBB> have to be <LINEBREAKERR>
housed <w VVN> </CONDENSATION> separately from other ,<DELETION> clean </DELETION> chimps.
</sub66><strat DELETION, DELETION, CONDENSATION, DELETION><duration 5.06><TC not respecting>

<sub67>
The Biomedical <w NP041> Primate <w NP042> <LINEBREAKERR>
Research <w NP043> Centre <w NP044> in Holland
</sub67><strat VERBATIM><duration 4.08><TC respecting>

<sub68>
was one of the first
to begin AIDS work with chimps.
</sub68><strat VERBATIM><duration 4.13><TC respecting>

RONALD BONTROP:
<sub69>
You see one
of our HIV -infected <DELETION> chi </DELETION> , chimpanzees.
</sub69><strat DELETION><duration 4.14><TC respecting>

<sub70>
<ADDITION> He </ADDITION> <DELETION> His named </DELETION> is called
Koon Koo .
He was infected in 1982,
</sub70><strat ADDITION, DELETION><duration 4.22><TC respecting>

<sub71>
one of the first chimps
that received <CONDENSATION> a the </CONDENSATION> serum sample
</sub71><strat CONDENSATION><duration 4.17><TC respecting>

<sub72>
from an AIDS patient from America.
</sub72><strat VERBATIM><duration 2.18><TC not respecting>

<sub73>
This chimp is still
completely healthy,
</sub73><strat VERBATIM><duration 3.20><TC respecting>

<sub74>
although he was infected
about 16 years ago.
</sub74><strat VERBATIM><duration 4><TC respecting>

NARRATOR:
<sub75>
It <SPECIFICATION> is 's </SPECIFICATION> possible that Koon Koo
and the other HIV chimps here
will develop AIDS and die.
</sub75><strat SPECIFICATION><duration 6.12><TC not respecting><3-liner>

<sub76>
If they don't,
</sub76><strat VERBATIM><duration 2.10><TC respecting>

<sub77>
they could be
in this <ALTSPCOMPOUND> bio-containment bio containment</ALTSPCOMPOUND>
housing
for another 20 years.
</sub77><strat VERBATIM+ALTSPCOMPOUND><duration_5.21><TC respecting>

<sub78>
<CONDENSATION> If ~~When~~ </CONDENSATION> it's warm <DELETION> ~~enough~~ </DELETION> <DELETION> ~~and not raining~~ </DELETION> , the roof folds back
<CONDENSATION> and they can see ~~to give them a glimpse of~~ </CONDENSATION> the outside world.
</sub78><strat CONDENSATION, DELETION, DELETION, CONDENSATION><duration 5.17><TC respecting>

<sub79>
AIDS <CONDENSATION> has ~~may have~~ meant </CONDENSATION> a stay of execution,
but <DELETION> ~~now~~ </DELETION> the chimps face a life sentence.
</sub79><strat CONDENSATION, DELETION><duration 5><TC not respecting>

~~TOM INSEL: Well I think we blew it. I think as a, as a scientific community we made a bad decision some years ago, but we made it under duress at the time when the AIDS epidemic was really raging in the United States and there was a bit of a scientific panic about how we were going to control this.~~

~~NARRATOR: Just in time for the arrival of AIDS Booee was on the move again. By the early 80s the initial excitement about teaching chimps sign language had faded and the American government withdrew its funding. For 13-year-old Booee it meant going back to the harsh world of biomedical research, this time to Jim Mahoney's lab.~~

~~JIM MAHONEY: Booee arrived along with almost 40 other chimps. He seemed to settle in to LEMSIP. Obviously there was an enormous change in his lifestyle, and he became just one of the chimps who was going to be used in hepatitis research and eventually, of course, he was supposed to have been used in AIDS.~~

~~NARRATOR: At LEMSIP Booee was put in a cage on his own, but this was one of the first centres to address the boredom that chimps faced in solitary confinement.~~

~~JIM MAHONEY: It was truly quite spectacular to see the way they would pick up a little plastic mirror and a comb or a, a tooth-brush and use the mirror to examine parts of their body and you'd find them cleaning their teeth and doing their eyes and just amazing, even down to playing a, a mouth organ and I remember there was one chimp in particular was, was incredible on the harmonica and he'd even hit it every now and again to knock the spit out, you know. Sometimes these chimps will take to human activities as though they'd always done them. Every now and again what we called a circus and that's when there were enough people that were free that we could put on a little party for the animals that would last maybe an hour and we'd blow up balloons and bring in all sorts of games and we'd go in in jump-suits and we'd paint the white jump-suits different colours and patterns and we'd wear light-coloured wigs and we'd just play with the chimps and let them play with the balloons.~~

~~NARRATOR: After 16 years at LEMSIP an American TV company heard about the sad story of Booee the signing chimp, now in biomedical research. They suggested he be reunited with the man who'd taught him sign language.~~

~~ROGER FOUTS: And they said wouldn't this be a great reunion to, to get you out. Do you think he will remember you, and I didn't think he would and, and it had been 16 years. I'd~~

~~changed, he'd changed.   I went in, I pulled the mask down, he recognised me immediately and we started playing games that we hadn't played in 16 years.~~

~~Booee, how you doing Booee, how you doing?   How you doing?   Do you remember? What do you want, what do you want, what do you want?   Booee, yes, Booee.~~

~~And not only did he recognise me but he remembered something I'd forgotten in that Booee had a nickname for him, for me and that he was such a fast signer, he was always very, very, I'd guess he'd be comparable of a fast talker, somebody who talks really, really fast, a jabber mouth or a jabber hands if you will, and his name sign was Booee and my name sign is Roger and what he would do is when he'd sign Booee or when he'd want us to do something rather than signing Booee/ Roger, he would take his name sign and just flick the top of his ear, it was like saying Rog or something, and I'd forgotten that.   He was the only chimp that's ever called me that.~~

~~Do you remember Roger?   Yes...~~

~~And he started doing this and suddenly I realised my gosh, that's my name, he remembered my name, and it was great fun.   I mean it was like no time had passed and all, he was the same little boy even though he's probably 140/150lbs now.   What gave me nightmares after that was when I told him I had to leave now.   I said I have to leave now, and rather than screaming or baying to come or to go he accepted it, but he accepted it with such weight and sorrow and that you could physically see him in this shot just, it was like somebody put a huge weight on his back and he was crushed and he moved to the back of his cage.~~

NARRATOR:
<sub80>
In the 40 years they've been used
in research, the lives
of captive chimps have <ERRCORRECTION> improved ~~been proved~~
</ERRCORRECTION>.
</sub80><strat VERBATIM+ERRCORRECTION><duration 7><TC not respecting><3-liner>

<sub81>
Bids to <CONDENSATION> help ~~support~~ </CONDENSATION> the chimps came from
<w PRP> <LINEBREAKERR>
groups <w NN> wanting to retire them, and <w CJC> <DELETION> ~~also~~ </DELETION>
<LINEBREAKERR>
<DELETION> ~~from~~ </DELETION> the Coulston Research Foundation.
</sub81><strat CONDENSATION, DELETION, DELETION><duration 8><TC not respecting>

<sub82>
Their founder is totally against
retiring these chimpanzees.
</sub82><strat VERBATIM><duration 5.04><TC not respecting>

DR FRED COULSTON (THE COULSTON FOUNDATION):
<sub83>
They are like human beings <DELETION> that are </DELETION> kept
in a hospital, <DELETION> in a clinic </DELETION> as they get older.
</sub83><strat DELETION, DELETION><duration 5><TC not respecting>

<sub84>
We will keep them in the same way . ,
We we <CONDENSATION> 'll will treat </CONDENSATION> them <DELETION> with
</DELETION>, for their diseases . ,
</sub84><strat CONDENSATION, DELETION><duration 5><TC not respecting>

<sub85>
We we will learn from them
how to treat human diseases
of comparable age groups.
</sub85><strat VERBATIM><duration 6.13><TC respecting><3-liner>

<sub86>
That's the true sanctuary.
</sub86><strat VERBATIM><duration 2.12><TC respecting>

<sub87>
To <DELETION> just </DELETION> take these valuable animals and <w CJC>
<LINEBREAKERR>
throw <SPECIFICATION> them 'em </SPECIFICATION> out in a field in, <DELETION>
in </DELETION> <DELETION> South </DELETION> Texas ?! .
</sub87><strat DELETION, SPECIFICATION, DELETION, DELETION><duration
4.24><TC not respecting>

NARRATOR:
<sub88>
Dr. Coulston wants to use the <w AT0> ageing <w AJ0> <LINEBREAKERR>
survivors <w NN> of the space programme
and their descendants
</sub88><strat VERBATIM><duration 7.07><TC not respecting><3-liner>

<sub89>
to study prostate cancer, diabetes
and any other human diseases that <w CJT> <LINEBREAKERR>
chimps <CONDENSATION> may might </CONDENSATION> develop in their old age.
</sub89><strat CONDENSATION><duration 7.13><TC not respecting><3-liner>

FRED COULSTON:
<sub90>
The point is ...
these are animals
</sub90><strat VERBATIM><duration 4.03><TC respecting>

<sub91>
that have value throughout <w PRP> their <w DPS> <LINEBREAKERR>
life <w NN> for the medical understanding
of diseases of old age.
</sub91><strat VERBATIM><duration 7.20><TC respecting><3-liner>

NARRATOR: Last month the Air Force announced the majority of the chimps would be
handed over to Dr. Coulston.
<sub92>
Funds couldn't be raised
to build a retirement sanctuary.
</sub92><strat VERBATIM><duration 4.13><TC not respecting>

<sub93>
Neither the drugs companies <DELETION> who'd used them </DELETION>,
<DELETION> nor </DELETION> the government <DELETION> who founded the colony
</DELETION> ,nor the public
<CONDENSATION> would were willing to </CONDENSATION> provide the money.
</sub93><strat DELETION, DELETION, DELETION, CONDENSATION><duration
6.17><TC not respecting><3-liner>

<sub94>
A lucky few have escaped
from research.
</sub94><strat VERBATIM><duration 5.03><TC respecting>

<sub95>
At the <ERRCORRECTION> Wildlife Waystation wildlife way station
</ERRCORRECTION>
outside Los Angeles ,
</sub95><strat VERBATIM+ERRCORRECTION><duration 4.21><TC_respecting>

<sub96>
one of the latest arrivals is Booee,
now 32 years old.
</sub96><strat VERBATIM><duration 4.12><TC respecting>

JIM MAHONEY
<sub97>
This is a <w DT0> typical <ADDITION> male <w AJ0> </ADDITION> chimpanzee <w
NN-AJ0> <LINEBREAKERR>
response <w NN> to strangers
coming into his territory.
</sub97><strat ADDITION><duration 6.13><TC not respecting><3-liner>

<sub98>
<DELETION> and </DELETION> I hope you are </SPECIFICATION> not n't
</SPECIFICATION> <DELETION> too </DELETION> offended <DELETION> by this
</DELETION>,
but that is a normal thing
for a chimp to want to do.
</sub98><strat DELETION, SPECIFICATION, DELETION, DELETION><duration
6.13><TC respecting><3-liner>

<sub99>
It doesn't make him
unnecessarily nasty <DELETION> or </DELETION> . ...
</sub99><strat DELETION><duration 4><TC respecting>

<sub100>
Look at him, he's laughing now . ,
You you see , <DELETION> you see </DELETION> ? .
</sub100><strat DELETION><duration 4.02><TC respecting>

<sub101>
<DELETION> There </DELETION> He he was displaying just a moment ago.
<DELETION> and </DELETION> now he's laughing <DELETION> and now he's laughing
</DELETION> with his head up.
</sub101><strat DELETION, DELETION, DELETION><duration 5><TC not respecting>

<sub102>
<DELETION> and </DELETION> He he wants to play tickle and chase.
</sub102><strat DELETION><duration 2.18><TC not respecting>

NARRATOR:
<sub103>
Jim Mahoney visits Booee regularly
to check his health.
</sub103><strat VERBATIM><duration 4.13><TC not respecting>

<sub104>
Booee escaped AIDS research,
but <DELETION> he </DELETION> was infected with hepatitis .
</sub104><strat DELETION><duration 4.23><TC not respecting>

<sub105>
so <DELETION> he </DELETION> <CONDENSATION> is has to be
</CONDENSATION> housed separately
from <DELETION> the </DELETION> other chimps <DELETION> here
</DELETION>.
</sub105><strat DELETION, CONDENSATION, DELETION, DELETION><duration
3.13><TC not respecting>

<sub106>
His life at the <ERRCORRECTION> Waystation way station </ERRCORRECTION>
depends <w VVZ> <LINEBREAKERR>
on <w PRP> donations <w NN> from the public.
</sub106><strat VERBATIM+ERRCORRECTION><duration 5.01><TC not respecting>

<sub107>
After all he's been through,
Booee's fate is still uncertain.
</sub107><strat VERBATIM><duration 5><TC not respecting>

<sub108>
He could live for another 20 years.
</sub108><strat VERBATIM><duration 2.13><TC not respecting>

<sub109>
The arguments over the morality
of experimenting on chimps continue,
</sub109><strat VERBATIM><duration 5.09><TC not respecting>

<sub110>
but both sides now agree .
</sub110><strat VERBATIM><duration 2.13><TC respecting>

<sub111>
<ADDITION> that </ADDITION> those Those we have used
deserve better treatment.
</sub111><strat ADDITION><duration 4.14><TC respecting>

MIKE KEELING:
<sub112>
I <DELETION> simply </DELETION> prioritise the advancement
of human health and welfare
</sub112><strat DELETION><duration 5><TC respecting>

<sub113>
at the cost of the chimp having
to donate some experimental effort
in their life . ,
</sub113><strat VERBATIM><duration 6.13><TC not respecting><3-liner>

<sub114>
But but <CONDENSATION> what I think everyone agrees is </CONDENSATION>
<ADDITION> that </ADDITION> if we <CONDENSATION> 're going to ask
<CONDENSATION> the chimp to do that .
</sub114><strat CONDENSATION, ADDITION, CONDENSATION><duration 4.07><TC
not respecting>

<sub115>
we <CONDENSATION> must ought to be willing to </CONDENSATION> back it up
<TRANSPOSITIONa> by </TRANSPOSITIONa> when they've done their service
</sub115><strat CONDENSATION, TRANSPOSITION><duration 4><TC not respecting>

<sub116>
<TRANSPOSITIONa> by </TRANSPOSITIONa> <CONDENSATIONa> giving
<CONDENSATIONa> <DELETION> we </DELETION> <CONDENSATIONa> provide
</CONDENSATIONa> <SPECIFICATION> them 'em </SPECIFICATION> a high
quality life
at a <CONDENSATION> stable stabilised </CONDENSATION> level of funding.
</sub116><strat TRANSPOSITION, CONDENSATION, DELETION, SPECIFICATION,
CONDENSATION><duration 5><TC not respecting>

JANE GOODALL:
<sub117>
Since chimpanzees are our <w DPS> closest <w AJS> <LINEBREAKERR>
living <w AJ0> relatives <w NN>, since <DELETION>in their own right </DELETION>
they're
wondrous and amazing beings,
</sub117><strat DELETION><duration 7.13><TC not respecting><3-liner>

<sub118>
we should feel such a weight
of guilt that we've brought them
from their native forests,
</sub118><strat VERBATIM><duration 6.20><TC not respecting><3-liner>

\<sub119\>
we've subjected them
to horrendous experimentation
\</sub119\>\<strat VERBATIM\>\<duration 4\>\<TC not respecting\>

\<sub120\>
and we \<SPECIFICATION\> are ~~'re~~ \</SPECIFICATION\> not prepared to
\<DELETION\> ~~get together~~ \</DELETION\> \<DELETION\> ~~and~~ \</DELETION\> sort out
\<br\>
a decent future for them.
\</sub120\>\<strat SPECIFICATION, DELETION, DELETION\>\<duration 4.15\>\<TC not
respecting\>

Appendix 3: Research Corpus – *The World's Most Wanted*

Transcript File

JANE CORBIN: It stood proud on the New York skyline, a symbol of the American dream. On Tuesday it was shattered. Within hours suspicion fell on one man.

PRESIDENT BUSH: There is no question he is what we would call a prime suspect.

CORBIN: The prime suspect is Osama Bin Laden whose murderous campaign against America had already earned him a place on the FBI's most wanted list.

BUSH: And if he thinks he can hide and run from the United States or allies, he will be sorely mistaken.

CORBIN: Tonight Panorama investigates the terrorist, Osama Bin Laden, the world's most wanted man. A religious warrior forged in the remote hills of Afghanistan, a defender of the purity of Islam, that's Bin Laden's claim. BIN LADEN VIDEO

Most of the world now sees him as a terrorist mastermind, a mass murderer of innocent civilians. But in some Muslim communities they have a different view of this man.

Dr SAAD AL FAGIH: It's a big irony that in his personal level he's very soft, very humble, very smooth. He speaks little.

CORBIN: Osama Bin Laden's beginnings were far from humble. He was born into a fabulously wealthy family in Jedda, Saudi Arabia, the 17th son with more than fifty brothers and sisters. But Osama Bin Laden turned his back on an ostentatious life of luxury to live as a fighter and a fugitive in the hills of Afghanistan.

His progress has been watched over the years by a fellow Saudi dissident who does not share Bin Laden's views on using violence.

Dr SAAD AL FAGIH: Being extremely devoted and having the history of being a successful merchant or contractor who is a millionaire and abandoning all this for the sake of what he believes in, has given him this big name in the Muslim world.

CORBIN: Afghanistan in the 1980s, a cockpit of superpower rivalry. The mujaheddin, Islamic fighters, supported clandestinely by America, were struggling to drive out the Soviet occupiers. Osama Bin Laden was one of thousands of young Saudis who flocked to the mountains to join the cause despite the perils.

[PART 2]

CORBIN: ... During his stay in the Sudan several intelligence agencies began tracking Bin Laden's activities. He was both building up his own militant organisation - Al Quaeda - and beginning to create a coalition of Islamic terror groups. In Washington the CIA set up a special task force to investigate what Osama Bin Laden was planning from his base in the Sudan. They used surveillance, secret agents and intercepts. They estimated he had $200 million, much of it in cash, some invested through front companies in Khartoum. The money and what it paid for was hard for CIA analysts to track.

STANLEY BEDLINGTON: He always deliberately maintained his distance from terrorist attacks. When you look at Osama Bin Laden and his network, his financial network, it's rather like peeling an onion skin by skin.

CORBIN: During the mid-90s some twenty countries were identified as places where terror groups were operating with support from Osama Bin Laden. There were attacks in the Philippines, the shooting of tourists in Egypt, bombings in Saudi Arabia which killed American soldiers. Bin Laden's aids were even trying to buy components for chemical and nuclear weapons on the black market.

BLEDLINGTON: He sort of acts as a financial godfather. I don't think, to my knowledge, that he's actually ordered a group to go out and carry out a specific operation. What he does is sort of select a group and then give it some financing and then give it its head as it were.

CORBIN: In 1998 Osama Bin Laden announced a formal coalition of some of the world's most feared terror groups. Amongst them, Egyptians held responsible for the massacre of 85 tourists, and he pronounced afatwa calling for attacks against American civilians.

~~BIN LADEN: "All these crimes committed by the Americans are but a clear declaration of war against Allah, his Prophet and against Muslims. Obeying the order to kill the Americans and their allies, both civilian and military, is the duty of every muslim."~~

CORBIN: In the Islamic world doors were closing against Bin Laden. Sudan succumbed to pressure to evict their troublesome guest. In Washington, Presidential advisers considered but rejected a raid to capture Bin Laden, now hiding in Afghanistan under the protection of the extremist Taliban.

Why was the decision taken in the end not to go into Afghanistan and take him out?

OLIVER REVELL: I believe they thought there would National Executive too much collateral damage, too many non-involved casualties and there would be a high casualty rate amongst the soldiers and it would require a great deal of logistical support, in and out.

CORBIN: So the decision was made to leave him there for the time being.

REVELL: Leave him there for the time being. BIN LADEN VIDEO

CORBIN: From Afghanistan Bin Laden continued to taunt the American President himself, Bill Clinton, with his propaganda material. But the terrorist leader is very wary about his security. Five times now, various individuals have tried to kill him.

[PART 3]

Dr SAAD AL FAGIH: ... That is not the case of many Moslems and especially Bin Laden team if we can use this phrase. Those people who believe in this, they may be even rich, prosperous people, they may be very highly educated, and we have a very clear example now.

Pilots, well trained, they elect to kill themselves in this exceptional manner.

CORBIN: As dawn broke over the shattered skyline of New York on the morning of September 12th the Intelligence community believed they had a concrete indication as to who was ultimately responsible for masterminding the devastating attack on America.

STANLEY BEDLINGTON: Immediately after the attack there were intercepts, one in particular, between people we presume to be supporters of Osama Bin Laden in either Washington and New York City or both who sent messages that were intercepted saying 'both jobs successful'.

CORBIN: More than 600 people were dead, over 5000 missing. President Bush toured the devastated streets. His authorities had established further indications of Bin Laden's guilt. A Yemeni linked to the bombing of the USS Cole was confirmed amongst the group of hijackers. Another member of the group was believed to have trained in a Bin Laden terror camp before going to America. For President Bush this was war.

BUSH: I can hear you, the rest of the world hears you (cheers from the people) and the people who knocked these buildings down will hear all of us soon. APPLAUSE

CORBIN: Osama Bin Laden denies he is responsible and he'll be hard to smoke out of his refuge in Afghanistan. In Kabul people are leaving, heading for the border. The Taliban have been given a three day ultimatum to give up the fugitive terrorist or face the wrath of America. But even if they can destroy Bin Laden, this won't end the terror threat.

Dr SAAD AL FAGIH: Now if they intend to destroy Bin Laden, Afghanistan, they would have done nothing because most of the bulk of the followers is outside Afghanistan, and they are much more intelligence now, much more sophisticated. They can revive themselves much easily, and that is the other dangerous side of it.

MIKE McMAHON: It's just a bad two days. It's gonna be a bad couple of months. I don't know how the city is ever going to recover from this. So many lives are going to be touched and lost. It's going to be incredible. Just the shame of these people thinking that they're going to go to heaven and all that stuff that they're killing us, killing people for no reason, for whatever their reasons are. It's senseless.

Subtitle File

*** Timecode: In: 10:0:8:12 Out: 10:0:13:8
*** Framecount: In: 900212 Out: 900333
It rose out of the New York skyline,
a symbol of the American dream.

*** Timecode: In: 10:0:13:9 Out: 10:0:16:1
*** Framecount: In: 900334 Out: 900401
On Tuesday, it was shattered.

*** Timecode: In: 10:0:17:13 Out: 10:0:19:16
*** Framecount: In: 900438 Out: 900491
Holy shit!

*** Timecode: In: 10:0:23:8 Out: 10:0:25:11
*** Framecount: In: 900583 Out: 900636
Holy...

*** Timecode: In: 10:0:25:12 Out: 10:0:27:18
*** Framecount: In: 900637 Out: 900693
Oh, my God!

*** Timecode: In: 10:0:27:19 Out: 10:0:31:12
*** Framecount: In: 900694 Out: 900787
Within hours,
suspicion fell on one man.

*** Timecode: In: 10:0:31:13 Out: 10:0:36:5
*** Framecount: In: 900788 Out: 900905
There is no question he's what
we would call a prime suspect.

*** Timecode: In: 10:0:37:16 Out: 10:0:42:11
*** Framecount: In: 900941 Out: 901061
The prime suspect is Osama Bin
Laden. His campaign against America

*** Timecode: In: 10:0:42:12 Out: 10:0:47:0
*** Framecount: In: 901062 Out: 901175
had earned him a place
on the FBI's most wanted list.

*** Timecode: In: 10:0:47:1 Out: 10:0:53:9
*** Framecount: In: 901176 Out: 901334
If he thinks he can hide from
the USA or allies, he is mistaken.

*** Timecode: In: 10:0:53:10 Out: 10:0:59:8
*** Framecount: In: 901335 Out: 901483
Tonight, Panorama investigates
the terrorist, Osama Bin Laden -

*** Timecode: In: 10:0:59:9 Out: 10:1:1:23
*** Framecount: In: 901484 Out: 901548
the world's most wanted man.

*** Timecode: In: 10:1:14:15 Out: 10:1:18:22
*** Framecount: In: 901865 Out: 901972
A religious warrior
in the hills of Afghanistan,

*** Timecode: In: 10:1:18:23 Out: 10:1:24:5
*** Framecount: In: 901973 Out: 902105
the defender of the purity
of Islam - that's Bin Laden's claim.

*** Timecode: In: 10:1:24:6 Out: 10:1:27:3
*** Framecount: In: 902106 Out: 902178
GUNSHOT

*** Timecode: In: 10:1:27:4 Out: 10:1:29:22
*** Framecount: In: 902179 Out: 902247
Most of the world sees him
as a terrorist mastermind,

*** Timecode: In: 10:1:29:23 Out: 10:1:32:13
*** Framecount: In: 902248 Out: 902313
a murderer of innocent civilians.

*** Timecode: In: 10:1:32:14 Out: 10:1:35:2
*** Framecount: In: 902314 Out: 902377
But in some Muslim communities,

*** Timecode: In: 10:1:35:3 Out: 10:1:37:16
*** Framecount: In: 902378 Out: 902441
they have a different view of him.

*** Timecode: In: 10:1:40:2 Out: 10:1:43:5
*** Framecount: In: 902502 Out: 902580
GUNSHOT

*** Timecode: In: 10:1:43:6 Out: 10:1:50:23
*** Framecount: In: 902581 Out: 902773
It's ironic that in his personal
life he is very soft, humble, smooth.

*** Timecode: In: 10:1:50:24 Out: 10:1:56:12
*** Framecount: In: 902774 Out: 902912
He speaks little. Osama Bin Laden's
beginnings were far from humble.

*** Timecode: In: 10:1:56:13 Out: 10:2:1:10
*** Framecount: In: 902913 Out: 903035
He was born into a wealthy family
in Jeddah, Saudi Arabia,

*** Timecode: In: 10:2:1:14 Out: 10:2:5:5
*** Framecount: In: 903039 Out: 903130
the 17th son
with 50 brothers and sisters.

*** Timecode: In: 10:2:5:6 Out: 10:2:10:12
*** Framecount: In: 903131 Out: 903262
But Osama Bin Laden turned his back
on an ostentatious life of luxury,

*** Timecode: In: 10:2:10:13 Out: 10:2:15:13
*** Framecount: In: 903263 Out: 903388
to live as a fighter and a fugitive
in the hills of Afghanistan.

*** Timecode: In: 10:2:15:14 Out: 10:2:19:21
*** Framecount: In: 903389 Out: 903496
His progress has been watched
over the years

*** Timecode: In: 10:2:19:22 Out: 10:2:22:10
*** Framecount: In: 903497 Out: 903560
by a fellow Saudi dissident,

*** Timecode: In: 10:2:22:11 Out: 10:2:27:8
*** Framecount: In: 903561 Out: 903683
who does not share Bin Laden's
views on using violence.

*** Timecode: In: 10:2:28:11 Out: 10:2:30:22
*** Framecount: In: 903711 Out: 903772
Being extremely devoted

*** Timecode: In: 10:2:30:23 Out: 10:2:38:6
*** Framecount: In: 903773 Out: 903956
and having the history of being
a merchant who is a millionaire

*** Timecode: In: 10:2:38:7 Out: 10:2:42:16
*** Framecount: In: 903957 Out: 904066
and abandoning this
for the sake of what he believes in,

*** Timecode: In: 10:2:42:17 Out: 10:2:45:13
*** Framecount: In: 904067 Out: 904138
made him famous in the Muslim world.

*** Timecode: In: 10:2:45:14 Out: 10:2:48:1
*** Framecount: In: 904139 Out: 904201
GUNFIRE

*** Timecode: In: 10:2:48:2 Out: 10:2:52:22
*** Framecount: In: 904202 Out: 904322
Afghanistan in the 1980s -
a cockpit of superpower rivalry.

*** Timecode: In: 10:2:52:24 Out: 10:2:57:3
*** Framecount: In: 904324 Out: 904428
The Mujahideen,
supported clandestinely by America,

*** Timecode: In: 10:2:57:4 Out: 10:3:1:15
*** Framecount: In: 904429 Out: 904540
was struggling to drive out
the Soviet occupiers.

*** Timecode: In: 10:3:1:16 Out: 10:3:5:21
*** Framecount: In: 904541 Out: 904646
Osama Bin Laden was
one of thousands of young Saudis

*** Timecode: In: 10:3:5:22 Out: 10:3:10:16
*** Framecount: In: 904647 Out: 904766
who flocked to the mountains to
join the cause, despite the perils.

[PART 2]

*** Timecode: In: 10:12:0:1 Out: 10:12:2:13
*** Framecount: In: 918001 Out: 918063
During his stay in the Sudan,

*** Timecode: In: 10:12:2:14 Out: 10:12:7:7
*** Framecount: In: 918064 Out: 918182
several intelligence agencies
began tracking his activities.

*** Timecode: In: 10:12:7:8 Out: 10:12:12:7
*** Framecount: In: 918183 Out: 918307
He was both building up his
militant organisation, al-Qaida,

*** Timecode: In: 10:12:12:8 Out: 10:12:16:24
*** Framecount: In: 918308 Out: 918424
and creating a coalition
of Islamic terror groups.

*** Timecode: In: 10:12:17:0 Out: 10:12:21:16
*** Framecount: In: 918425 Out: 918541
In Washington,
the CIA set up a special task force

*** Timecode: In: 10:12:21:17 Out: 10:12:26:21
*** Framecount: In: 918542 Out: 918671
to investigate what Osama Bin Laden
was planning from his base.

*** Timecode: In: 10:12:26:22 Out: 10:12:31:5
*** Framecount: In: 918672 Out: 918780
They used surveillance,
secret agents and intercepts.

*** Timecode: In: 10:12:31:6 Out: 10:12:33:24
*** Framecount: In: 918781 Out: 918849
They estimated he had ¤200 million,

*** Timecode: In: 10:12:34:0 Out: 10:12:38:22
*** Framecount: In: 918850 Out: 918972
much of it in cash, some invested
through front companies in Khartoum.

*** Timecode: In: 10:12:38:23 Out: 10:12:43:17
*** Framecount: In: 918973 Out: 919092
The money, and what it paid for,
was hard for CIA analysts to track.

*** Timecode: In: 10:12:45:8 Out: 10:12:53:16
*** Framecount: In: 919133 Out: 919341
He always deliberately maintained
his distance from terrorist attacks.

*** Timecode: In: 10:12:53:17 Out: 10:12:58:18
*** Framecount: In: 919342 Out: 919468
When you look at his financial
network, it is like peeling an onion.

*** Timecode: In: 10:12:58:19 Out: 10:13:3:14
*** Framecount: In: 919469 Out: 919589
In the mid-'90s, 20 countries
were identified as places

*** Timecode: In: 10:13:3:15 Out: 10:13:8:11
*** Framecount: In: 919590 Out: 919711
where terrorists were operating
with support from Osama Bin Laden.

*** Timecode: In: 10:13:8:12 Out: 10:13:12:3
*** Framecount: In: 919712 Out: 919803
There were attacks
in the Philippines,

*** Timecode: In: 10:13:12:4 Out: 10:13:14:22
*** Framecount: In: 919804 Out: 919872
the shooting of tourists in Egypt,

*** Timecode: In: 10:13:14:23 Out: 10:13:19:16
*** Framecount: In: 919873 Out: 919991
bombings in Saudi Arabia,
which killed American soldiers.

*** Timecode: In: 10:13:19:17 Out: 10:13:25:18
*** Framecount: In: 919992 Out: 920143
Bin Laden's aides even tried to buy
components for chemical weapons.

*** Timecode: In: 10:13:25:19 Out: 10:13:28:6
*** Framecount: In: 920144 Out: 920206
He acts as a financial godfather.

*** Timecode: In: 10:13:28:7 Out: 10:13:33:15
*** Framecount: In: 920207 Out: 920340
To my knowledge, I don't think that
he's ordered a specific operation.

*** Timecode: In: 10:13:33:16 Out: 10:13:39:9
*** Framecount: In: 920341 Out: 920484
He selects a group, then finances
it and gives it its head, as it were.

*** Timecode: In: 10:13:39:10 Out: 10:13:43:8
*** Framecount: In: 920485 Out: 920583
In 1998,
Bin Laden announced a coalition

*** Timecode: In: 10:13:43:9 Out: 10:13:47:12
*** Framecount: In: 920584 Out: 920687
of some of the world's
most feared terror groups,

*** Timecode: In: 10:13:47:13 Out: 10:13:52:12
*** Framecount: In: 920688 Out: 920812
amongst them Egyptians, responsible
for the massacre of 85 tourists.

*** Timecode: In: 10:13:52:13 Out: 10:13:57:7
*** Framecount: In: 920813 Out: 920932
He pronounced a fatwa, calling for
attacks against American civilians.

*** Timecode: In: 10:14:13:14 Out: 10:14:18:7
*** Framecount: In: 921339 Out: 921457
In the Islamic world, doors
were closing against Bin Laden.

*** Timecode: In: 10:14:18:8 Out: 10:14:22:2
*** Framecount: In: 921458 Out: 921552
Sudan evicted
their troublesome guest.

*** Timecode: In: 10:14:22:3 Out: 10:14:26:21
*** Framecount: In: 921553 Out: 921671
In Washington, presidential
advisers considered, but rejected,

*** Timecode: In: 10:14:26:22 Out: 10:14:29:10
*** Framecount: In: 921672 Out: 921735
a raid to capture Bin Laden,

*** Timecode: In: 10:14:29:11 Out: 10:14:34:14
*** Framecount: In: 921736 Out: 921864
now hiding in Afghanistan,
under the protection of the Taliban.

*** Timecode: In: 10:14:34:15 Out: 10:14:39:17
*** Framecount: In: 921865 Out: 921992
Why was the decision taken not to
go into Afghanistan and get him?

*** Timecode: In: 10:14:39:18 Out: 10:14:46:8
*** Framecount: In: 921993 Out: 922158
I believe they thought there would
be too many none-involved casualties

*** Timecode: In: 10:14:46:9 Out: 10:14:49:20
*** Framecount: In: 922159 Out: 922245
and a high casualty rate
amongst soldiers

*** Timecode: In: 10:14:49:21 Out: 10:14:53:7
*** Framecount: In: 922246 Out: 922332
and it would require
logistical support.

*** Timecode: In: 10:14:53:8 Out: 10:14:58:1
*** Framecount: In: 922333 Out: 922451
So, the decision was made to leave
him there? Yes, for the time being.

*** Timecode: In: 10:14:58:2 Out: 10:15:3:14
*** Framecount: In: 922452 Out: 922589
Bin Laden continued to taunt the
American President, Bill Clinton,

*** Timecode: In: 10:15:3:15 Out: 10:15:6:19
*** Framecount: In: 922590 Out: 922669
with his propaganda material.

*** Timecode: In: 10:15:9:0 Out: 10:15:11:23
*** Framecount: In: 922725 Out: 922798
GUNFIRE

*** Timecode: In: 10:15:11:24 Out: 10:15:16:15
*** Framecount: In: 922799 Out: 922915
But the terrorist leader
IS very wary about his security.

*** Timecode: In: 10:15:16:16 Out: 10:15:20:23
*** Framecount: In: 922916 Out: 923023
Five times now,
individuals have tried to kill him.

[PART 3]

*** Timecode: In: 10:24:41:8 Out: 10:24:46:3
*** Framecount: In: 937033 Out: 937153
and especially the Bin Laden team,
if you can use this phrase.

*** Timecode: In: 10:24:46:4 Out: 10:24:48:19
*** Framecount: In: 937154 Out: 937219
Those people who believe in this

*** Timecode: In: 10:24:48:20 Out: 10:24:53:20
*** Framecount: In: 937220 Out: 937345
may be even rich or prosperous
people and maybe highly educated.

*** Timecode: In: 10:24:53:21 Out: 10:24:58:16
*** Framecount: In: 937346 Out: 937466
We have a very clear example now -
pilots, well trained.

*** Timecode: In: 10:24:58:17 Out: 10:25:1:7
*** Framecount: In: 937467 Out: 937532
They elect to kill themselves

*** Timecode: In: 10:25:1:8 Out: 10:25:4:17
*** Framecount: In: 937533 Out: 937617
in this exceptional manner.

*** Timecode: In: 10:25:4:18 Out: 10:25:9:18
*** Framecount: In: 937618 Out: 937743
As dawn broke over the shattered
skyline of New York on 12 September,

*** Timecode: In: 10:25:9:19 Out: 10:25:12:7
*** Framecount: In: 937744 Out: 937807
the intelligence community believed

*** Timecode: In: 10:25:12:8 Out: 10:25:17:16
*** Framecount: In: 937808 Out: 937941
they had a concrete indication as
to who was ultimately responsible

*** Timecode: In: 10:25:17:17 Out: 10:25:20:21
*** Framecount: In: 937942 Out: 938021
for masterminding the attack
on America.

*** Timecode: In: 10:25:20:22 Out: 10:25:24:12
*** Framecount: In: 938022 Out: 938112
After the attack
there were intercepts,

*** Timecode: In: 10:25:24:13 Out: 10:25:29:3
*** Framecount: In: 938113 Out: 938228
between people we presumed
to be Bin Laden supporters

*** Timecode: In: 10:25:29:4 Out: 10:25:31:24
*** Framecount: In: 938229 Out: 938299
in Washington, New York or both,

*** Timecode: In: 10:25:32:0 Out: 10:25:36:15
*** Framecount: In: 938300 Out: 938415
who sent messages
that were intercepted saying,

*** Timecode: In: 10:25:36:16 Out: 10:25:39:14
*** Framecount: In: 938416 Out: 938489
"Both jobs successful".

*** Timecode: In: 10:25:39:15 Out: 10:25:42:18
*** Framecount: In: 938490 Out: 938568
More than 600 people were dead.

*** Timecode: In: 10:25:42:19 Out: 10:25:45:9
*** Framecount: In: 938569 Out: 938634
Over 5,000 were missing.

*** Timecode: In: 10:25:45:9 Out: 10:25:49:14
*** Framecount: In: 938634 Out: 938739
President Bush toured
the devastated streets.

*** Timecode: In: 10:25:49:15 Out: 10:25:54:5
*** Framecount: In: 938740 Out: 938855
Further implications of Bin Laden's
guilt had been found.

*** Timecode: In: 10:25:54:6 Out: 10:25:59:6
*** Framecount: In: 938856 Out: 938981
A Yemeni linked to the USS Cole
was confirmed as a hijacker.

*** Timecode: In: 10:25:59:7 Out: 10:26:3:2
*** Framecount: In: 938982 Out: 939077
Another member was believed
to have trained

*** Timecode: In: 10:26:3:3 Out: 10:26:6:10
*** Framecount: In: 939078 Out: 939160
in a Bin Laden camp
before going to America.

*** Timecode: In: 10:26:6:11 Out: 10:26:9:1
*** Framecount: In: 939161 Out: 939226
For President Bush this was war.

*** Timecode: In: 10:26:9:2 Out: 10:26:11:7
*** Framecount: In: 939227 Out: 939282
I can hear you!

*** Timecode: In: 10:26:11:8 Out: 10:26:14:18
*** Framecount: In: 939283 Out: 939368
CHEERING

*** Timecode: In: 10:26:14:19 Out: 10:26:19:14
*** Framecount: In: 939369 Out: 939489
I can hear you.
The rest of the world hears you.

*** Timecode: In: 10:26:19:15 Out: 10:26:21:20
*** Framecount: In: 939490 Out: 939545
And the people...

*** Timecode: In: 10:26:21:21 Out: 10:26:24:1
*** Framecount: In: 939546 Out: 939601
CHEERING

*** Timecode: In: 10:26:24:2 Out: 10:26:28:7
*** Framecount: In: 939602 Out: 939707
And the people
who knocked these buildings down

*** Timecode: In: 10:26:28:8 Out: 10:26:31:9
*** Framecount: In: 939708 Out: 939784
will hear all of us soon.
CHEERING

*** Timecode: In: 10:26:31:10 Out: 10:26:35:0
*** Framecount: In: 939785 Out: 939875
Osama Bin Laden
denies he is responsible.

*** Timecode: In: 10:26:35:1 Out: 10:26:38:23
*** Framecount: In: 939876 Out: 939973
He'll be hard to smoke out
of his refuge in Afghanistan.

*** Timecode: In: 10:26:38:24 Out: 10:26:42:14
*** Framecount: In: 939974 Out: 940064
In Kabul,
people are heading for the border.

*** Timecode: In: 10:26:42:15 Out: 10:26:45:5
*** Framecount: In: 940065 Out: 940130
The Taliban have three days

*** Timecode: In: 10:26:45:6 Out: 10:26:49:3
*** Framecount: In: 940131 Out: 940228
to give up the fugitive
or face America's wrath.

*** Timecode: In: 10:26:49:4 Out: 10:26:54:23
*** Framecount: In: 940229 Out: 940373
Even if they can destroy Bin Laden
this won't end the terror threat.

*** Timecode: In: 10:26:54:24 Out: 10:27:0:3
*** Framecount: In: 940374 Out: 940503
Now if they intend to destroy Bin
Laden they will have done nothing.

*** Timecode: In: 10:27:0:4 Out: 10:27:4:19
*** Framecount: In: 940504 Out: 940619
Most of the bulk of the followers
is outside Afghanistan.

*** Timecode: In: 10:27:4:20 Out: 10:27:7:10
*** Framecount: In: 940620 Out: 940685
There is much more intelligence now,

*** Timecode: In: 10:27:7:11 Out: 10:27:12:24
*** Framecount: In: 940686 Out: 940824
much more sophisticated and that is
the other dangerous side of it.

Merge with annotation

JANE CORBIN:
<sub1>
It <PARAPHRASE> rose out of ~~stood proud on~~ </PARAPHRASE> the New York skyline,

a symbol of the American dream.
</sub1><strat PARAPHRASE><duration 4.21><TC not respecting>

<sub2>
On Tuesday, it was shattered.
</sub2><strat VERBATIM><duration 2.17><TC respecting>

<sub3>
Holy shit!
</sub3><strat ADDITION-AUDIO><duration 2.03><TC respecting>

<sub4>
Holy...
</sub4><strat ADDITION-AUDIO><duration 2.03><TC respecting>

<sub5>
Oh, my God!
</sub5><strat ADDITION-AUDIO><duration 2.06><TC respecting>

<sub6>
Within hours ,
suspicion fell on one man.
</sub6><strat VERBATIM><duration 3.18><TC respecting>

PRESIDENT BUSH:
<sub7>
There is no question he <CONDENSATION> 's ~~is~~ </CONDENSATION> what <w PNQ>
<LINEBREAKERR>
we would call a prime suspect.
</sub7><strat CONDENSATION><duration 4.17><TC_not respecting>

CORBIN:
<sub8>
The prime suspect is Osama <w NP031> Bin <w NP032> <LINEBREAKERR>
Laden <w NP033> . <PARAPHRASE> His ~~whose~~ </PARAPHRASE> <DELETION>
~~murderous~~ </DELETION> campaign against America
</sub8><strat PARAPHRASE, DELETION><duration 4.20><TC not respecting>

<sub9>
had <DELETION> already </DELETION> earned him a place
on the FBI's most wanted list.
</sub9><strat DELETION><duration 4.13><TC respecting>

BUSH:
<sub10>
<DELETION> And </DELETION> If if he thinks he can hide <DELETION> and
</DELETION> <DELETION> run</DELETION> from <w PRP> <LINEBREAKERR>

the <w AT0> <CONDENSATION> USA <w NP0> United States </CONDENSATION>
or allies, he <CONDENSATION> is will be</CONDENSATION> <DELETION> sorely
</DELETION> mistaken.
</sub10><strat DELETION, DELETION, DELETION, CONDENSATION,
CONDENSATION, DELETION><duration 6.08><TC respecting>

CORBIN:
<sub11>
Tonight, Panorama investigates
the terrorist, Osama Bin Laden -,
</sub11><strat VERBATIM><duration 5.23><TC respecting>

<sub12>
the world's most wanted man.
</sub12><strat VERBATIM><duration 2.14><TC respecting>

<sub13>
A religious warrior
<DELETION> forged </DELETION> in the <DELETION> remote </DELETION> hills
of Afghanistan,
</sub13><strat DELETION, DELETION><duration 4.07><TC respecting>

<sub14>
<PARAPHRASE> the a </PARAPHRASE> defender of the purity
of Islam -, that's Bin Laden's claim.
</sub14><strat PARAPHRASE><duration 5.07><TC respecting>

 BIN LADEN VIDEO
<sub15>
GUNSHOT
</sub15><strat ADDITION-AUDIO><duration 2.22><TC respecting>

<sub16>
Most of the world <DELETION> now </DELETION> sees him
as a terrorist mastermind,
</sub16><strat DELETION><duration 2.18><TC not respecting>

<sub17>
a <DELETION> ~~mass~~ </DELETION> murderer of innocent civilians.
</sub17><strat DELETION><duration 2.15><TC not respecting>

<sub18>
But in some Muslim communities <u>.</u>
</sub18><strat VERBATIM><duration 2.13><TC not respecting>

<sub19>
they have a different view of <CONDENSATION> <u>him</u> ~~this man~~ </CONDENSATION>.
</sub19><strat CONDENSATION><duration 2.13><TC not respecting>

<sub20>
GUNSHOT
</sub20><strat ADDITION-AUDIO><duration 3.03><TC respecting but too long>

Dr SAAD AL FAGIH:
<sub21>
It's <CONDENSATION> <u>ironic</u> ~~a big irony~~ </CONDENSATION> that in <w PRP> his <w DPS> personal <w AJ0> <LINEBREAKERR>
<ERRCORRECTION> <u>life</u> <w NN> ~~level~~ </ERRCORRECTION> he <SPECIFICATION> <u>is</u> ~~'s~~ </SPECIFICATION> very soft, <DELETION> ~~very~~ </DELETION> humble, <DELETION> ~~very~~</DELETION> smooth.
</sub21><strat CONDENSATION, ERRCORRECTION, SPECIFICATION, DELETION, DELETION><duration 7.17><TC respecting>

<sub22>
He speaks little. ~~CORBIN:~~ Osama Bin Laden <w NP0> 's <w POS <LINEBREAKERR>
beginnings <w NN> were far from humble.
</sub22><strat VERBATIM><duration 5.13><TC not respecting>

<sub23>
He was born into a <DELETION> ~~fabulously~~ </DELETION> wealthy family
in <ALTSP> <u>Jeddah</u> ~~Jedda~~ </ALTSP> Saudi Arabia,
</sub23><strat DELETION><duration 4.22><TC respecting>

<sub24>
the 17th son
with <DELETION> ~~more than~~ </DELETION> <u>50</u> ~~fifty~~ brothers and sisters.
</sub24><strat DELETION><duration 3.16><TC respecting>

<sub25>
But Osama Bin Laden turned his back
on an ostentatious life of luxury <u>.</u>
</sub25><strat VERBATIM><duration 5.06><TC not respecting>

<sub26>
to live as a fighter and a fugitive
in the hills of Afghanistan.
</sub26><strat VERBATIM><duration 5.00><TC not respecting>

<sub27>
His progress has been watched
over the years
</sub27><strat VERBATIM><duration 4.07><TC respecting>

</sub28>
by a fellow Saudi dissident .
</sub28><strat VERBATIM><duration 2.13><TC respecting>

<sub29>
who does not share Bin Laden <w NP0> 's <w POS> <LINEBREAKERR>
views <w NN> on using violence.
</sub29><strat VERBATIM><duration 4.22><TC respecting>

Dr SAAD AL FAGIH:
<sub30>
Being extremely devoted
</sub30><strat VERBATIM><duration 2.11><TC respecting>

<sub31>
and having the history of being
a <DELETION> successful </DELETION> merchant <DELETION> or </DELETION>
<DELETION> contractor </DELETION> who is a millionaire
</sub31><strat DELETION, DELETION, DELETION><duration 7.08><TC_respecting>

<sub32>
and abandoning <DELETION> all <DELETION> this
for the sake of what he believes in,
</sub32><strat DELETION><duration 4.09><TC not respecting>

<sub33>
<CONDENSATION> made has given him famous this big name </CONDENSATION> in
the Muslim world.
</sub33><strat CONDENSATION><duration 2.21><TC not respecting>

<sub34>
GUNFIRE
</sub34><strat ADDITION-AUDIO><duration 2.12><TC respecting but too long>

\<sub35\>
CORBIN:
Afghanistan in the 1980s –,
a cockpit of superpower rivalry.
\</sub35\>\<strat VERBATIM\>\<duration 4.20\>\<TC not respecting\>

\<sub36\>
The \<ALTSP\> Mujahideen mujaheddin \</ALTSP\> ,\<DELETION\> Islamic fighters
\</DELETION\> ,
supported clandestinely by America,
\</sub36\>\<strat DELETION\>\<duration 4.04\>\<TC not respecting\>

\<sub37\>
\<ERROR\> was were \</ERROR\> struggling to drive out
the Soviet occupiers.
\</sub37\>\<strat VERBATIM\>\<duration 4.11\>\<TC respecting\>

\<sub38\>
Osama Bin Laden was \<w VBD\> \<LINEBREAKERR\>
{one of thousands of young Saudis}_Comp
\</sub38\>\<strat VERBATIM\>\<duration 4.05\>\<TC not respecting\>

\<sub39\>
who flocked to the mountains to \<w TO0\> \<LINEBREAKERR\>
join \<w VVI\> the cause, despite the perils.
\</sub39\>\<strat VERBATIM\>\<duration 4.19\>\<TC not respecting\>

\<sub40\>
GUNFIRE
\</sub40\>\<strat ADDITION-AUDIO\>\<duration 3.15\>\<TC respecting but too long\>

\<sub41\>
During his stay in the Sudan ,
\</sub41\>\<strat VERBATIM\>\<duration 2.12\>\<TC respecting\>

\<sub42\>
several intelligence agencies
began tracking \<CONDENSATION\> his Bin Laden's \</CONDENSATION\> activities.
\</sub42\>\<strat CONDENSATION\>\<duration 4.18\>\<TC not respecting\>

\<sub43\>
He was both building up his \<w DPS\> \<DELETION\> ~~own~~ \</DELETION\>
\<LINEBREAKERR\>
militant \<w AJ0\> organisation \<w NN\>, \<ALTSP\> al-Qaida ~~Al Quaeda~~ \</ALTSP\> .
\</sub43\>\<strat DELETION\>\<duration 4.24\>\<TC not respecting\>

\<sub44\>
and \<CONDENSATION\> creating ~~beginning to create~~ \</CONDENSATION\> a coalition
\<br\>
of Islamic terror groups.
\</sub44\>\<strat CONDENSATION\>\<duration 4.16\>\<TC respecting\>

\<sub45\>
In Washington ,
the CIA set up a special task force
\</sub45\>\<strat VERBATIM\>\<duration 4.16\>\<TC respecting\>

\<sub46\>
to investigate what Osama Bin Laden
was planning from his base \<DELETION\> ~~in the Sudan~~ \</DELETION\> .
\</sub46\>\<strat DELETION\>\<duration 5.04\>\<TC not respecting\>

\<sub47\>
They used surveillance,
secret agents and intercepts.
\</sub47\>\<strat VERBATIM\>\<duration 4.08\>\<TC not respecting\>

\<sub48\>
They estimated he had ¤ ~~$~~ 200 million,
\</sub48\>\<strat VERBATIM\>\<duration 2.18\>\<TC not respecting\>

\<sub49\>
much of it in cash, some invested
through front companies in Khartoum.
\</sub49\>\<strat VERBATIM\>\<duration 4.22\>\<TC not respecting\>

\<sub50\>
The money, and what it paid for ,
was hard for CIA analysts to track.
\</sub50\>\<strat VERBATIM\>\<duration 4.19\>\<TC not respecting\>

STANLEY BEDLINGTON:
<sub51>
He always deliberately maintained
his distance from terrorist attacks.
</sub51><strat VERBATIM><duration 8.08><TC respecting>

<sub52>
When you look at <w PRP> <DELETION> Osama Bin Laden and his network,
</DELETION> his <w DPS> financial <w AJ0> <LINEBREAKERR>
network <w NN>, it <SPECIFICATION> is 's </SPECIFICATION> <DELETION> rather
</DELETION> like peeling an onion <DELETION> skin by skin </DELETION>.
</sub52><strat DELETION, SPECIFICATION, DELETION,
DELETION><duration_5.01><TC not respecting>

CORBIN:
<sub53>
<CONDENSATION> In During </CONDENSATION> the <ALTSP> mid-'90s mid-90s
</ALTSP> . <DELETION> some <DELETION> 20 twenty countries
were identified as places
</sub53><strat CONDENSATION, DELETION><duration 4.20><TC respecting>

<sub54>
where <CONDENSATION> terrorists terror groups </CONDENSATION3> were operating
with support from Osama Bin Laden.
</sub54><strat CONDENSATION><duration 4.21><TC not respecting>

<sub55>
There were attacks
in the Philippines,
</sub55><strat VERBATIM><duration 3.16><TC respecting>

<sub56>
the shooting of tourists in Egypt,
</sub56><strat VERBATIM><duration 2.18><TC not respecting>

<sub57>
bombings in Saudi Arabia ,
which killed American soldiers.
</sub57><strat VERBATIM><duration 4.18><TC not respecting>

<sub58>
Bin Laden's <ERRCORRECTION> aides ~~aids~~ </ERRCORRECTION>
<CONDENSATION> ~~were~~ even tried ~~trying~~ </CONDENSATION>to buy
components for chemical <DELETION> ~~and~~ </DELETION> <DELETION> ~~nuclear~~
</DELETION> weapons <DELETION> ~~on the black market~~ </DELETION>.
</sub58><strat CONDENSATION , DELETION, DELETION, DELETION><duration
6.01><TC respecting>

BLEDLINGTON:
<sub59>
He <DELETION> ~~sort of~~ </DELETION> acts as a financial godfather.
</sub59><strat DELETION><duration 2.11><TC Not respecting> aides

<sub60>
<TRANSPOSITIONa> ~~I don't think,~~ <TRANSPOSITIONa> To ~~to~~ my knowledge,
<TRANSPOSITIONa> I don't think </TRANSPOSITIONa> that <w CJT>
<LINEBREAKERR>
he's <DELETION> ~~actually~~ </DELETION> ordered <DELETION> ~~a group to go out and
carry out~~ </DELETION> a specific operation.
</sub60><strat TRANSPOSITION, DELETION, DELETION><duration 5.08><TC not
respecting>

<sub61>
<CONDENSATIONa> He ~~What he does is~~ <CONDENSATIONa> <DELETION> ~~sort of~~
</DELETION> <CONDENSATIONa> selects ~~select~~ <CONDENSATIONa> a group,
<DELETION> ~~and~~ </DELETION> then <CONDENSATIONa> <CONDENSATIONb>
finances <w VVZ> ~~give~~ <LINEBREAKERR>
it_Obj ~~some financing~~ <CONDENSATIONa> </CONDENSATIONb> and <DELETION>
~~then~~ </DELETION> <CONDENSATIONa> gives ~~give~~ </CONDENSATIONa> it its head,
as it were.
</sub61><strat CONDENSATION, DELETION, DELETION, CONDENSATION,
DELETION><duration 5.18><TC respecting>

CORBIN:
<sub62>
In 1998,
<DELETION> ~~Osama~~ </DELETION> Bin Laden announced a <DELETION> ~~formal~~
</DELETION> coalition
</sub62><strat DELETION, DELETION><duration 3.23><TC respecting>

<sub63>
of some of the world's
most feared terror groups , ~~.~~
</sub63><strat VERBATIM><duration 4.03><TC respecting>

<sub64>
amongst ~~Amongst~~ them ~~,~~ Egyptians, <DELETION> ~~held~~ </DELETION> responsible
<LINEBREAKERR>
for the massacre of 85 tourists . ~~,~~
</sub64><strat DELETION><duration 4.24><TC not respecting>

<sub65>
<DELETION> ~~and~~ </DELETION> He ~~he~~ pronounced a fatwa, calling for
attacks against American civilians.
</sub65><strat DELETION><duration 4.19><TC not respecting>

~~BIN LADEN: "All these crimes committed by the Americans are but a clear declaration of war against Allah, his Prophet and against Muslims. Obeying the order to kill the Americans and their allies, both civilian and military, is the duty of every muslim."~~

CORBIN:
<sub66>
In the Islamic world, doors_Subj<LINEBREAKERR>
{were closing against Bin Laden}_Predicate.
</sub66><strat VERBATIM><duration 4.18><TC not respecting>

<sub67>
Sudan <CONDENSATION> evicted ~~succumbed to pressure to evict~~ </CONDENSATION>

their troublesome guest.
</sub67><strat CONDENSATION><duration 3.19><TC respecting>

<sub68>
In Washington, presidential <w AJ0> ~~Presidential~~ <LINEBREAKERR
advisers <w NN> considered, but rejected .
</sub68><strat VERBATIM><duration 4.18><TC not respecting>

<sub69>
a raid to capture Bin Laden,
</sub69><strat VERBATIM><duration 2.13><TC respecting>

<sub70>
now hiding in Afghanistan .
under the protection of the <DELETION> ~~extremist~~ </DELETION> Taliban.
</sub70><strat DELETION><duration 5.03><TC not respecting>

<sub71>
Why was the decision taken <DELETION> ~~in the end~~ </DELETION> not to <w TO0>
<LINEBREAKERR>
go <w VVI> into Afghanistan and <CONDENSATION> get~~take~~ him~~out~~
</CONDENSATION> ?
</sub71><strat DELETION, CONDENSATION><duration 5.02><TC not respecting>

OLIVER REVELL:
<sub72>
I believe they thought there would <w VM0> <LINEBREAKERR>
<ADDITION> be <w VBI> </ADDITION> <ERRORTRANS> ~~National Executive~~
</ERRORTRANS> <DELETION> ~~too much collateral damage~~ </DELETION> ~~,~~too many
<ERROR> none-involved ~~non-involved~~ </ERROR> casualties
</sub72><strat ADDITION, DELETION><duration 6.15><TC respecting>

<sub73>
and <DELETION> ~~there would be~~ </DELETION>a high casualty rate
amongst <DELETION> ~~the~~ </DELETION>soldiers
</sub73><strat DELETION, DELETION><duration 3.11><TC not respecting>

<sub74>
and it would require
<DELETION> ~~a great deal of~~ </DELETION> logistical support ~~,~~ <DELETION> ~~in and out~~
</DELETION>.
</sub74><strat DELETION, DELETION><duration 3.11><TC not respecting>

CORBIN:
<sub75>
So, the decision was made to leave <w VVI> <LINEBREAKERR>
him_Obj there? ~~REVELL:~~ <ADDITION> Yes </ADDITION> , <DELETION> ~~Leave him there~~ </DELETION> for the time being.
</sub75><strat ADDITION, DELETION><duration 4.18><TC not respecting>

~~BIN LADEN VIDEO~~
<no corresponding subtitle>

CORBIN:
<sub76>
<DELETION> ~~From Afghanistan~~ </DELETION> Bin Laden continued to taunt the <w
AT0> <LINEBREAKERR>
American <w AJ0> President <w NN> <DELETION> ~~himself~~ </DELETION>, Bill
Clinton,
</sub76><strat DELETION, DELETION><duration 5.12><TC not respecting>

<sub77>
with his propaganda material.
</sub77><strat VERBATIM><duration 3.04><TC respecting>

<sub78>
GUNFIRE
</sub78><strat_ADDITION-AUDIO><duration_2.23><TC_respecting but too long>

<sub79>
But the terrorist leader
IS ~~is~~ very wary about his security.
</sub79><strat VERBATIM><duration 4.16><TC not respecting>

<sub80>
Five times now,
<DELETION>~~various~~</DELETION> individuals have tried to kill him.
</sub80><strat DELETION><duration 4.07><TC respecting>

<sub81>
and especially <ADDITION> the </ADDITION> Bin Laden team ,
if <PARAPHRASE> you ~~we~~ </PARAPHRASE> can use this phrase.
</sub81><strat ADDITION, PARAPHRASE><duration 4.20><TC not respecting>

<sub82>
Those people who believe in this~~,~~
</sub82><strat VERBATIM><duration 2.15><TC not respecting>

<sub83>
<DELETION> ~~they~~ </DELETION> may be even rich <ADDITION> or </ADDITION> ,
prosperous <w AJ0> <LINEBREAKERR>
people <w NN> , <ADDITION> and </ADDITION> <CONDENSATION> maybe ~~they~~
~~may be~~</CONDENSATION> <DELETION> ~~very~~ </DELETION> highly educated . ,
</sub83><strat DELETION, ADDITION, ADDITION, CONDENSATION,
DELETION><duration 5.00><TC not respecting>

<sub84>
<DELETION> ~~and~~</DELETION> We ~~we~~ have a very clear example now --
pilots ~~Pilots~~ , well trained . ,
</sub84><strat DELETION><duration 4.20><TC respecting>

<sub85>
They ~~they~~ elect to kill themselves
</sub85><strat VERBATIM><duration 2.15><TC respecting>

<sub86>
in this exceptional manner.
</sub86><strat VERBATIM><duration 3.09><TC respecting>

CORBIN:
<sub87>
As dawn broke over <w PRP> the <w AT0> shattered <w AJ0> <LINEBREAKERR>
skyline <w NN> of New York on <TRANPOSITIONa> 12 <TRANPOSITIONa>
<DELETION> ~~the morning of~~ </DELETION> September <TRANPOSITIONa> ~~12th~~
</TRANPOSITIONa>~~.~~
</sub87><strat TRANSPOSITION, DELETION><duration 5.00><TC not respecting>

<sub88>
the intelligence ~~Intelligence~~ community believed
</sub88><strat VERBATIM><duration 2.13><TC not respecting>

<sub89>
they had a concrete indication as <w PRP21> <LINEBREAKERR>
to <w PRR22> who was ultimately responsible
</sub89><strat VERBATIM><duration 5.08><TC not respecting>

<sub90>
for masterminding the <DELETION> ~~devastating~~ </DELETION> attack
on America.
</sub90><strat DELETION><duration 3.04><TC not respecting>

STANLEY BEDLINGTON:
<sub91>
<DELETION> ~~Immediately~~ </DELETION> After ~~after~~ the attack
there were intercepts,
</sub91><strat DELETION><duration 3.15><TC respecting>

<sub92>
<DELETION> ~~one~~ </DELETION> </DELETION> ~~in particular~~ </DELETION> ~~,~~
</DELETION> between people we presumed
to be <CONDENSATION> Bin Laden supporters ~~of Osama Bin Laden~~
</CONDENSATION>
</sub92><strat DELETION, DELETION, CONDENSATION><duration 4.15><TC
respecting>

<sub93>
in <DELETION> either </DELETION> Washington, <DELETION> and </DELETION>
New York <CONDENSATION> City </CONDENSATION> or both ,
</sub93><strat DELETION, DELETION, CONDENSATION><duration 2.20><TC
respecting>

<sub94>
who sent messages
that were intercepted saying,
</sub94><strat VERBATIM><duration 4.15><TC respecting>

<sub95>
" : Both both jobs successful " : .
</sub95><strat VERBATIM><duration 2.23><TC respecting>

CORBIN:
<sub96>
More than 600 people were dead . ,
</sub96><strat VERBATIM><duration 3.03><TC respecting>

<sub97>
Over over <ALTSP> 5,000 5000 </ALTSP> <ADDITION> were </ADDITION> missing.
</sub97><strat ADDITION><duration 2.15><TC respecting>

<sub98>
President Bush toured
the devastated streets.
</sub98><strat VERBATIM><duration 4.05><TC respecting>

<sub99>
<PARAPHRASE> Further implications of <w PRP> Bin Laden <w NP0> 's <w POS>
<LINEBREAKERR>
guilt <w NN> His authorities had been <CONDENSATION> found established
</CONDENSATION> further indications of Bin Laden's guilt </PARAPHRASE>.
</sub99><strat PARAPHRASE, CONDENSATION><duration 4.15><TC not respecting>

<sub100>
A Yemeni linked to <DELETION> the bombing of </DELETION> the USS Cole
was confirmed <CONDENSATION> as a hijacker amongst the group of hijackers
</CONDENSATION>.
</sub100><strat DELETION, CONDENSATION><duration 5.00><TC respecting>

<sub101>
Another member <DELETION> ~~of the group~~ </DELETION> was believed
to have trained
</sub101><strat DELETION><duration 3.20><TC respecting>

<sub102>
in a Bin Laden <DELETION> ~~terror~~ </DELETION> camp
before going to America.
</sub102><strat DELETION><duration 3.07><TC not respecting>

<sub103>
For President Bush this was war.
</sub103><strat VERBATIM><duration 2.15><TC not respecting>

BUSH:
<sub104>
I can hear you!
</sub104><strat ADDITION-AUDIO><duration 2.05><TC respecting>

<sub105>
CHEERING
</sub105><strat ADDITION-AUDIO><duration 3.10><TC respecting>

<sub106>
I can hear you . ~~,~~
The ~~the~~ rest of the world hears you .
</sub106><strat VERBATIM><duration 4.20><TC respecting>

(cheers from the people)
<no corresponding subtitle>

<sub107>
And the people...
</sub107><strat ADDITION-AUDIO><duration 2.05><TC respecting>

<sub108>
CHEERING
</sub108><strat ADDITION-AUDIO><duration 2.05><TC respecting>

<sub109>
And ~~and~~ the people
who knocked these buildings down
</sub109><strat VERBATIM><duration 4.05><TC respecting>

<sub110>
will hear all of us soon.
 APPLAUSE CHEERING
</sub110><strat VERBATIM+ADDITION-AUDIO><duration 3.01><TC respecting>

CORBIN:
<sub111>
Osama Bin Laden <LINEBREAKERR>
denies he is responsible.
</sub111><strat VERBATIM><duration 3.15><TC respecting>

<sub112>
<DELETION> and </DELETION> He' he' ll be hard to smoke out
of his refuge in Afghanistan.
</sub112><strat DELETION><duration 3.22><TC not respecting>

<sub113>
In Kabul ,
people are <DELETION> leaving </DELETION> , heading for the border.
</sub113><strat DELETION><duration 3.15><TC not respecting>

<sub114>
The Taliban have <CONDENSATION> been given a three days day ultimatum
</CONDENSATION>
</sub114><strat CONDENSATION><duration 2.15><TC respecting>

<sub115>
to give up the fugitive <DELETION> terrorist </DELETION>
or face <CONDENSATION> America's the wrath of America </CONDENSATION>.
</sub115><strat DELETION, CONDENSATION><duration 3.22><TC not respecting>

<sub116>
<DELETION> But </DELETION> Even even if they can destroy Bin Laden,
this won't end the terror threat.
</sub116><strat DELETION><duration 5.19><TC respecting>

Dr SAAD AL FAGIH:
<sub117>
Now if they intend to destroy Bin <w NP021> <LINEBREAKERR>
Laden<w NP022> , <DELETION> Afghanistan, </DELETION> they
<CONDENSATION> will would have done </CONDENSATION> nothing .
</sub117><strat DELETION, CONDENSATION><duration 5.04><TC not respecting>

<sub118>
<DELETION> ~~because~~ <DELETION> <u>Most</u> ~~most~~ of the bulk of the followers
<u>is outside Afganistan .</u> ~~,~~
</sub118><strat DELETION><duration 4.15><TC not respecting>

<sub119>
<DELETION> ~~and~~ </DELETION> <ERRCORRECTION> <u>There is</u> ~~they are~~
</ERRCORRECTION> much more intelligence now,
</sub119><strat DELETION><duration 2.15><TC not respecting>

<sub120>
much more sophisticated~~.~~ <DELETION> ~~They can revive themselves much easily~~
</DELETION> ~~,~~
and that is the other dangerous side of it.
</sub120><strat DELETION><duration 5.13><TC not respecting>

Appendix 4: Tagset used for POS tagging & segmentation rules

The tagset used is the UCREL Claws5 tagset (see Garside et al., 1997, pp. 256-257) with minor alterations.

AJ0   adjective (unmarked) (e.g. *magnificent*, *spectacular*)

AJS   superlative adjective (e.g. *closest*)

AT0   article (e.g. *the*, *a*, *an*)

AV0   adverb (unmarked) (e.g. *largely*, *here*)

AVQ   wh-adverb (e.g *why*)

CJC   coordinating conjunction (e.g. *and*, *or*)

CJT   the conjunction *that*

DPS   possessive determiner form (e.g. *my*, *his*)

DT0   general determiner (e.g. *these*, *many*)

NN   noun (either singular or plural) (e.g. *life*, *monuments*)

NP0   proper noun (e.g. *Bill Clinton*)

PNP   personal pronoun (e.g. *they*, *he*)

PNQ   wh-pronoun (e.g. *what*, *who*)

POS   the possessive or genitive *'s* or *'*

PRP   preposition (e.g. *about*, *from*)

TO0   the infinitive marker *to*

VBB   the 'base forms' of the verb *be* (except the infinitive), i.e. *am*, *are*

VBD   past form of the verb *be,* i.e. *was, were*

VBG   -ing form of the verb *be*, i.e. *being*

VBI   infinitive of the verb *be*

VBN	past participle of the verb *be*, i.e. *been*
VBZ	-s form of the verb *be*, i.e. *is, 's*
VDD	past form of the verb *do*, i.e. *did*
VDG	-ing form of the verb *do*, i.e. *doing*
VDI	infinitive of the verb *do*
VDN	past participle of the verb *do*, i.e. *done*
VHB	base form of the verb *have* (except the infinitive), i.e. *have*
VHD	past tense form of the verb *have*, i.e. *had, 'd*
VHG	-ing form of the verb *have*, i.e. *having*
VHI	infinitive of the verb *have*
VHN	past participle of the verb *have*, i.e. *had*
VHZ	-s form of the verb *have*, i.e. *has, 's*
VM0	modal auxiliary verb (e.g. *may*, *would*)
VVD	past tense form of lexical verb (e.g. *conducted*, *thought*, *took*)
VVG	*-ing* from of lexical verb (e.g. *uncovering*, *looking*)
VVI	infinitive of lexical verb (e.g. *join*)
VVN	past participle form of lexical verb (e.g. *written*, *started*)
VVZ	*-s* form of lexical verb (e.g. *depends*)
XX0	the negative *not* or *n't*

Furthermore, as noted by the researchers, any of the tags may in theory be modified by the addition of a pair of numbers to it: e.g. PRP 13 PRP 23 PRP33. This notation signifies that the tag occurs in a sequence of similar tags, representing a sequence of words which for grammatical purposes are treated as a single unit (for instance *in terms of*).

Appendix 5: Deletion in *Atlantis Uncovered*

Number	Transcript	Subtitle	Deletions	Type of deletion
Sub 10	When scientists *first* flew over these peaks 60 years ago they thought they were the tops of volcanoes.	When scientists flew over here 60 years ago, they thought these peaks were the tops of volcanoes.	first (1)	Additional information (time, adjunct)
Sub 13	*I could see Nakbe on the horizon and* I knew *that* it needed to be investigated because of the implications that it had for the development of civilisation.	I knew Nakbe had to be investigated because of its implications for the development of civilisation.	- I could see Nakbe on the horizon (2) - and (3) - that (4)	- Additional information when there is lots of text and not much time - Conjunction - Conjunction
Sub 15	to a *great and* magnificent ancient civilisation	to a magnificent ancient civilisation?	- great (5) - and (6)	- Similar meaning - Conjunction
Sub 16	*and* I felt *like* it was necessary to get to Nakbe and understand that.	I felt it was necessary to get to Nakbe and understand it.	- and (7) - Like (8)	- Conjunction - Grammatically redundant item
	It's a wonderful opportunity to view the the human process in action through an ancient window.	It's a wonderful opportunity to view thehuman process in action through an ancient window.	- the (82)	- Repetition/hesitation
Sub 23	but if they dig deep enough *at places like Nakbe* perhaps they'll find they're wrong.	But if they dig deep enough, they may find they're wrong.	- at places like Nakbe (9)	Additional information (place, adjunct)

Sub 24	Archaeologists have been discovering lost cities *like Nakbe* since the first Europeans went in search of colonies abroad.	Archaeologists have discovered lost cities since Europeans first went in search of colonies abroad.	- like Nakbe (10)	Additional information (example/comparison, adjunct)
Sub 29	who went out to, *to* some horrible place *in the middle of nowhere* and discovered some ancient kingdom *or lost civilisation*	He went out to some horrible place and discovered some ancient kingdom.	- to (11) - in the middle of nowhere (12) - or (13) - lost civilisation (14)	- Repetition/hesitation - Additional information (place, adjunct) - Conjunction - Similar meaning
Sub 30	and *after all*, if you have people finding forgotten tombs in Ancient Egypt	And, if you have people finding forgotten tombs in ancient Egypt	after all (15)	Discourse marker (padding expression / filler)
Sub 31	and *people finding* lost civilisations in the middle of the tropical rainforest of Mexico	and lost civilisations in the middle of a tropical rainforest in Mexico,	people finding (16)	Repetition – emphasis (see subtitle 30)
Sub 32	*certainly* must have peeked people's curiosity at the time.	it must have piqued people's curiosity at the time.	certainly (17)	Comment disjunct
Sub 35	ancient civilisations on both sides of the Atlantic *chose to* build pyramids.	ancient civilisations on both sides of the Atlantic chose pyramids.	chose to (18)	Grammatically redundant item (because of subtitle 34 - *Of all the structures they could have built,*)
Sub 38	Archaeologists had no way of dating *most of* these wonders.	Archaeologists had no way of dating these wonders.	most of (19)	Quantifier

Sub 41	This is *all* part of a long tradition of architectural development and *a long tradition of* intense religious belief.	This was part of a long tradition of architectural development and intense religious belief.	- all (20) - a long tradition of (21)	- Quantifier - Repetition – emphasis
Sub 44	Standing alone in the Oaxaca Valley in *southern* Mexico is a place even more revealing than Nakbe: Monte Alban.	Standing alone in the Oaxaca Valley in Mexico is a place even more revealing than Nakbe - Monte Alban.	- southern (22)	Additional information (origin, adjective)
Between sub 51-52	*...use anything that's in the corner of this wall...*	-----------	...use anything that's in the corner of this wall... (23)	Overlapping speech
Sub 55	*where we have* a wall *and then we have* another house foundation *that was* placed over that wall	a wall, another house foundation placed over it,	- where we have (24) - and (25) - then (26) - we have (27) - that was (28)	- Grammatically redundant information -Conjunction - Conjunct - Grammatically redundant information - Grammatically redundant items (relative pronoun + auxiliary verb from relative clause)
Sub 56	and *then* a plaster floor *that was put* over that *later*	and a plaster floor over that.	- then (29) - that was put (30) - later (31)	- Conjunct - Grammatically redundant items (relative pronoun + auxiliary verb + verb from relative clause) - Additional information (time, adjunct)

Sub 57	*and* in several areas *we've got 3 or* 4 levels of houses.	Sometimes, four levels of houses.	- and (32) - we've got (33) - 3 (34) - or (35)	- Conjunction - Grammatically redundant information - Numeral - Conjunction
Sub 58	Layer after painstaking layer *throughout the valley*	Layer after painstaking layer,	throughout the valley (36)	- Additional information (place, adjunct)
Sub 60	*It's been fascinating*, it's been a surprise every day	It's been a surprise every day.	It's been fascinating (37)	- Similar meaning
Sub 61	As soon as we started it was *always oh* *look at this*, here's another wall, *oh, oh look at this,* I've got plaster,	As soon as we began, it was "Here's a wall!" "I've got plaster!"	- always (38) -oh (39) - look at this (40) - oh, oh, look at this (41)	- Additional information (time, adjunct) - Exclamation - Repetition - emphasis - Repetition - emphasis
Sub 62	*oh* look at this, *that we've got,* I've got a, *a* burial.	"Look at this - I've got a burial."	- oh (42) - that we've got (43) - a (44)	-Exclamation - Redundant information - Repetition/hesitation
Sub 63	Gary, we've got another offering *here*	Gary? I've got another offering. >	here (45)	Additional information (place, adjunct)
Between sub 65-66	*Oh boy.*	--------	Oh boy. (46)	Exclamation
Sub 67	*It might be*, he might have been buried with a dog.	He might have been buried with a dog.	It might be (47)	False start

Sub 68	*I think there's a dog in here...* *Oh my.* Yeah, *yeah* that's a dog.	Yeah, that's a dog.	- I think there's a dog in here... (48) - Oh my. (49) - yeah (50)	- Overlapping speech - Exclamation - Repetition - emphasis
Between sub 68-69	*That's a dog.*	---------	That's a dog. (51)	Overlapping speech
Between sub 68-69	*That's a dog.*	---------	That's a dog. (52)	Overlapping speech
Sub 69	*Little by little,* archaeologists *throughout the valley* have revealed *no less than* the birth of a new civilisation:	Archaeologists have revealed the birth of a new civilisation -	- Little by little, (53) - throughout the valley (54) - no less than (55)	- Additional information (manner, adjunct) - Additional information (place, adjunct) - Emphatic device
Sub 71	*Well* what we see *here* is *a lot of* signs of continuity in terms of how people built their houses	We see signs of continuity in terms of how people built their houses,	- Well (56) -here (57) - a lot of (58)	-Discourse marker - Additional information (place, adjunct) - Quantifier
Sub 72	how people buried their dead *and* these things bear continuities to patterns	how people buried their dead... These things bear continuities	And (59)	Conjunction
Sub 73	to patterns that we see back to the earliest occupations at Monte Alban *and sometimes even before.*	to patterns in the earliest occupations at Monte Alban.	- and (60) - sometimes (61) - even before (62)	- Conjunction - Additional information (time/frequency, adjunct) - Additional information (time, adjunct)

Sub 74	Archaeology has revealed the startling fact that in one _small_ corner of the world	Archaeology has revealed the startling fact that in one corner of the world	small (63)	Additional information (size, adjective)
Sub 75	the _long_ road to a _similar_ monument building society was taken by at least two _different_ groups of people independently,	the road to a monument-building society was taken by at least two groups of people independently.	- long (64) - similar (65) - different (66)	- Additional information (size, adjective) - Additional information (judgement, adjective) - Additional information (judgement, adjective)
Sub 82	but how did that agriculture come about? _That was a big question._	But how did that agriculture come about?	That was a big question. (67)	Additional information when there is lots of text and not much time
Sub 86	A life of foraging _high_ in the mountains where water was a long way off.	a life of foraging in the mountains, where water was a long way off.	high (68)	Additional information (place, adjunct)
Sub 90	_but_ the first thing they grew wasn't food. They grew bottle gourds which they used to carry water.	The first thing they grew wasn't food. They grew bottle gourds, which they used to carry water.	but (69)	Conjunction
Sub 92	It was some 6,000 years after they began planting gourds that people here _finally_ chose to settle in villages.	It was some 6,000 years after they began planting gourds that people here chose to settle in villages.	finally (70)	Conjunct
Sub 93	By that time they were growing _very_ different foods to _those of_ the near eastern farmers:	Then, they were growing different foods to the Near Eastern farmers -	- very (71) - those of (72)	-Intensifier - Grammatically redundant item (inferred from the text)

Sub 108	*The National Socialists in Germany,* the Nazis, had the notion of Aryan supremacy	the Nazis had the notion of Aryan supremacy	The National Socialists in Germany, (73)	- Similar meaning
Sub 115	we are better off knowing that civilisations developed *more or less* independently	we're better off knowing that civilisations developed independently	more or less (74)	Hedge
Sub 117	*because* we know what happens down the line when we believe that,	We know what happens down the line when we believe that.	because (75)	Conjunction
Sub 118	*so* I'm not going to tell you that belief in Atlantis is necessarily the first step towards genocide, or Holocaust,	I'm not going to say that belief in Atlantis is necessarily the first step towards genocide or holocaust.	so (76)	Conjunction
Sub 119	but *what I'm telling you is* we are on a very slippery slope if we believe in fantasies	But we are on a very slippery slope if we believe in fantasies.	what I'm telling you is (77)	Emphatic devices – cleft
Sub 120	*and that* those fantasies lead us *down* to places we *really* don't want to go.	Fantasies lead us to places we don't want to go.	- and (78) - that (79) -down (80) - really (81)	- Conjunction - Conjunction - Grammatically redundant item - Intensifier

Appendix 6: Deletion in *Chimps on Death Row*

Number	Transcript	Subtitle	Deletions	Type of deletion
Sub 3	There's no question that the use of chimpanzees presents a greater moral dilemma	The use of chimpanzees presents a greater moral dilemma	- There's no question (1) - that (2)	- Content disjunct - Conjunction
Sub 4	than the use of any other non-human animal for research.	than the use of any other animal for research.	non-human (3)	- Redundant
Sub 7	I hope some day that, that when we look back on this time we're, we're keeping chimpanzees in captivity against their will	I hope we look back on this time where we keep chimpanzees captive	- some day (4) - that (5) -that (6) - when (7) - we're (8) - against their will (9)	- Additional information (time, adjunct) - Conjunction - Repetition/hesitation - Error - Repetition/hesitation - Additional information (manner, adjunct)
Sub 8	and with the same shame that I look back on my great-grandfather who was a slaver holder	with the shame that I feel for my great-grandfather as a slaveholder.	- and (10) - same (11)	- Error - Additional information (value, adjective)
Sub 9	and I hope that some day my great-grand-children will look back on us with shame.	My great-grandchildren should look at us with shame.	- and (12) - that (13) - some day (14)	- Conjunction - Conjunction - Additional information (time, adjunct)
Sub 11	which set them apart from monkeys and the rest of the animal kingdom.	which set them apart from the rest of the animal kingdom.	- monkeys (15) - and (16)	- Phrases conveying similar/subordinate meaning - Conjunction

Sub 13	They have brains and central nervous systems <u>more</u> like ours <u>than that of any other living creature</u>	They have brains and central nervous systems similar to ours.	more [like ours] than that of any other living creature (17)	- Secondary importance information, too much to transcribe
Sub 15	<u>and</u> in addition, <u>without any question</u>, they have emotions like joy and sorrow, fear and no, mental as well as physical suffering.	In addition, they have emotions like joy, sorrow and fear, and know mental AND physical suffering.	- and (18) - without any question (19)	- Conjunction - Content disjunct
Sub 22	which are a plague or a pest of <u>the</u> human kind	which are a plague or a pest to humankind.	the (20)	Error
Sub 24	of HIV, AIDS, malaria and hepatitis <u>research</u> of course.	of HIV, AIDS, malaria and hepatitis, of course.	research (21)	Grammatically redundant information (*research* also used in subtitle 23)
Sub 28	whereon <u>beforehand</u> it is known that the chimpanzee may die within the experiments.	where it is known that the chimp may die within the experiment.	beforehand (22)	Additional information (time, adjunct)
Sub 30	A chimpanzee will reach the honourable age of 50 <u>years</u> in its life	The chimpanzee will reach the honourable age of 50 in its life	years (23)	Redundant item
Sub 31	and we <u>also</u> have the policy not to euthanise these chimpanzees.	and we do not euthanise these chimpanzees.	also (24)	Conjunct
Sub 35	and what future awaits the <u>thousands of</u> survivors.	and what future awaits the survivors.	thousands of (25)	Quantifier
Sub 38	a zoo of dogs, pigs and bears was assembled to test the effects of high altitude flight <u>on a mammal's body.</u>	a zoo of dogs, pigs and bears was assembled to test the effects of high-altitude flight.	on a mammal's body (26)	Less important information and there is lots of text and not much time

Sub 41	We'd used them in just about everything that, that you can use a chimp for.	We used them in everything you can use a chimp for.	- just about (27) - that (28) - that (29)	- Intensifier - Conjunction - Repetition/hesitation
Sub 42	There was a surplus developing of animals that had been exposed to hepatitis B	There was a surplus of animals that had been exposed to hepatitis B,	developing (30)	Redundant item
Sub 46	I guess at that point in time there was consideration of euthanasia as a population control method.	Euthanasia was considered as a population control method.	- I guess (31) - at that point in time (32)	- Hedge - Additional information (time, adjunct)
Sub 54	As it turns out the only animal that's known that gets infected with HIV 1,	The only animal that's known that gets infected with HIV 1,	As it turns out (33)	Redundant item
Sub 55	which is the form of the virus that causes AIDS in humans, was the chimpanzee.	which causes AIDS in humans, was the chimpanzee.	is the form of the virus that (34)	Not primary information, much text to transcribe
Sub 57	Far from having a surplus, primate centres around the world stepped up their breeding programmes	Far from having a surplus, primate centres stepped up their breeding programmes	around the centre (35)	Additional information (place, adjunct)
Sub 58	to compensate for all the animals they thought would soon be dying of AIDS.	to compensate for the animals they thought would soon be dying of AIDS.	all (36)	Quantifier
Sub 59	When we first started putting the material in the animal we were continuously looking for disease.	When we first put material in the animal, we were continuously looking for disease.	the (37)	Shouldn't have been deleted

Sub 60	The disease didn't come, you know, for 2 and 3 and 4 years, but as we know it's a very latent infection	The disease didn't come for two, three, four years, but it's a very latent infection.	- you know (38) - and (39) - and (40) - as we know (41)	- Padding expression - Conjunction - Conjunction - Clause intended to elicit support from the speaker
Sub 61	and what we had to consider was the fact that it may lay dormant for years before it actually caused infection in the chimp,	We had to consider that it may lay dormant for years before causing infection in the chimp.	- and (42) - the fact (43) - actually (44)	- Conjunction - Redundant item - Padding expression
Sub 62	so probably for 5 or 6 years people were just thinking well, we just haven't had enough time,	For five or six years, people just thought we hadn't had enough time.	- so (45) - probably (46) - well (47) - just (48)	- Conjunction - Content disjunct - Padding expression - Exclusive adverb (focus adjunct)
Sub 63	but once we reached that time there was the opinion that more than likely this isn't going to be a disease model.	Once we reached that time, the opinion was that this wouldn't be a disease model.	- but (49) - more than likely (50)	- Conjunction - Content disjunct
Sub 66	They are still a health hazard to humans and have to be housed separately from other, clean chimps.	They are a health hazard and are housed separately from other chimps.	- still (51) - to humans (52) - clean (53)	- Additional information (time, adjunct) - Redundant information, can be inferred - Redundant information, can be inferred
Sub 69	You see one of our HIV-infected chi, chimpanzees	You see one of our HIV-infected chimpanzees.	chi (54)	-False start
Sub 70	His named is called Koo.	He is called Koon.	His named (55)	Error
Sub 78	When it's warm enough and not raining the roof folds back to give them a glimpse of the outside world.	If it's warm, the roof folds back and they can see the outside world.	- enough (56) - and not raining (57)	- Intensifier - Secondary importance information, too much to transcribe
Sub 79	AIDS may have meant a stay of execution, but *now* the chimps face a life sentence.	AIDS has meant a stay of execution, but the chimps face a life sentence.	now (58)	- Additional information (time, adjunct)

Sub 81	Bids to support the chimps came from groups wanting to retire them, and also from the Coulston Research Foundation.	Bids to help the chimps came from groups wanting to retire them and the Coulston Research Foundation.	- also (59) - from (60)	- Conjunct - Grammatically redundant item
Sub 83	They are like human beings that are kept in a hospital, in a clinic as they get older.	They are like human beings kept in a hospital as they get older.	- that are (61) - in a clinic (62)	- Grammatically redundant items (relative pronoun + auxiliary verb from relative clause) - Near-synonym repetition
Sub 84	We will keep them in the same way, we will treat them with, for their diseases,	We will keep them in the same way. We'll treat them for their diseases.	with (63)	False start (error corrected)
Sub 87	To just take these valuable animals and throw 'em out in a field in, in South Texas.	To take these valuable animals and throw them out in a field in Texas?!	-just (64) -in (65) - South (66)	- Exclusive adverb - Repetition/hesitation - Additional information (location, noun as adjective)
Sub 93	Neither the drugs companies who'd used them nor the government who founded the colony, nor the public were willing to provide the money.	Neither the drugs companies, the government nor the public would provide the money.	- who'd used them (67) - nor (68) - who founded the colony (69)	- Secondary importance information/much time, not much text - Conjunction - Secondary importance information/much time, not much text
Sub 98	and I hope you aren't too offended by this, but that is a normal thing for a chimp to want to do.	I hope you are not offended, but that is a normal thing for a chimp to want to do.	- and (70) - too (71) - by this (72)	- Conjunction - Intensifier - Redundant item
Sub 99	It doesn't make him unnecessarily nasty or...	It doesn't make him unnecessarily nasty.	or... (73)	Unfinished speech
Sub 100	Look at him, he's laughing now, you see, you see.	Look at him, he's laughing now. You see?	you see (74)	Repetition – emphasis

Sub 101	<u>There</u> he was displaying just a moment ago <u>and</u> now he's laughing <u>and now he's laughing</u> with his head up	He was displaying just a moment ago. Now he's laughing with his head up.	- there (75) - and (76) - and now he's laughing (77)	- Deictic expression - Conjunction - Repetition/hesitation
Sub 102	<u>and</u> he wants to play tickle and chase.	He wants to play tickle and chase.	and (78)	- Conjunction
Sub 104	Booee escaped AIDS research, but <u>he</u> was infected with hepatitis	Booee escaped AIDS research, but was infected with hepatitis,	he (79)	- Grammatically redundant item
Sub 105	so <u>he</u> has to be housed separately from <u>the</u> other chimps <u>here</u>.	so is housed separately from other chimps.	- he (80) - the (81) - here (82)	- Grammatically redundant item - Grammatically redundant item - Additional information (place, adjunct)
Sub 112	I <u>simply</u> prioritise the advancement of human health and welfare	I prioritise the advancement of human health and welfare	simply (83)	Exclusive adverb
Sub 116	<u>we</u> provide 'em a high quality life at a stabilised level of funding.	by giving them a high quality life at a stable level of funding.	we (84)	Error
Sub 117	Since chimpanzees are our closest living relatives, since <u>in their own right</u> they're wondrous and amazing beings,	Since chimpanzees are our closest living relatives, since they're wondrous and amazing beings,	in their own right (85)	- Redundant item, cliché
Sub 120	and we're not prepared to <u>get together and</u> sort out a decent future for them.	and we are not prepared to sort out a decent future for them.	- get together (86) - and (87)	- Less important information, can be inferred too, too much text - Conjunction

Appendix 7: Generalised Phrase Structure Grammar categories

AP = adjectival phrase

NP = noun phrase

PP = prepositional phrase

S = sentence

S′ (Sbar) = dependent clause

VP = verb phrase

Vaux = auxiliary verb

Vtrans= transitive verb

Vcop = copular verb

- The following notations allow us to deal with coordination (in all cases a similar phrase precedes the coordinating conjunction):

APplus = coordinating conjunction + AP

NPplus = coordinating conjunction + NP

Splus = coordinating conjunction + S

VPplus = coordinating conjunction + VP

- We have also used two non-standard notations to denote:

NP-subj = a subject NP

NP-obj= a direct object NP

Appendix 8: Segmentation in *Atlantis Uncovered*

No.	Transcript	Subtitle	Type of line-break error	Alternative
Sub 6	but a few survivors escaped in ships and brought civilisation to the primitive peoples around the world.	But some survivors escaped in ships (35) and brought civilisation <u>to the</u> (31) <u>primitive peoples</u> around the world. (35)	**PRP** AT0/(AJ0) **NN**	Tentative suggestion with major editing: But some survivors escaped (26) in ships and brought civilisation (33) to the primitive peoples (24) around the world. (17)
Sub 10	When scientists first flew over these peaks 60 years ago they thought they were the tops of volcanoes.	When scientists flew over here (30) 60 years ago, they thought <u>these</u> (32) <u>peaks</u> were the tops of volcanoes. (33)	DT0/NN	When scientists flew over here (30) 60 years ago, they thought (26) these were the tops of volcanoes. (33)
Sub 14	What were the processes that led a simple hunting and gathering group	What were the processes that led <u>a</u> (34) simple hunting and gathering <u>group</u> (34)	AT0/NN	What were the processes that led (32) a simple hunting and gathering group (36)
Sub 15	to a great and magnificent ancient civilisation	<u>to a magnificent</u> (16) <u>ancient civilisation</u>? (21)	**PRP** AT0 AJ0 /AJ0 **NN**	to a magnificent civilisation? (30)
Sub 24	Archaeologists have been discovering lost cities like Nakbe since the first Europeans went in search of colonies abroad.	Archaeologists have discovered (30) lost cities since Europeans <u>first</u> (33) <u>went</u> in search of colonies abroad. (34)	AV0/VVD	Archaeologists have discovered (30) lost cities since Europeans (27) first went in search of colonies. (33)
Sub 26	they stumbled upon the spectacular remains of ancient societies.	they stumbled <u>upon the spectacular</u> (34) <u>remains</u> of ancient societies. (29)	**PRP** AT0 AJ0/**NN**	Can't suggest

Sub 27	The locals often knew nothing about the people who'd built these monuments.	The locals often knew nothing <u>about</u> (35) <u>those</u> who had built these monuments. (36)	PRP/DT0	The locals knew nothing about those (35) who had built these monuments. (30)
Sub 28	Archaeology was conducted largely by adventurers and rogues	Archaeology was <u>conducted</u> (25) <u>largely</u> by adventurers and rogues. (34)	VVD/AV0	Archaeology was conducted largely (33) by adventurers and rogues. (26)
Sub 36	Of all the ways of disposing of their dead they had mummified them.	Of all the ways of disposing <u>of</u> (31) <u>their dead</u>, they had mummified them. (36)	**PRP**/DPS **NN**	Of all the ways of disposing(28) of their dead, they mummified them. (35)
Sub 37	Instead of alphabets they had written in strange pictures - hieroglyphs.	Instead of alphabets, they <u>had</u> (30) <u>written</u> in strange pictures – (29) hieroglyphs. (12)	VHD/VVN	Instead of alphabets, they wrote (32) in strange pictures - hieroglyphs. (34)
Sub 40	It was logical to think they'd all started as one unique culture that spread,	It was logical to think they'<u>d</u> (30) all <u>started</u> as one unique culture (33) that spread. (12)	VHD/VVN	It was logical to think (23) they'd all started (18) as one unique culture that spread. (34)
Sub 44	Standing alone in the Oaxaca Valley in southern Mexico is a place even more revealing than Nakbe: Monte Alban.	Standing alone in the Oaxaca Valley (35) in Mexico is a place even <u>more</u> (30) <u>revealing</u> than Nakbe - Monte Alban. (35)	AV0/AJ0	Can't suggest
Sub 45	The first archaeologists who came here couldn't tell who had built it.	The first archaeologists who <u>came</u> (33) <u>here</u> couldn't tell who had built it. (36)	VVH/AV0	The first archaeologists here (29) couldn't tell who had built it. (31)
Sub 49	older even than the first Mayan monuments of this scale.	Older, even, <u>than the first</u> (27) <u>Mayan monuments</u> of this scale. (30)	**PRP** AT0 AJ0/AJ0 **NN**	Can't suggest
Sub 52	Gary and Linda Feinman are uncovering a new site - El Palmillo.	Gary and Linda Feinman <u>are</u> (26) <u>uncovering</u> a new site - El Palmillo. (36)	VBB/VVG	Can't suggest
Sub 53	...there's a second pot underneath, but that also appears to be broken.	..The second part underneath, <u>but</u> (33) that also appears to be broken. (31)	CJC	..The second part underneath, (29) but that also appears to be broken. (35)

Sub 70	the Zapotec, who evolved as gradually and as independently as did the Maya.	the Zapotec – (13) who evolved as gradually <u>and</u> (28) as independently as did the Maya. (33)	CJC	the Zapotec, (12) who evolved as gradually(24) and independently as did the Maya. (34)
Sub 71	Well what we see here is a lot of signs of continuity in terms of how people built their houses,	We see signs of continuity <u>in terms</u> (35) <u>of</u> how people built their houses, (33)	PRP31 PRP32/PRP33	Can't suggest
Sub 73	to patterns that we see back to the earliest occupations at Monte Alban and sometimes even before.	to patterns <u>in the earliest</u> (27) <u>occupations</u> at Monte Alban. (27)	PRP AT0 AJS/NN	Can't suggest
Sub 75	the long road to a similar monument building society was taken by at least two different groups of people independently,	the road <u>to a monument-building</u> (31) <u>society</u> was taken by at least two (33) groups of people independently. (31)	**PRP** AT0 AJ0/**NN**	Tentative suggestion with major editing involving subtitle A74 that the long road (18) to a monument-building society (30) was taken by at least two groups (32) of people independently. (24)
Sub 77	There's a simple reason why so many ancient peoples built pyramids.	There's a simple reason <u>why so</u> <u>many</u> (35) <u>ancient peoples</u> built pyramids. (31)	**AVQ** AV0 DT0/(AJ0) NN	Can't suggest
Sub 78	Before engineers had invented the dome, the spire or structural steel	Before engineers had invented <u>the</u> (33) <u>dome</u>, the spire or structural steel, (36)	AT0/NN	Before engineers invented the dome, (35) the spire or structural steel, (30)
Sub 79	a sloping pile was the only high structure you could build.	a sloping pile was <u>the</u> <u>only</u> (27) <u>high structure</u> you could build. (31)	**AT0** AJ0/AJ0 **NN**	Can't suggest
Sub 80	If you wanted a high showy monument when all there was to build with was heavy stone	If you wanted <u>a high, showy</u> (27) <u>monument</u>, when all there was (28) to build with was heavy stone, (30)	**AT0** AJ0 AJ0/**NN**	If you wanted a high monument, (30) when all there was to build with (32) was heavy stone, (16)

Sub 90	but the first thing they grew wasn't food. They grew bottle gourds which they used to carry water.	The first thing they grew wasn't (32) food. They grew bottle gourds, (30) which they used to carry water. (31)	Verb/ Complement	Tentative suggestion with major editing The first thing they grew (25) wasn't food, but bottle gourds, (31) which they used to carry water. (31)
Sub 91	This is the first known plant domestication in the Americas, prompted not by hunger but by practicality.	It was <u>the Americas' first-known</u> (32) <u>plant domestication</u> - prompted <u>by</u> (33) <u>practicality</u> rather than hunger. (32)	**AT0** AJ0/NN PREP/**NN**	Can't suggest Tentative suggestion with major editing It was the Americas' first-known (32) plant domestication (19) prompted not by hunger (22) but by practicality. (20)
Sub 92	It was some 6,000 years after they began planting gourds that people here finally chose to settle in villages.	It was some 6,000 years after <u>they</u> (34) <u>began planting gourds</u> that <u>people</u> (33) <u>here chose to settle in villages</u>. (33)	Subject/ Predicate x2	Tentative suggestion with major editing Some 6,000 years (16) after they began planting gourds, (33) people here (11) chose to settle in villages. (28)
Sub 93	By that time they were growing very different foods to those of the near eastern farmers:	Then, they were growing <u>different</u> (33) <u>foods</u> to the Near Eastern farmers – (35)	AJ0/NN	They were growing different foods (33) to the Near Eastern farmers – (29)
Sub 95	Had the Atlantians brought farming to the Americas there is no reason why they wouldn't have brought the same crops	Had the Atlanteans brought farming (34) to the Americas, why <u>wouldn't</u> (29) they <u>have brought</u> the same crops (32)	VM0/VHI VVN	Tentative suggestion with major editing Had the Atlanteans (18) brought farming to the Americas, (32) wouldn't they have brought the crops (36) that existed in the Near East? (30)

Sub 102	Every year crowds flock to ancient sites in search of lost wisdom.	Every year, crowds flock <u>to ancient</u> (35) <u>sites</u> in search of lost wisdom. (31)	**PRP** AJ0/**NN**	Can't suggest
Sub 103	Science continues to be ignored by a public yearning for the romance of a more mysterious past.	Science continues to be ignored (31) by a public yearning <u>for the</u> (28) <u>romance</u> of a more mysterious past. (34)	**PRP** AT0/**NN**	Science continues to be ignored (31) by a public yearning for the romance (36) of a more mysterious past. (26)
Sub 107	We've seen myths of that kind in our own time have tragic consequences.	We've seen myths of that kind <u>in our</u> (36) <u>own time</u> have tragic consequences – (35)	**PRP** DPS/ DT0 **NN**	We've seen such myths in our time (33) have tragic consequences – (26)
Sub 111	What is less well known is that prominent Nazis believed that the master race originated in Atlantis.	What is less well-known is <u>that</u> (31) prominent Nazis believe <u>that the</u> (32) <u>master race</u> originated in Atlantis. (35)	CJT CJT	Can't suggest Can't suggest
Sub 112	These claims to an ancestral heritage in Atlantis	These claims to (15) an ancestral heritage in Atlantis (33)	**PRP**/AT0 **NN**	These claims (12) to an ancestral heritage in Atlantis (36)
Sub 113	fed the Nazis belief in the supremacy of the Aryan master race.	fed the Nazis' belief <u>in the</u> (28) <u>supremacy</u> of the Aryan master race. (35)	**PRP** AT0/**NN**	fed the Nazis' belief (21) in the supremacy of the Aryan race. (35) fed the Nazis' belief in the supremacy (36) of the Aryan master race. (25)
Sub 114	When we come to something like the lost continent of Atlantis	When we come to something <u>like</u> (30) <u>the</u> lost <u>continent</u> of Atlantis, (31)	**PRP**/AT0 **NN**	When we come to something (25) like the lost continent of Atlantis, (36)
Sub 118	so I'm not going to tell you that belief in Atlantis is necessarily the first step towards genocide, or Holocaust,	I'm not going to say that belief (32) in Atlantis is necessarily <u>the first</u> (35) <u>step</u> towards genocide or holocaust. (35)	**AT0** AJ0/**NN**	I'm not saying that belief (26) in Atlantis is the first step (29) towards genocide or holocaust. (30)
Sub 119	but what I'm telling you is we are on a very slippery slope if we believe in fantasies	But we are <u>on a very</u> slippery (29) <u>slope</u> if we believe in fantasies. (33)	**PRP** AT0 AV0 AJ0/**NN**	But we are on a very slippery slope (35) if we believe in fantasies. (27)

Appendix 9: Segmentation in *Chimps on Death Row*

No.	Transcript	Subtitle	Type of line-break error	Alternative
Sub 1	The chimpanzees at the Holloman Air Force Base have been used in research for 40 years.	The chimpanzees of the <u>Holloman</u> (31) <u>Air Force Base</u> have been used (29) in research for 40 years. (25)	NP0	Can't suggest
Sub 3	There's no question that the use of chimpanzees presents a greater moral dilemma	The use of chimpanzees <u>presents</u> (31) <u>a greater moral dilemma</u> (23)	Verb/object	The use of chimpanzees (22) presents a greater moral dilemma (32)
Sub 8	and with the same shame that I look back on my great-grandfather who was a slaver holder	with the shame that I feel <u>for my</u> (33) <u>great-grandfather</u> as a slaveholder. (35)	**PRP** DPS/**NN**	Can't suggest
Sub 13	They have brains and central nervous systems more like ours than that of any other living creature.	They have brains and <u>central</u> (28) <u>nervous systems</u> similar to ours. (32)	AJ0/AJ0 NN	Can't suggest
Sub 14	They show characteristics like altruism, compassion and so forth	They show characteristics <u>like</u> (30) <u>altruism</u>, compassion and so forth. (34)	PRP/NN	They show characteristics (25) like altruism and compassion. (29)
Sub 21	You're only allowed to use chimpanzees to study diseases	We're only allowed <u>to use</u> (25) <u>chimpanzees</u> to study diseases (29)	Verb/Object	We're only allowed (18) to use chimpanzees to study diseases (36)
Sub 25	For the chimps these human diseases are rarely fatal	For the chimps, <u>these human</u> (27) <u>diseases</u> are rarely fatal, (26)	**DT0** AJ0/**NN**	For the chimps, these human diseases (36) are rarely fatal, (17)

Sub 27	In this Institute we have the policy not to engage in experiments	In this institute, we have <u>the</u> (30) <u>policy</u> not to engage in experiments (35)	AT0/NN	In our institute, we have the policy (36) not to engage in experiments (28)
Sub 42	There was a surplus developing of animals that had been exposed to hepatitis B	There was a surplus of animals <u>that</u> (35) <u>had been exposed to hepatitis B</u>, (32)	CJT	Tentative suggestion with major editing A surplus of animals (20) had been exposed to hepatitis B, (32)
Sub 45	The surplus raised an awful question: was it acceptable to kill the chimps they didn't need?	The surplus raised <u>an awful</u> (27) <u>question</u> - was it acceptable <u>to</u> (31) <u>kill</u> the chimps they didn't need? (33)	**AT0** AJ0/**NN** TO0/VVI	The surplus raised an awful question (36) - was it acceptable to kill (27) the chimps they didn't need? (28) The surplus raised an awful question (36) - was it acceptable (19) to kill the chimps they didn't need? (36)
Sub 58	to compensate for all the animals they thought would soon be dying of AIDS.	to compensate for the animals <u>they</u> (34) <u>thought would soon be dying of AIDS</u>. (36)	Subject/ Predicate	to compensate for the animals (29) they thought would soon die of AIDS. (36)
Sub 59	When we first started putting the material in the animal we were continuously looking for disease.	When we first put material (26) in the animal, we <u>were</u> continuously (35) <u>looking</u> for disease. (20)	VBD/VVG	When we first put (17) the material in the animal, (27) we kept looking for disease. (28)
Sub 61	and what we had to consider was the fact that it may lay dormant for years before it actually caused infection in the chimp,	We had to consider that it <u>may</u> (30) <u>lay</u> dormant for years before (28) causing infection in the chimp. (31)	VM0/VVI	Can't suggest
Sub 62	so probably for 5 or 6 years people were just thinking well, we just haven't had enough time,	For five or six years, <u>people</u> just (34) <u>thought we hadn't had enough time</u>. (34)	Subject/ Predicate	For six years, people just thought (34) we hadn't had enough time. (26)

Sub 63	but once we reached that time there was the opinion that more than likely this isn't going to be a disease model.	Once we reached that time, (26) the opinion was that (20) this wouldn't be a disease model. (33)	CJT	Once we reached that time, (26) the opinion was (15) that this wasn't a disease model. (33)
Sub 64	There are now 200 HIV-positive chimps around the world	There are now 200 <u>HIV-positive</u> (30) <u>chimps</u> around the world (23)	AJ0/NN	There are 200 HIV-positive chimps (33) around the world (16)
Sub 66	They are still a health hazard to humans and have to be housed separately from other, clean chimps.	They are a health hazard and <u>are</u> (32) <u>housed</u> separately from other chimps. (36)	VBB/VVN	Can't suggest
Sub 67	The Biomedical Primate Research Centre in Holland	The <u>Biomedical Primate</u> (22) <u>Research Centre</u> in Holland (26)	NP0	Can't suggest
Sub 81	Bids to support the chimps came from groups wanting to retire them, and also from the Coulston Research Foundation.	Bids to help the chimps came <u>from</u> (33) <u>groups</u> wanting to retire them <u>and</u> (33) <u>the Coulston Research Foundation</u>. (33)	PRP/NN CJC	Bids to help the chimps came (28) from groups wanting to retire them (34) and the Coulston Foundation. (28)
Sub 87	To just take these valuable animals and throw 'em out in a field in, in South Texas.	To take these valuable animals <u>and</u> (34) <u>throw them out in a field in Texas?!</u> (36)	CJC	To take these valuable animals (30) and throw them in a field in Texas?! (36)
Sub 88	Dr. Coulston wants to use the ageing survivors of the space programme and their descendants	Dr Coulston wants to use <u>the ageing</u> (35) <u>survivors</u> of the space programme (32) and their descendants (21)	**AT0** AJ0/**NN**	Can't suggest
Sub 89	to study prostate cancer, diabetes and any other human diseases that chimps might develop in their old age.	to study prostate cancer, diabetes (34) and any other human diseases <u>that</u> (33) <u>chimps may develop in their old age</u>. (36)	CJT	to study prostate cancer, diabetes (34) and any other human diseases (24) that chimps may develop in old age. (35)

Sub 91	that have value throughout their life for the medical understanding of diseases of old age.	that have value <u>throughout</u> <u>their</u> (32) <u>life</u> for the medical understanding (34) of diseases of old age. (23)	**PRP** DPS/**NN**	Can't suggest
Sub 97	This is a typical chimpanzee response to strangers coming into his territory	This is <u>a</u> typical <u>male</u> <u>chimpanzee</u> (33) <u>response</u> to strangers (21) coming into his territory. (26)	**DT0** AJ0 (NN-AJ0)/**NN**	Tentative suggestion with major editing: This is a typical male response (31) to strangers entering his territory. (36)
Sub 106	His life at the way station depends on donations from the public.	His life at the Waystation <u>depends</u> (34) <u>on</u> donations from the public. (29)	VVZ/PRP (phrasal verb)	His life at the Waystation (26) depends on donations from the public. (28)
Sub 117	Since chimpanzees are our closest living relatives, since in their own right they're wondrous and amazing beings,	Since chimpanzees are <u>our</u> <u>closest</u> (33) <u>living relatives</u>, since they're (31) wondrous and amazing beings, (28)	**DPS** AJS/AJ0 **NN**	Can't suggest

Appendix 10: Segmentation in *The World's Most Wanted*

No.	Transcript	Subtitle	Type of line-break error	Alternative
Sub 7	There is no question he is what we would call a prime suspect.	There is no question he's <u>what</u> (30) <u>we would call a prime suspect</u>. (30)	PNQ	There is no question (20) he's what we call a prime suspect. (34)
Sub 8	The prime suspect is Osama Bin Laden whose murderous campaign against America	The prime suspect is <u>Osama Bin</u> (30) <u>Laden</u>. His campaign against America (35)	NP0	The prime suspect is Osama Bin Laden. (37) His campaign against America (28)
Sub 10	And if he thinks he can hide and run from the United States or allies, he will be sorely mistaken.	If he thinks he can hide <u>from</u> (29) <u>the USA</u> or allies, he is mistaken. (34)	**PRP**/AT0 **NP0**	If he thinks he can hide (24) from the USA or allies, he is wrong. (36)
Sub 21	It's a big irony that in his personal level he's very soft, very humble, very smooth.	It's ironic that <u>in his personal</u> (32) <u>life</u> he is very soft, humble, smooth. (37)	**PRP** DPS AJ0/**NN**	It's ironic that in his personal life (37) he is very soft, humble, smooth. (32)
Sub 22	He speaks little. Osama Bin Laden's beginnings were far from humble.	He speaks little. <u>Osama Bin Laden's</u> (35) <u>beginnings</u> were far from humble. (32)	NN POS/NN	Osama Bin Laden's beginnings (28) were far from humble. (21)
Sub 29	who does not share Bin Laden's views on using violence.	who does not share Bin <u>Laden's</u> (30) <u>views</u> on using violence. (24)	NN POS/NN	who does not share (18) Bin Laden's views on using violence. (36) who does not share Bin Laden's views(36) on using violence. (18)
Sub 38	Bin Laden was one of thousands of young Saudis	Osama Bin Laden <u>was</u> (19) <u>one of thousands of young Saudis</u> (32)	Verb/ Complement	Osama Bin Laden(15) was one of thousands of young Saudis (36)

Sub 39	who flocked to the mountains to join the cause despite the perils.	who flocked to the mountains to (31) join the cause, despite the perils. (35)	TO0/VVI	who flocked to the mountains to join (33) the cause, despite the perils. (30)
Sub 43	He was both building up his own militant organisation - Al Quaeda -	He was both building up his (27) militant organisation, al-Qaida, (32)	**DPS**/AJ0 **NN**	He was both building up (23) his militant organisation, al-Qaida, (36)
Sub 52	When you look at Osama Bin Laden and his network, his financial network, it's rather like peeling an onion skin by skin.	When you look at his financial (30) network, it is like peeling an onion. (37)	**PRP** DPS AJ0/**NN**	When you look at his financial network, (37) it is like peeling an onion. (28)
Sub 60	I don't think, to my knowledge, that he's actually ordered a group to go out and carry out a specific operation.	To my knowledge, I don't think that (35) he's ordered a specific operation. (34)	CJT	I don't think that he's ordered a specific operation.
Sub 61	What he does is sort of select a group and then give it some financing and then give it its head as it were.	He selects a group, then finances (33) it and gives it its head, as it were. (37)	Verb/Object	He selects a group, then finances it (36) and gives it its head, as it were. (34)
Sub 66	In the Islamic world doors were closing against Bin Laden.	In the Islamic world, doors (27) were closing against Bin Laden. (31)	Subject/ Predicate	In the Islamic world, (21) doors were closing against Bin Laden. (37)
Sub 68	Washington, Presidential advisers considered but rejected	In Washington, presidential (27) advisers considered, but rejected, (34)	AJ0/NN	Tentative suggestion involving major editing with subtitle C69 In Washington, presidential advisers(36) considered, but rejected a raid (31) to capture Bin Laden (21)

Sub 71	Why was the decision taken in the end not to go into Afghanistan and take him out?	Why was the decision taken not <u>to</u> (33) <u>go</u> into Afghanistan and get him? (32)	TO0/VVI	Why was the decision taken not to go (36) into Afghanistan and get him? (29)
Sub 72	I believe they thought there would National Executive too much collateral damage, too many non-involved casualties	I believe they thought there <u>would</u> (34) <u>be</u> too many none-involved casualties (36)	VM0/VBI	I believe they thought there would be (37) too many none-involved casualties (33)
Sub 75	So the decision was made to leave him there for the time being. Leave him there for the time being.	So, the decision was made to <u>leave</u> (34) <u>him</u> there? Yes, for the time being. (35)	Verb/Object + Question and Answer on the same line	So, they decided to leave him there? (36) - Yes, for the time being. (25)
Sub 76	From Afghanistan Bin Laden continued to taunt the American President himself, Bill Clinton, with his propaganda material.	Bin Laden continued to taunt <u>the</u> (32) <u>American President</u>, Bill Clinton, (33)	**AT0/** (AJ0) **NN**	Bin Laden continued to taunt (28) the American President, Bill Clinton, (37)
Sub 83	they may be even rich, prosperous people, they may be very highly educated,	may be even <u>rich</u> or <u>prosperous</u> (30) <u>people</u> and maybe highly educated. (33)	AJ0/NN	may be even rich or prosperous (30) and maybe highly educated. (26)
Sub 87	As dawn broke over the shattered skyline of New York on the morning of September 12th	As dawn broke <u>over</u> <u>the</u> <u>shattered</u> (32) <u>skyline</u> of New York on 12 September, (36)	**PRP** AT0 AJ0/**NN**	Can't suggest or Tentative suggestion with deletion: As dawn broke over the skyline (30) of New York on 12 September, (28)

Sub 89	they had a concrete indication as to who was ultimately responsible for masterminding the devastating attack on America.	they had a concrete indication as (33) to who was ultimately responsible (33)	PRP21/PRP22	they had a concrete indication (30) as to who was ultimately responsible (36)
Sub 99	His authorities had established further indications of Bin Laden's guilt.	Further implications of Bin Laden's (35) guilt had been found. (21)	**PRP** NN POS/**NN**	Further implications (20) of Bin Laden's guilt had been found. (36)
Sub 117	Now if they intend to destroy Bin Laden, Afghanistan, they would have done nothing	Now if they intend to destroy Bin (33) Laden they will have done nothing. (34)	NP0	Now if they plan to destroy Bin Laden (37) they will have done nothing. (28)

Bibliography

Abney, Steven P. (1991). "Parsing by chunks". In Berwick, Robert C., Steven P. Abney & Carol. Tenny (eds.) *Principle-Based Parsing: Computation and Psycholinguistics.* Dordrecht, London: Kluwer Academic, pp. 257–278.

Araujo, Vera Lucia Santiago (2004). "Closed subtitling in Brazil". In Orero, Pilar (ed.), *Topics in audiovisual translation*. Amsterdam: John Benjamins, pp. 199-213.

Audio Description Associates LLC (no date). *The visual made verbal.* Takoma Park, MD.

Baker, Mona (1995). "Corpora in Translation Studies: An Overview and Some Suggestions for Future Research". *Target* **7(2)**, pp. 223-243.

Baker, Robert G. (1981). *Guidelines for the Subtitling of Television Programmes for the Deaf and Hard of Hearing.* Independent Broadcasting Authority, Southampton University and Independent Television Companies Association.

Baker, Robert, Andrew Lambourne, and Guy Rowston (1982). *Handbook for Television Subtitlers*. Crawley Court, Winchester, Hampshire: Independent Broadcasting Authority, Southampton University and ITV.

Bartoll, Eduard (2004). "Parameters for the classification of subtitles". In Orero, Pilar (ed.) *Topics in Audiovisual Translation*. Amsterdam, Philadelphia: John Benjamins, pp. 53-60.

BBC (1998). *BBC Subtitling Guide*. London: BBC.

BCI (2005). *BCI Guidelines – Subtitling*. Broadcasting Commission of Ireland. Available on-line at http://www.bci.ie/codes/access_codes.html (last accessed 28 October 2008)

Bennett, Paul (1995). *A Course in Generalized Phrase Structure Grammar*. London and New York: Routledge.

Biber, Douglas (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Bloor, Tom and A. Meriel Bloor (2004). *The functional analysis of English*. 2nd edition. London: Arnold.

Bobeldijk, Marcel (2004). "Technology from a deaf person's perspective". Speech held at the occasion of the European Conference "Access to the Information Society" in the Residence Palace, Brussels, on 25 February 2004. Available on-line http://www.ictrnid.org.uk/euc/txtmb.html (last accessed 28 October 2008).

Brondeel, Herman (1994). "Teaching Subtitling Routines". *Meta*, **39.1**, pp. 26-33.

Daelemans, Walter, Anja Hothker and Erik Tjong Kim Sang (2004). "Automatic Sentence Simplification for Subtitling in Dutch and English". *In Proceedings of the 4th International Language Resources and Evaluation Conference (LREC 2004), Lisbon, Portugal*, pp. 1045-1048.

Davis, Hallowel and S. Richard Silverman (1978). *Hearing and deafness*. 4th edition. New York, London : Holt, Rinehart and Winston.

Deafness Forum of Australia (2004). *Captioning Quality Code of Practice*. Available on-line http://www.deafnessforum.org.au/word/Captioning%20Quality%20Code%20 of%20Practice.DOC (last accessed 28 October 2008)

De Beaugrande, Robert-Alain and Wolfgang Ulrich Dressler (1981). *Introduction to text linguistics*. London and New York: Longman.

Delabastita, Dirk (1989). "Translation and mass-communication". *Babel* **35:4**, pp. 193-218.

De Linde, Zoe (1995). "Read my lips: Subtitling principles, practices, and problems". *Perspectives: Studies in Translatology* **3.1**, pp. 9-20.

De Linde, Zoe (1997). *Linguistic and Visual Complexity of Television Subtitles*. Bristol: University of Bristol. Unpublished PhD thesis.

De Linde, Zoe and Neil Kay (1999). *The Semiotics of Subtitling*. Manchester: St. Jerome.

Den Boer, Corien (2001). "Live interlingual subtitling". In Gambier, Yves and Henrik Gottlieb (eds.) *(Multi)Media Translation: Concepts, Practices, and Research*. Amsterdam, Philadelphia: John Benjamins, pp. 167-172.

Described and Captioned Media Program (2007). "Captioning Key – Guidelines and Preferred Techniques". Available on-line at
http://www.dcmp.org/caai/nadh7.pdf (last accessed on April 25th 2008)

Díaz Cintas, Jorge (1999). "Dubbing or subtitling: the eternal dilemma". *Perspectives: Studies in Translatology* **7.1**, pp. 31-40.

Díaz Cintas, Jorge (2001). *La traducción audiovisual: El subtitulado*. Salamanca, España: Ediciones Almar.

Diaz Cintas, Jorge (2001b). "Teaching subtitling at university" In Cunico, Sonia (ed.) *Training Translators and Interpreters in the New Millennium*. Portsmouth: University of Portsmouth, pp. 29-44.

Díaz Cintas, Jorge (2003) "Audiovisual Translation in the Third Millennium". In Anderman, Gunilla and Margaret Rogers (eds.) *Translation today: trends and perspectives*. Clevedon: Multilingual Matters, pp. 192-204.

Díaz Cintas, Jorge (2003b). *Teoría y práctica de la subtitulación: inglés/español.* Barcelona: Ariel.

Díaz Cintas, Jorge (2004). "Subtitling: the long journey to academic acknowledgment". *The Journal of Specialised Translation*, **Issue 1**, pp. 50-70. On-line journal http://www.jostrans.org/issue01/art_diaz_cintas.php (last accessed 28 October 2008)

Díaz Cintas, Jorge and Aline Remael (2007). *Audiovisual Translation: Subtitling.* Manchester: St Jerome.

Erard, Michael (2001). "The King of Closed Captions. The self-appointed watchdog of a growing industry". *The Atlantic Online*. Available at http://www.theatlantic.com/doc/200109/erard (last accessed 28 October 2008)

Eugeni, Carlo (forthcoming). "Respeaking political debate for the deaf: the Italian case". Baldry, Anthony and Elena Montagna (eds.) *Interdisciplinary perspectives on multimodality: theory and practice. Proceedings of the third international conference on multimodality.* Campobasso: Palladino.

Eugeni, Carlo (forthcoming b). "Respeaking BBC news, a strategic approach". *The Sign Language Translator and Interpreter (SLTI)*, **3:1**.

Franco, Eliana P.C. and Vera Lucia Santiago Araujo (2003). "Reading Television: Checking Deaf People's Reactions to Closed Subtitling in Fortaleza, Brazil". *The Translator* **9.2**, pp. 249-267.

Gambier, Yves (1994). "Audio-visual communication: typological detour". In Dollerup, Cay and Anette Lindegaard (eds.) *Teaching Translation and Interpreting 2: Insights, Aims, Visions.* Amsterdam and Philadelphia: John Benjamins, pp. 275-283.

Gambier, Yves (2003). "Introduction: Screen Transadaptation: Perception and Reception." *The Translator* **9:2**, pp. 171-190.

Gambier, Yves (2004). "Tradaptation cinématographique". In Orero, Pilar (ed.) *Topics in Audiovisual Translation*. Amsterdam, Philadelphia: John Benjamins, pp. 169-181.

Garside, Roger, Geoffrey Leech and Anthony McEnery. (1997). *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London and New York: Longman.

Georgakopoulou, Panayota (2003). *Reduction Levels in Subtitling – DVD Subtitling: A Compromise of Trends*. Guildford: University of Surrey. Unpublished PhD thesis.

Gottlieb, Henrik (1992). "Subtitling – A new university discipline". In Dollerup, Cay and Anne Loddegaard (eds.) *Teaching Translation and Interpreting: Training, Talent and Experience*. Amsterdam and Philadelphia: John Benjamins. pp. 161-170.

Gottlieb, Henrik (1994). "Subtitling: people translating people". In Dollerup, Cay and Anette Lindegaard (eds.) *Teaching Translation and Interpreting 2: Insights, Aims, Visions*. Amsterdam and Philadelphia: John Benjamins, pp. 261-274.

Gottlieb, Henrik (1994b). "Subtitling: Diagonal Translation". *Perspectives: Studies in Translatology* **2:1**, pp. 101-121.

Gottlieb, Henrik (1998). "Subtitling". In Baker, Mona (ed.) *Encyclopedia of Translation Studies.* London and New York: Routledge, pp. 244-248.

Gottlieb, Henrik (2005). "Texts, translation and subtitling – in theory and in Denmark". In Gottlieb, Henrik *Screen Translation: Eight studies in subtitling, dubbing and voice over.* Copenhagen: Center for Translation Studies, Department of English, University of Copenhagen, pp. 1-40.

Gottlieb, Henrik (2005b). "Subtitling Against the Current: Danish Concepts, English Minds". In Gottlieb, Henrik *Screen Translation: Eight studies in subtitling, dubbing and voice over.* Copenhagen: Center for Translation Studies, Department of English, University of Copenhagen, pp. 195-219.

Greenbaum, Sidney (1996). *The Oxford English Grammar*. Oxford: Oxford University Press.

Gutt, Ernst-August (2000). *Translation and relevance: cognition and context*. Oxford: Blackwell.

Halliday, M.A.K. (1985). *An introduction to functional grammar*. London: Edward Arnold.

Halliday, M.A.K. (1989). *Spoken and written language.* 2nd edition. Oxford: Oxford University Press.

Halliday, M.A.K. (1994). *An introduction to functional grammar*. 2nd edition. London, New York, Sydney, Auckland: Edward Arnold.

Halliday, M.A.K. and Ruqaiya Hasan (1976). *Cohesion in English.* London: Longman.

Heiss, Christine and Marcello Soffritti (2008). "Forlixt 1 – The Forlì Corpus of Screen Translation: Exploring microstructures." In Chiaro, Delia, Christine Heiss and Chiara Bucaria (eds.) *Between Text and Image: Updating research in screen translation*. Amsterdam/Philadelphia: John Benjamins, pp. 51-62.

House, Juliane (1977). *A model for translation quality assessment*. Tübingen : TBL Verlag Gunter Narr.

IFF Research Ltd and Jane Sancho-Aldridge (1997). *Good news for deaf people: subtitling of national news programmes*. London: Independent Television Commission.

ITC (1997). *ITC Guidance on Standards for Subtitling*. London.

ITC (1999). *ITC Guidance on Standards for Subtitling*. London.

Ivarsson, Jan and Mary Carroll (1998). *Subtitling*. Simrishamn: TransEdit.

Ivarsson, Jan and Mary Carroll (1998b). *Code of Good Subtitling Practice*. Available on-line at http://www.esist.org/Code.pdf (last accessed 28 October 2008)

Jakobson, Roman (1959). "On linguistic aspects of translation". In Venuti, Lawrence (ed.) (2000) *The Translation Studies Reader*. London: Routledge, pp. 113-118.

James, Heulwen (1998). "Screen Translation Training and European Co-operation". In Gambier, Yves (ed.) *Translating for the Media, Papers from the International Conference "Languages and the Media", Berlin, November 22-23, 1996*. Centre for Translation and Interpreting, University of Turku, pp. 243-258.

Jaskanen, Susanna (1999). *On the inside track to Loserville, USA: strategies used in translating humour in two Finnish versions of "Reality Bites"*. Helsinki: University of Helsinki, Faculty of Arts, Department of English. Unpublished MA thesis. Available on-line at
http://ethesis.helsinki.fi/julkaisut/hum/engla/pg/jaskanen/index.html
(last accessed 28 October 2008)

Jensema, Carl and Ralph McCann (1995). *Presentation Speed and Vocabulary in Closed Captioned Television*. Institute for Disabilities Research and Training.

Jensema, Carl, Ralph McCann and Scott Ramsey (1996). "Closed-captioned television presentation and vocabulary. *American Annals of the Deaf* **141:4**, pp. 284-292.

Jensema, Carl (1998). "Viewer reaction to different television captioning speeds". *American Annals of the Deaf* **143:4**, pp.318-324.

Jurafsky, Daniel and James H. Martin (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. New Jersey: Prentice Hall.

Karamitroglou, Fotios (1998). "A Proposed Set of Subtitling Standards in Europe". *Translation Journal* **2:2**. On-line journal
http://accurapid.com/journal/04stndrd.htm (last accessed 28 October 2008)

Karamitroglou, Fotios (2001). "Ο Υπολογιστής κι ο Μάγειρας ή μην Πυροβολείτε Τον Υποτιτλιστή (The Subtitler and the Cook or don't Shoot the Subtitler)". *Translatum Journal* **1**. On-line journal
http://www.translatum.gr/journal/1/karam.htm (last accessed 28 October 2008)

Kennedy, Graeme (1998). *An introduction to corpus linguistics*. London: Longman.

Kenny, Dorothy (1998). "Corpora in translation studies". In Baker, Mona (ed.) *Routledge Encyclopedia of Translation Studies*. London: Routledge, pp. 50-53.

Koljonen, Tiina (1995). "TV-översättning - konsten att komprimera". In Saari, Mirja and Jan Lindström (eds.) *Svenskan i Finland 3: Föredrag vid tredje sammankomsten för beskrivningen av svenskan i Finland, Helsingfors 14-15 Oktober 1994*. Helsinki: Helsingfors, pp. 179-192.

Kovačič, Irena (1992). *Jezikoslovni pogled na podnaslovno prevajanje televizijskih oddaj (Linguistic Aspects of Subtitling Television Programmes).* Faculty of Philosophy, University of Ljubljana. Unpublished PhD thesis.

Kovačič, Irena (1994) "Relevance as a factor in subtitling reductions". In Dollerup, Cay and Anette Lindegaard (eds.) Teaching Translation *and Interpreting 2: Insights, Aims, Visions*. Amsterdam and Philadelphia: John Benjamins, pp. 245-251.

Kovačič, Irena (1998). "Six Subtitlers - Six Subtitling Texts". In Bowker, Lynne, Michael Cronin, Dorothy Kenny and Jennifer Pearson (eds.) *Unity in Diversity? Current Trends in Translation Studies*. Manchester: St. Jerome, pp. 75-82.

Kovačič, Irena (1998b). "Language in the media – A new challenge for translator trainers". In Gambier, Yves (ed.) *Translating for the media*. Turku: University of Turku, Centre for Translation and Interpreting, pp. 123-129.

Kyle, Jim (1992). *Switched On: Deaf People's Views on Television Subtitling; A Report Prepared for the ITC and BBC by the Centre for Deaf Studies, University of Bristol*. Bristol: Centre for Deaf Studies, University of Bristol.

Laviosa, Sara (2002). *Corpus-based translation studies: theory, findings, applications*. Amsterdam, New York: Rodopi.

Leech, Geoffrey (1991). "The state of the art in corpus linguistics". In Aijmer, Karin and Bengt Altenberg (eds.) *English corpus linguistics: studies in honour of Jan Svartvik*. London: Longman.

Leech, Geoffrey (1997). "Grammatical tagging". In Garside, Roger, Geoffrey Leech and Anthony McEnery (ed.) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London and New York: Longman, pp. 19-33.

Lomheim, Sylfest (1995). "L'écriture sur l'écran, stratégies de sous-titrage à NRK; une étude de cas". In Gambier, Yves (ed.) *Communication audiovisuelle et transferts linguistiques. Audiovisual Communication and Language Transfer. Proceedings from the International Forum in Strasbourg, 22-24.6.1995*. Special issue of *Translatio – Nouvelles de la FIT – FIT Newsletter* **14:3-4**, pp. 288-293.

Lomheim, Sylfest (1999). "The Writing on the Screen. Subtitling: A Case Study from Norwegian Broadcasting (NRK), Oslo". In Anderman, Gunilla and Margaret Rogers (eds.) *Word, Text, Translation. Liber Amicorum for Peter Newmark.* Clevedon: Multilingual Matters, pp. 190-207.

Luyken, Georg-Michael, Thomas Herbst, Jo Langham-Brown, Helen Reid and Hermann Spinhof (1991). *Overcoming language barriers in television: Dubbing and subtitling for the European audience.* Manchester: European Institute for the Media.

Melero, Maite, Antoni Oliver and Toni Badia (2006). "Automatic Multilingual Subtitling in the eTITLE project". In *Proceedings of ASLIB Translating and the Computer 28*. Available on-line at http://mutis.upf.es/~mmelero/publicacions/aslib2006.pdf (last accessed 28 October 2008)

Marleau, Lucien (1982). "Les sous-titres… Un mal necessaire". *Meta* **27.3**, pp. 271-285.

Marschark, Marc (1997). *Raising and Educating a Deaf Child*. New York and Oxford: Oxford University Press.

McEnery, Tony and Andrew Wilson (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Moores, Donald F. (1987). *Educating the deaf: psychology, principles and practices*. 3rd edition. Boston: Houghton Mifflin.

Motulsky, Harvey (1999). *Analyzing Data with GraphPad Prism: A companion to GraphPad Prism version 3*. San Diego CA: GraphPad Software Inc. Available on-line at http://www.graphpad.com/articles/AnalyzingData.pdf (last accessed 28 October 2008)

Nedergaard-Larsen, Birgit (1993). "Culture-bound problems in subtitling". *Perspectives: Studies in Translatology* **1.2**, pp. 207-241.

Neves, Joselia (2004). *vKv Subtitling System / Sistema de Legendagem Vozes K se Vêem*. In Neves, Joselia (2005). *Audiovisual Translation: Subtitling for the Deaf and Hard-of-Hearing*. London: University of Surrey Roehampton. Unpublished PhD thesis.

Neves, Joselia (2005). *Audiovisual Translation: Subtitling for the Deaf and Hard-of-Hearing*. London: University of Surrey Roehampton. Unpublished PhD thesis.

Neves, Joselia (2007). " A world of change in a changing world". In Díaz Cintas, Jorge, Pilar Orero and Aline Remael (eds.) *Media for All: Subtitling for the Deaf, Audio Description, and Sign Language*. Amsterdam: Rodopi, pp. 89-98.

Neves, Joselia (2007b). "Of Pride and Prejudice". *The Sign Language Translator and Interpreter (SLTI)*, **1:2**, pp. 251-274.

Newmark, Peter (1988). *A textbook of translation*. Hemel Hempstead: Prentice-Hall International.

NSM Research (2001). *Subtitling on cable, satellite and digital terrestrial television channels. Summary report prepared for RNID*. Oxford: NSM Research.

Nyberg, Eric and Teruko Mitamura (1997). "A Real-Time MT System for Translating Broadcast Captions". *MT Summit VI, Machine Translation Past Present Futute, San Diego, CA, October 29 - November 1, 1997*, pp. 51-57. Available on-line at: http://www.lti.cs.cmu.edu/Research/Kant/PDF/mts6.pdf (last accessed 28 October 2008)

O'Connell, Eithne (1998). "Choices and Constraints in Screen Translation". In Bowker, Lynne, Michael Cronin, Dorothy Kenny and Jennifer Pearson (eds.) *Unity in Diversity? Current Trends in Translation Studies*. Manchester: St. Jerome, pp. 65-71.

Ofcom (2005). *Subtitling – An issue of speed*. London: Office of Communications.

Ofcom (2008). *Code on Television Access Services*. London: Office of Communications. Available on-line at http://www.ofcom.org.uk/tv/ifi/codes/ctas/ctas.pdf (last accessed 28 October 2008)

Ofcom (2008b). *Television channels required to provide television access services in 2009.* London: Office of Communications. Available on-line at http://www.ofcom.org.uk/tv/ifi/guidance/tv_access_serv/tv_accesss_statemen t09/tvaccessservices09.pdf (last accessed 28 October 2008)

Olohan, Maeve (2004). *Introducing corpora in translation studies*. London and New York: Routledge.

Orero, Pilar (2004). *Topics in Audiovisual Translation*. Amsterdam, Philadelphia: John Benjamins.

Paul, Peter V. and Stephen P. Quigley (1990). *Education and Deafness*. New York: Longman.

Paul, Peter V. and Stephen P. Quigley (1994). *Language and Deafness*. 2nd edition. San Diego: Singular Publishing Group.

Pedersen, Jan (2003). *A corpus-linguistic investigation into quantitative and qualitative reduction in subtitles*. Örebro University, Sweden. Unpublished background study.

Pedersen, Jan (2007). *Scandinavian Subtitles: A comparative study of subtitling norms in Sweden and Denmark with a focus on Extralinguistic Cultural References.* Stockholm: Stockholm University. Unpublished PhD thesis. Abstract available on-line at http://www.diva-portal.org/su/theses/spikblad.xsql?dbid=7159 (last accessed 28 October 2008).

Perego, Elisa (2008). "What Would We Read Best? Hypotheses and Suggestions for the Location of Line Breaks in Films Subtitles". *The Sign Language Translator and Interpreter* **2.1**, pp. 35-63.

Perego, Elisa and Fabio Del Missier (2008). "Is a reading situation better than another for subtitled film viewers?" In *Handbook of Abstracts, Audiovisual Translation: Multidisciplinary Approaches (La traduction audiovisuelle: Approches pluridisciplinaires), 19-21 June 2008, Montpellier, France,* p. 34.

Piperidis, Stelios, Iason Demiros, Prokopis Prokopidis, Peter Vanroose, Anja Hoethker, Walter Dealemans, Elsa Sklavonou, Manos Konstantinou and Yannis Karavidas (2004). "Multimodal Multilingual Resources in the Subtitling Process". In *Proceedings of the 4th International Language Resources and Evaluation Conference (LREC 2004), Lisbon, Portugal,* pp. 205-208.

Popowich, Fred, Paul McFetridge, Davide Turcato and Janine Toole (2001). "Machine Translation of Closed Captions". *Machine Translation* **15**, pp. 311-341.

Pym, Anthony (1999). "Statistics for Novices: An introduction for myself and for a few students contemplating research." Unpublished notes. Available on-line at http://www.tinet.org/~apym/on-line/research_methods/stats/stats.html (last accessed 28 October 2008).

Remael, Aline (p. 2004). "A place for film dialogue analysis in subtitling courses". In Orero, Pilar (ed.), *Topics in audiovisual translation*. Amsterdam: John Benjamins, pp. 21-34.

RNID (1999). *Subtitling for deaf and hard of hearing people: Report for the Royal National Institute for Deaf People*. London: RNID.

Ross, Mark, Diane Brackett and Antonia Maxon (1982). Hard of hearing children in regular schools. New Jersey: Prentice Hall.

Shuttleworth, Mark and Moira Cowie (1997). *Dictionary of Translation Studies*. Manchester: St. Jerome.

Smith, Stephen (1998). "The language of subtitling". In Gambier, Yves (ed.) *Translating for the media*. Turku: University of Turku, Centre for Translation and Interpreting, pp. 139-149.

Softel (2001). *FAQ - Live or 'Real-time' Subtitling for TV.* Available on-line at http://www.softel-usa.com/downloads/Softel_Live_Subtitling_FAQ.pdf (last accessed 28 October 2008)

Sperber, Dan and Deirdre Wilson (1986). *Relevance: communication and cognition.* Oxford: Blackwell.

Titford, Christopher (1982). "Sub-Titling – Constrained Translation". In *Lebende Sprachen* **27:3**, pp. 113–116.

Tjong Kim Sang, Erik, Walter Daelemans and Anja Hothker (2004). "Reduction of Dutch Sentences for Automatic Subtitling". In *Proceedings of CLIN-2003, University of Antwerp, Antwerp Papers in Linguistics, 111*, pp. 109-123.

Tomaszkiewicz, Teresa (1998). "L'économie discursive dans le cinéma sous-titré". *In Ruffino Giovanni (ed.) Atti del XXI Congresso internazionale di linguistica e filologia romanza: Centro di studi filologici e linguistici siciliani, Universita di Palermo, 18-24 settembre 1995*. Tubingen: M. Niemeyer, pp. 281-289.

Toole, Janine, Davide Turcato, Fred Popowich, Dan Fass and Paul McFetridge (1998). "Time-Constrained Machine Translation". In Farwell, David, Laurie Gerber and Edward Hovy (eds.) *Machine Translation and the Information Soup. Third Conference of the Association for Machine Translation in the Americas AMTA' 98, Langhorne, PA, USA, October 28-31, 1998*. Berlin: Springer, pp.103-112.

Valentini, Christina (2006). "A Multimedia Database for the Training of Audiovisual Translators". *The Journal of Specialised Translation*, **Issue 6**, pp. 68-84. On-line journal http://www.jostrans.org/issue06/art_valentini.php (last accessed 28 October 2008)

Valentini, Christina (2008). "Forlixt 1 – The Forlì Corpus of Screen Translation: Exploring macrostructures." In Chiaro, Delia, Christine Heiss and Chiara Bucaria (eds.) *Between Text and Image: Updating research in screen translation*. Amsterdam/Philadelphia: John Benjamins, pp. 37-50.

Watson, Linda, Susan Gregory and Stephen Powers (1999). *Deaf and Hearing Impaired Pupils in Mainstream Schools*. London: David Fulton.