Faculty of Arts
Department of Translators and Interpreters

Faculty of Medicine, Health and Human Sciences
Department of Linguistics

Subtitles for Access to Education

The impact of subtitles, subtitle language and linguistic proficiency on cognitive load, comprehension, reading and processing in different styles of asynchronous, online university lectures

Thesis submitted for the degree of doctor in Linguistics at Macquarie University by:

Thesis submitted for the degree of doctor in Applied Linguistics: Media Accessibility at the University of Antwerp to be defended by:

Senne M. Van Hoecke
BA Applied Linguistics, University of Ghent, 2018
MA Interpreting, University of Ghent, 2019

Supervisors:
Prof. Dr. Iris Schrijver – University of Antwerp
Prof. Dr. Isabelle Robert – University of Antwerp
Prof. Dr. Jan-Louis Kruger – Macquarie University
Prof. Dr. Marc Orlando – Macquarie University

Adjunct supervisor:
Prof. Dr. Rauno Parrila – Australian Catholic University

Period: 2019 – 2023

Antwerp/Sydney, 31 May 2023

Statement of Originality

This thesis is being submitted to Macquarie University and the University of Antwerp in accordance with the Cotutelle agreement dated 23 September 2021.

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material written by another person except where due reference is made in the thesis itself, nor does it contain any material previously published other than the published papers included and explicitly mentioned in the thesis itself.

31 May 2023

Senne M. Van Hoecke

Summary

Online lectures play a key role in today's education. They can usually be viewed and reviewed at the leisure of the learner and they are able to reach wider audiences across the globe. These student populations are also more multicultural and multilingual than before. As a consequence, an increasing number of higher education institutions are starting to use English as a medium of instruction (EMI). However, EMI might negatively influence learning performance for students with limited proficiency. Subtitles may help to overcome this language barrier, but how these fare in different online lectures is still underexplored. Moreover, few guidelines exist on what to present and how to design online lectures, meaning it is not an easy endeavor for a lecturer to produce an effective subtitled online lecture. Fortunately, cognitive theory lists a number of instructional principles that relate to cognitive processing and learning, and can provide some sort of guidance for lecturers. However, these instructional principles can be contradictory and do not offer a uniform answer as to what to consider when producing an online lecture. Research is therefore required to compare different styles of lectures and examine the effect and interactions subtitles have in these lectures.

This project examines the impact of the presence and language of subtitles on comprehension, perceived cognitive load, reading and cognitive processing in different styles of online lectures. This goal relates to three knowledge gaps: (1) the effect of subtitles on comprehension and cognitive load in education is still a matter of contention; (2) how subtitles are read and processed in different contexts remains underexplored; and (3) the question of how lecture styles impact viewing and processing, and how lecture styles interact with subtitles has so far remained largely unanswered. In addition to addressing these knowledge gaps, this project contributes to the methodological foundations of audiovisual translation (AVT) research by presenting a stepwise approach to prepare experimental AVT research.

Specifically, this project consists of a number of experiments to first set out a methodological approach and thoroughly prepare the material used in further experiments. Following this preparatory process, two eye tracking studies were conducted to address the knowledge gaps above. The first experiment was conducted with L2 English speaking students in Belgium and explored the effect of the presence and language of subtitles (intralingual/English vs interlingual/Dutch) in two distinct lecture styles (talking head vs voice-over PowerPoint). It focused mainly on comprehension, perceived cognitive load and visual attention distribution of students watching these recorded lectures. Additionally, students were interviewed to examine their perceptions of different styles of lectures and subtitles in online lectures. The second experiment was conducted with L1 English speaking students in Australia and explored the effect of three different lecture styles (talking head, voice-over PowerPoint and composite/picture-in-picture) with English subtitles on comprehension and perceived cognitive load, but also on viewing, reading and cognitive processing. With these two experiments, the project attempts to comprehensively answer questions about the impact of subtitles and lecture styles in online education.

Table of Contents

List of Tables

List of Figures

Acknowledgments

First and foremost, I wish to express my sincere gratitude to my supervisors:

Professor Iris Schrijver has supervised me closely for my entire period working at the University of Antwerp. She has gone above and beyond to not only help me with this thesis, but also guide me in becoming a full-fledged academic. Thank you for encouraging and pushing me to become a better researcher, as well as creating opportunities for me. I also wish to thank you for all the pleasant talks, lunches and drinks we had. I could not have wished for a more knowledgeable and caring supervisor.

Professor Isabelle S. Robert has always been watching me from the shadows, providing feedback and guidance whenever necessary. You have challenged me at times when I needed challenging, you have delivered just criticism when it was necessary and have made me a stronger researcher, academic and person in the process. Thank you for being a great supervisor.

Professor Jan-Louis Kruger, with whom professor Iris Schrijver put me into contact to commence the cotutelle PhD agreement between the University of Antwerp and Macquarie University, has taken care of me during my time at Macquarie University. In this period, I had the pleasure to enjoy challenging and thoughtful discussions about research with you, alongside a few responsible drinks and the occasional braai. When I returned to Antwerp, your support continued as if I was still at Macquarie University. Thank you for challenging and supporting me as a supervisor and entertaining me as a friend.

I wish to sincerely thank all three of you for your unwavering and invaluable support throughout this entire PhD, as supervisors, as colleagues and as friends. I could not have done this without you and sincerely hope our paths will cross again in the future so I may one day have the pleasure to be working with you again.

I also wish to extend my gratitude to professor Marc Orlando and professor Rauno Parrila. Both of you have supported me during my stay at Macquarie University. Admittedly, we did not have the chance to meet often as I had already made considerable progress with my PhD before I arrived at Macquarie University. Still, the few meetings we had were always enlightening and encouraging for me. It has been a pleasure to work with you and I hope I will see you once again in the future.

I wish to thank my colleagues both at the University of Antwerp and Macquarie University for welcoming me into the office and sharing a wonderful time with me there. I have thoroughly enjoyed being a part of both institutions and was fortunate to have great, supportive colleagues all around. A special word of thanks goes to professor Aline Remael, who had a significant hand in the development of the research project, and professor Frank Albers, who aided in writing and recording the lectures used in this thesis.

Last but not least, I wish to thank my family, friends and especially my partner, Luca, for their never-ending support in these past years. To my parents, thank you for making me who I am and providing me with every opportunity in life. To my big brother, thank you for being my best friend, helping me whenever necessary. To my friends, thank you for all the joy you have brought me in the past years, distracting me from work and woes when needed. To my partner, thank you for putting up with me in stressful periods during which I, admittedly, am not a pleasant person to live with; thank you for being there for me every single time, without fail; thank you for your love; and thank you for being you.

Part I: Research Context

Chapter 1 – Introduction

The introduction provides an overview of the thesis. The following section offers a general introduction to the project, its research background and its overarching research question. Section 2 of this chapter specifies the thesis structure with a brief summary of each chapter.

1 General introduction

In the present day, higher education institutions are faced with two major challenges: (1) increasingly multilingual and multicultural student populations due to globalization; and (2) an increased demand for online and recorded lectures. The past decade has already seen numerous higher education institutions offering more English-taught programs to address the first challenge (Wächter & Maiworm, 2014). The change of instructional language, however, may have a negative impact on students and/or lecturers with limited proficiency in English. One way to compensate for this drawback might be the introduction of subtitles in education.

Subtitles are probably the most common form of audiovisual translation (AVT). They provide a written version of sound in film, television or other audiovisual multimedia. Twenty years ago, subtitles in education might have seemed far-fetched. However, with the development and constant improvement of speech-to-text services, the concept of subtitles in education has never looked more plausible. Even more realistic is the introduction of subtitles in asynchronous (recorded), online education. In conjunction with the growing offer of recorded lectures in response to the second challenge mentioned above, the introduction of subtitles in education is only a small step away and is already a reality at many institutions.

Because of the widespread and ever-expanding use of subtitles for audiovisual content, the topic has received considerable interest in research since the turn of the millennium, yet the knowledge of subtitles and their impact on cognitive processing, cognitive load and comprehension, especially in education, is limited. There are a number of reasons for this: Firstly, the main topic of interest so far has been either accessibility or language learning (Díaz Cintas, 2020). Subtitles were initially introduced as a way to provide access to media for the hearing impaired (Downey, 2008). Subtitles for the deaf and hard of hearing (SDH) has therefore remained at the heart of research on subtitles. Going from accessibility to education, the other frequently visited topic has been language learning. The visual presentation of auditory information in the same or in a different language than the source has been of great interest for those passionate about language learning. As such, the studies proving benefits of subtitles for vocabulary acquisition and general language learning are plentiful (e.g., Bird & Williams, 2002; Danan, 2004; Montero Perez, 2020). The second reason for the limited knowledge of subtitles and processing, cognitive load and comprehension is that the field of research into subtitles, and, more broadly, AVT, is still in early development. Consequently, methodologies regularly differ from one study to another, leading to inconsistencies with regard to their findings. A third and final reason we would like to mention is the inherent complexity of subtitles and the processing thereof. Subtitles rarely occur in isolation. Instead, they are part of a highly dynamic and multimodal context. They are regularly accompanied by sound and video. Not to mention that subtitles are already "dynamic" (due to their limited duration), which makes the reading of subtitles distinctly different from the reading of a static text.

In sum, the research on the use and effects of subtitles is far from complete. Specifically with regard to subtitles in education, a number of questions remain: What are the effects of subtitles on cognitive load and comprehension in education? How are subtitles and other concurrent on-screen content viewed, read and processed? How does the design of a lecture affect the reception, perception and processing of subtitles?

The effects of subtitles on cognitive load and comprehension in lectures is a topic that has been touched upon a number of times before. Results, however, are still inconclusive. Some studies report benefits of subtitles with no effect or a decrease in cognitive load and/or increased comprehension (e.g., Kruger et al., 2013, 2014; Kruger & Steyn, 2014; Vulchanova et al., 2015), whereas others report detrimental effects of subtitles, such as an increase in perceived cognitive load (e.g., Craig et al., 2002; Diao et al., 2007). While the effects of subtitles are still a matter of contention, it has been made clear that they depend on a large number of factors, e.g., the language of the subtitles and proficiency of the audience (Bisson et al., 2014; Hefer, 2013a, 2013b), the presence of other content (Liao et al., 2021; van der Zee et al., 2017) or the actual reading of subtitles (Kruger & Steyn, 2014).

When it comes to this reading of the subtitles, there are surprisingly few studies that have detailed the reading process of subtitles (e.g., d'Ydewalle & De Bruycker, 2007; Liao et al., 2021; Perego et al., 2010). The reading of static text has been very well documented (Radach & Kennedy, 2013; Rayner, 1998; Reichle, 2021), but with the addition of concurrent visual and auditory content, and a pre-determined reading pace, the reading of subtitles bears many dissimilarities compared to the reading of static texts. It is well-established that subtitles are being read largely automatically (d'Ydewalle & Gielen, 1992), but how this reading progresses specifically and how the subtitles are integrated with other information has not yet been answered comprehensively. The use of accurate eye trackers can provide valuable insight into the matter. Two studies that do go into considerable detail on the processing of subtitles were conducted by Liao and her colleagues (Liao et al., 2022; Liao et al., 2021). Using fine-grained eye-tracking measures, they revealed a change in reading behavior depending on the speed of the subtitles, the presence and language of concurrent auditory content, and the presence of visual content.

The fact that concurrent visual content changes the reading behavior ties into the last question asked in this project: How does the design of a lecture affect the reception, perception and processing of subtitles? There is a considerable body of research present on instructional principles and how components of lectures possibly affect cognitive load and comprehension of students (Mayer, 2014a; Sweller et al., 1998, 2019). These principles are regularly embedded in large theoretical frameworks, such as the cognitive load theory (Sweller et al., 2011) or the cognitive theory of multimedia learning (Mayer, 2014a). While they can be used to assume how a specific lecture with subtitles would affect students, so far there are, to our knowledge, only two studies (Chan, 2020; van der Zee et al., 2017) that have considered the design of a lecture alongside the use of subtitles in education. Both studies reported effects of lecture design on comprehension. In the study by Chan (2020), the effect of lecture design was only present when students had access to native language subtitles. These findings highlight the importance of lecture design in overall comprehension, but also with regard to the effects of subtitles on the learner.

In conclusion, this thesis aimed to provide a more comprehensive understanding of the effects of subtitles and lecture design on cognitive load, comprehension, and reading and processing. Using eye tracking, psychometric questionnaires and comprehension tests, among others, it answered the question: What is the effect of subtitle presence, subtitle language and students' language proficiency levels on cognitive load and processing, comprehension and viewing/reading behavior in different styles of asynchronous, online lectures? In addition to this one goal, this thesis wishes to contribute to the methodological foundation of the field of research on audiovisual translation. It does so by employing, detailing and recommending a stepwise approach for the preparation of quasi-experimental and experimental AVT research. Fulfilling these two goals, this thesis will hopefully be a valuable addition to the research conducted within the field.

2 Thesis structure

The present manuscript is written as a thesis by publication. The thesis is comprised of seven chapters, including five standalone published/submitted papers. Since these are standalone papers, some degree of repetition throughout this thesis is to be expected.

Chapter 2 presents an overarching literature review relevant to this thesis. It provides the foundation and presents the reader with the research background pertaining to the different articles constituting the thesis. Therefore, although every paper has its own literature review, Chapter 2 provides a more substantial, thorough and highly detailed version of the literature relevant to all papers. To avoid confusion, the literature review is designed in such a way that it commences with research on static reading and eye movements and gradually introduces new components until the multimodal environment, that is subtitles in education, is attained.

Chapter 3 presents the research aims and hypotheses that lie at the heart of this thesis. Additionally, it provides insight into the aims and methodologies of each respective study that was conducted as part of this thesis. Methodologies and approaches changed over the course of this thesis (e.g., a transition from global eye movement analyses to local eye movement analyses using different eye-tracking systems). Chapter 3 intends to shed some light on how and why these changes were implemented while simultaneously providing more detailed information on core materials that stayed the same throughout the thesis.

Chapter 4 consists of two standalone papers, which report on four pilot studies. These studies were conducted to lay a solid foundation for the further studies that were planned as part of this thesis. Their secondary aim was to improve overall methodological practices in the present field of research. Though it is not a research question, it is a de facto aim of the present thesis to contribute to a more valid, replicable and generalizable research practice for the future of the field.

Chapter 5 is comprised of two standalone papers, which report on a large-scale eye-tracking study conducted with L2 English speakers in Belgium. It examined the effects of subtitle presence and language in different asynchronous, online lecture formats on comprehension, cognitive load and global eye movements. Additionally, it provides valuable insight into learner's perception investigated in a qualitative part of the study.

Chapter 6 consists of a single standalone paper, which presents the results of a large-scale eye-tracking study with L1 English speakers in Australia. It predominantly focuses on local eye movements and how cognitive processing is affected by lecture styles in a subtitled, asynchronous, online lecture.

Chapter 7 summarizes the findings, contributions and limitations of this PhD study, and also suggests some possible future research avenues.

3 References

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*(4), 509–533. doi:10.1017/S0142716402004022

Bisson, M.-J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics, 35*(2), 399–418. doi:10.1017/s0142716412000434

Chan, W. S. (2020). *An investigation of subtitles as learning support in university education* [Unpublished doctoral dissertation]. Macquarie University, Sydney, Australia.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94*(2), 428–434. doi:10.1037//0022-0663.94.2.428

d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist, 12*(3), 196–205. doi:10.1027/1016-9040.12.3.196

d'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In K. Rayner (Ed.), *Eye movements and visual cognition* (pp. 415–427). New York, NY: Springer.

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta, 49*(1), 67–77. doi:10.7202/009021ar

Diao, Y., Chandler, P., & Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology, 120*(2), 237–261.

Díaz Cintas, J., & Szarkowska, A. (2020). Introduction: Experimental research in audiovisual translation - Cognition, reception, production. *The Journal of Specialised Translation, 33*, 3–16.

Downey, G. J. (2008). *Closed captioning subtitling, stenography, and the digital convergence of text with television*. Baltimore, MD: John Hopkins University Press.

Hefer, E. (2013a). Reading first and second language subtitles: Sesotho viewers reading in Sesotho and English. *Southern African Linguistics and Applied Language Studies, 31*(3), 359–373. doi:10.2989/16073614.2013.837610

Hefer, E. (2013b). Reading second language subtitles: A case study of Afrikaans viewers reading in Afrikaans and English. *Perspectives, 21*(1), 22–41. doi:10.1080/0907676x.2012.722652

Kruger, J.-L., Hefer, E., & Matthew, G. (2013). *Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts.* Paper presented at the 2013 Conference on Eye Tracking South Africa, Cape Town, South Africa.

Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research, 7*(5), 1–15. doi:10.16910/jemr.7.5.4

Kruger, J.-L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly, 49*(1), 105–120. doi:10.1002/rrq.59

Liao, S., Yu, L., Kruger, J.-L., & Reichle, E. D. (2022). The impact of audio on the reading of intralingual versus interlingual subtitles: Evidence from eye movements. *Applied Psycholinguistics, 43*(1), 237–269. doi:10.1017/S0142716421000527

Liao, S., Yu, L., Reichle, E. D., & Kruger, J.-L. (2021). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading, 25*(5), 417–435.

Mayer, R. E. (2014a). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43–71). New York, NY: Cambridge University Press.

Mayer, R. E. (2014b). Principles based on social cues in multimedia learning: Personalization, Voice, Image, and Embodiment principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 345–368). New York, NY: Cambridge University Press.

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, 1–25. doi:10.1017/s0272263119000706

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology, 13*(3), 243–272. doi:10.1080/15213269.2010.502873

Radach, R., & Kennedy, A. (2013). Eye movements in reading: Some theoretical context. *The Quarterly Journal of Experimental Psychology, 66*(3), 429–452. doi:10.1080/17470218.2012.750676

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Reichle, E. D. (Ed.) (2021). *Computational models of reading: A handbook*. New York, NY: Oxford University Press.

Sweller, J., Kalyuga, S., & Ayres, P. (2011). *Cognitive load theory* (1 ed.). New York, NY: Springer.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. doi:1040-726X/98/0900-0251

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292. doi:10.1007/s10648-019-09465-5

van der Zee, T., Admiraal, W., Paas, F., Saab, N., & Giesbers, B. (2017). Effects of subtitles, complexity, and language proficiency on learning from education videos. *Journal of Media Psychology, 29*(1), 18–30.

Wächter, B., & Maiworm, F. (2014). *English-taught programmes in European higher education: The state of play in 2014*. Bonn, Germany: Lemmens.

Chapter 2 – Literature review

This literature review aims to give a complete overview of cognitive processing in a subtitled educational environment. The focus here will be eye movements and effects, disregarding complex neurology. It is written in such a way that it gradually introduces more factors in the discussion of said processing, starting from the core processes of reading in general to reading of subtitles specifically and the effects of subtitles, concluding with processing in multimodal contexts.

1 The process of reading

1.1 The basics

As much as reading may seem trivial to the reader, it is a highly complex skill that involves many different stages between extracting visual verbal information with the eyes to comprehending the text. This section will not discuss the decryption of visual verbal information, the language processing involved or the cognitive structuring of the information in order for one to understand the visual words. It will instead focus on the very basics of the process of reading and its first stage, namely the eye movements during the reading of words and/or logograms. As stated in the eye-mind hypothesis (Just & Carpenter, 1980), the mind attends to where the eye is fixated (more detailed information on the eye-mind hypothesis can be found in Chapter 2, Section 3.3.3). This hypothesis highlights the importance of eye movements for the understanding of reading. Initially, four distinct eye movement topics will be discussed: 1) The retina and visual acuity; 2) Saccades and fixations; 3) Perceptual span; and 4) Eye movement control. For a more extensive review of eye movements in reading, see Rayner (1998), Radach and Kennedy (2013) and Reichle (2021).

1.1.1 The retina and visual acuity

Before explaining specific eye movements, it might be helpful to have a basic understanding of the eye and its limitations (Figure 1). When images in the form of light reflection enter our eye through the pupil, the light is reflected on our retina. The retina consists of multiple layers. One of the layers in the retina contains two types of photoreceptor cells: rods and cones. Rods serve to detect motion and are specialized to function in dim-light conditions, providing only black-and-white vision. Cones are designed for bright conditions, discerning colors and allowing high-detailed vision (Rayner et al., 2012). The highest density of cones is present in the foveal pit (Kolb, 2005). The density of cones decreases and the density of rods increases as we move away from the central point in the foveal pit, fovea centralis or fovea for short, towards the parafovea, which contains a mixture of cones and rods, and eventually the perifovea, which only contains rods (Rayner et al., 2012).

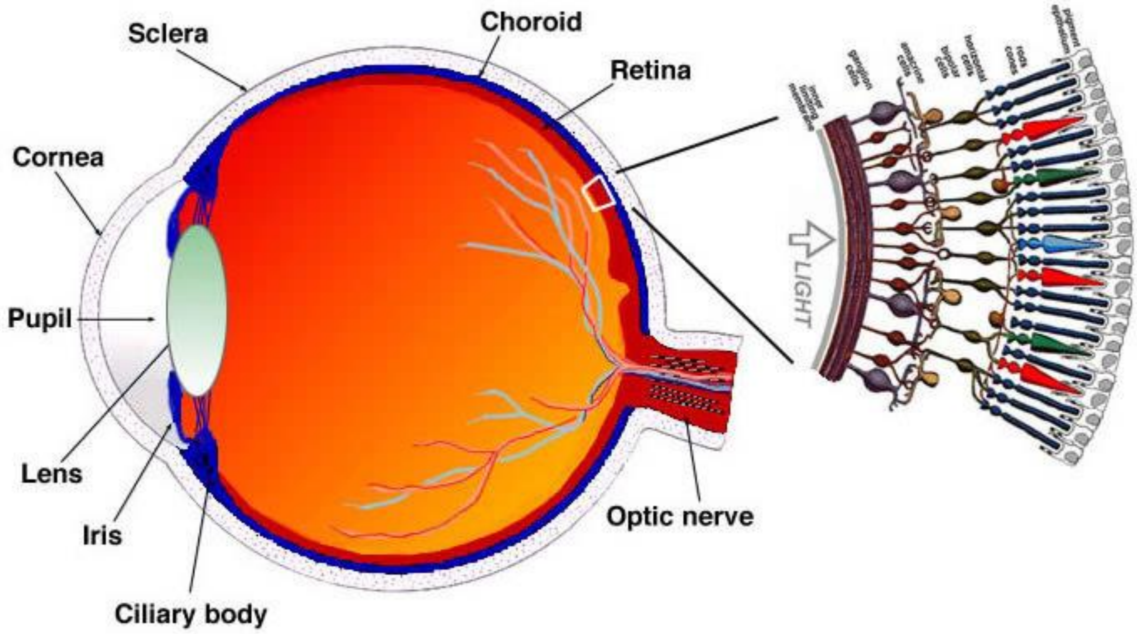


Figure 1: A section through the human eye with enlarged retina (Kolb, 2005)

However, the fovea only captures about 2° of our vision (about 6-8 characters in reading from the point of fixation) and the parafovea about 5° (about 15 characters from the point of fixation) (Frey & Bosse, 2018; Rayner, 1998). Everything beyond the parafovea is the peripheral region. As the cone density decreases when we move away from the fovea, so does the acuity of our vision. This implies that, for certain tasks where we need to be able to discern small details and thus require high visual acuity, e.g., reading, we need to compensate for the size of our fovea by continuously moving our eyes, fixating on new letters or words as we continue along a text. In sum, to read something, we need to move our eyes.

1.1.2 Saccades and fixations

A second important aspect related to the study of eye movements are saccades and fixations. When reading, our eyes do not glide over the text in a continuous pattern, they instead jump back and forth in a large numbers of rapid eye movements, called saccades. Saccades are usually between 4 to 8 characters long during text reading and take between 20ms to 50ms to complete, though this is highly dependent on the length of the saccade (Rayner, 1998). During saccades, no information is acquired and the visual information is blurred. This phenomenon is called saccadic suppression (Matin, 1974). For a simple demonstration, look into a mirror moving your eyes from left to right. You are able to make eye contact with yourself but will never be able to see your eyes moving (Krekelberg, 2010). The only moments you will be able to see yourself is during the short periods the eyes remain still, called fixations. Fixations in reading typically lie between 200ms and 300ms (Rayner, 1998), even if the range is much larger going from 100ms and shorter fixations to 400ms and longer fixations (Rayner et al., 2012). The duration of a fixation is highly dependent on a number of different factors, e.g., text (complexity/type), font and word (frequency/complexity) (Rayner et al., 2012).

Because of saccadic suppression, all information in reading is acquired during fixations (Wolverton & Zola, 1983). Based on previous research (cf. Rayner, 1998), it seems that the reader only needs 50ms to 70ms of fixation time to extract the necessary visual information required for reading. Reichle et al. (2003, p. 446), however, emphasize that this does not mean that the word is recognized and processed within that time, rather the information for reading reaches the processing system. Consequently, the actual duration of a fixation varies. The majority of studies (e.g., Findelsberger et al., 2019; Henderson et al., 1999; Just & Carpenter, 1980; Rayner et al., 2004) show that higher linguistic complexity generally results in longer fixations. Two major categories of influence can be identified: Firstly, fixation duration is influenced by word complexity. This word complexity is composed of word frequency, word length, word ambiguity, word predictability (based on context), word familiarity and factors regarding age of acquisition. Fixations tend to be longer for lower frequencies, longer lengths, higher ambiguity, lower predictability and lower familiarity (Clifton et al., 2007; Rayner et al., 2012; Reichle et al., 2009). These influences can, however, not be seen in isolation as, for example, low-frequency words can also influence and lead to longer fixations in subsequent words as a result of a spillover effect (Findelsberger et al., 2019). The second major influence is syntax. In terms of syntactical influences, it has been shown that syntactic ambiguity and inference regarding antecedents referring to a previously mentioned entity may lead to longer fixations (Clifton et al., 2007; Rayner et al., 2012). Furthermore, fixations tend to be longer at the end of a sentence or clause than in the middle, i.e., clause wrap-up effect (for an overview on specific studies, see Rayner et al., 2012).

Even though no information is acquired during saccades, they are a vital part of reading and can give significant insight into (attentive) reading patterns. However, not all saccades are usually considered in research and, additionally, some eye movements should be distinguished from saccades. When we say our eyes are still during fixations, this is not completely true. Our eyes have a constant tremor called nystagmus (Rayner, 1998), which is often disregarded in research. Furthermore, our eyes also tend to drift during fixations. These drifts are frequently followed by microsaccades, small corrections to refixate the eyes on the correct position (Rayner, 1998). These are generally also disregarded in reading research. Secondly, saccades also need to be distinguished from pursuit eye movements (i.e., slower and sometimes anticipating eye movements), vergence eye movements (i.e., inward eye movements to fixate on something closer like your nose), or vestibular eye movements (i.e., the rotation of the eye to compensate for movement of the head) (Rayner, 1998). These three are evidently less relevant for the present thesis, but are still mentioned here for the purpose of completeness.

More relevant for this thesis are the distinctions that can be made between different saccades. Saccades generally move forward when reading a text but sometimes they also move backwards in a text, against the direction of reading. Depending on whether these occur within a single word or between words, they are called refixations or regressions/regressive saccades, respectively. These are strong indicators of reading difficulty and are likely to be related to problems with language processing (Reichle et al., 2003, p. 448). About 10-15% of the saccades while reading an average text are regressions (Rayner, 1998). The frequency of regressions increases with reading difficulty (Inhoff et al., 2019), e.g., incongruency between verb and noun (Sturt & Kwon, 2018) or ambiguous sentence structures (Mitchell et al., 2008). In a static text, regressions allow the reader to reevaluate sentence structures, double-check the reference of pronouns and better understand the text as a whole. In fleeting texts, such as subtitles, the time constraint and text segmentation of the text might make it more difficult for a reader and might thus change regressions and eye movement significantly.

Another important component of reading and saccades in reading is the concept of saccade latency. To read continuously, the reader needs to repeatedly devote time to planning the next fixation and, consequently, saccade. The time needed to plan and execute a saccade is called saccade latency (Rayner, 1998). Based on previous research, saccade latency is estimated to be between 180ms and 250ms, meaning every next saccade is likely to be planned within 100ms of a fixation (Reichle et al., 2003). Considering both the average fixation duration and saccade latency, it is highly likely that the planning of saccades can be done in conjunction with language processing (Rayner, 1998).

1.1.3 Perceptual span

Another important concept in reading is the perceptual span. Perceptual span can be defined as the region from which readers gather useful information (Rayner et al., 2010). The perceptual span is different from the visual span, i.e., the number of letters that can be recognized horizontally during a fixation in reading, and the visual attention span, i.e., the number of visual elements that can be processed simultaneously during a fixation (not limited to reading only) (Frey & Bosse, 2018).

Previously, the fovea and parafovea regions of the retina were discussed. While the fovea is our main source of accurate information during reading, the parafovea has its role in reading as well. The moving window technique can provide insight into the matter (McConkie & Rayner, 1975). This technique makes use of a window that moves with the eye movements of the reader. The size of the window can be determined by the researcher. The text inside the window is clear, whereas the text on the outside of the window is blurry. When the window is too small, the reading pace decreases. As the window grows in size, the reading pace normalizes up to a certain window size, that is the global perceptual span, i.e., an area that includes both the high visual acuity foveal view and the low visual acuity parafoveal view. In terms of standard alphabetical characters, the foveal view extends to about six to eight characters left and right, whereas the parafoveal area extends to up to 15 characters (Häikiö et al., 2009; Pollatsek et al., 1993). It is important to mention that the perceptual span is asymmetrical. It extends to 15 characters in the direction of reading, e.g., to the right in left-to-right script like English, and is assumed to only extend to about three or four characters in the opposite direction and not up or down (McConkie & Rayner, 1975). In more recent research, however, reading performance was affected by changes in letters up to two words or approximately 11 characters away from the fixation in the opposite direction of reading, meaning the specific size of the perceptual span is still debated (Jordan et al., 2016). Additionally, the size of the perceptual span is not constant. It is influenced by processing demands caused by, for example, text readability (Rayner, 1986), and reader characteristics, for example, average reading speed (Rayner et al., 2010). The perceptual span can also grow as reading experience grows, e.g., children vs. adults (Häikiö et al., 2009). The perceptual span consists of three different regions of information retrieval: information on word length (the largest region), information on letter features (retrieving general shapes of letters) and information on letter identities (the smallest region) (Häikiö et al., 2009).

With the perceptual span in mind, it has been shown that the processing of a word that is in the parafovea can already begin before the word is being fixated (Schotter et al., 2012). Information can be extracted parafoveally and can later be integrated with a subsequent and frequently shorter foveal fixation. In some cases, entire (short) words can be identified parafoveally. If this is the case, these words are often skipped. In an average reading task, about 30% of the words are never fixated (Schotter et al., 2012).

1.1.4 Eye movement control

Having established how the reader extracts information and how more than just the fixated information can be processed, the question that remains is: how does our brain combine all this and consistently make decisions to continue the reading process? When do we move our eyes and where do we move them to?

In Section 1.1.2 of this chapter, we already established that fixation duration is strongly influenced by linguistic complexity, more specifically word complexity and syntactical complexity. This largely determines when we move our eyes. In addition, this decision is influenced by the reading goal the reader of a text has. As shown in Swets et al. (2008), depending on whether comprehension questions after a self-paced reading test were generic or very specific, the reading speed of participants changed. Slower reading occurred when more specific information had to be retrieved and retained from the text. Another influence is the reading speed of an individual, i.e., the speed at which someone is able to read and process words. Each person has a different reading speed. Reading speed and reading accuracy is assumed to increase through practice. This means that more frequent readers become better at correctly reading words and can do so at faster rates (Kuhn et al., 2010; Lobier et al., 2013). Reading speed can be increased artificially, though this generally comes at the expense of reading accuracy, meaning more words are incorrectly read and processed. As this is important for the reading of fleeting text, like subtitles, a more detailed discussion can be found in Section 2.2 of this chapter. This list of influences is not exhaustive, but can give an initial picture of how fixation duration changes as our minds allot more or less time to process more or less complex matter.

The question of where to move our eyes is a little more difficult. On a word level, these decisions seem to be influenced by length of the fixated word and parafoveal information of the word length of the next word and the predictability of the next word (higher predictability words more frequently leading to word skipping), among others. On a higher level (textual), the reading goal and time granted to read a text, for example, play a role. In Liao et al. (2021), participants were found to skip more words in subtitles overall, and especially at the end of sentences, when subtitles were faster (i.e., time to read the text was shorter). When fixating a new word, the first fixation does tend to be between the beginning and the middle of the word (for a more detailed overview, see Reichle et al., 2003 and Rayner et al., 2012).

1.2 Models of eye movement control in reading

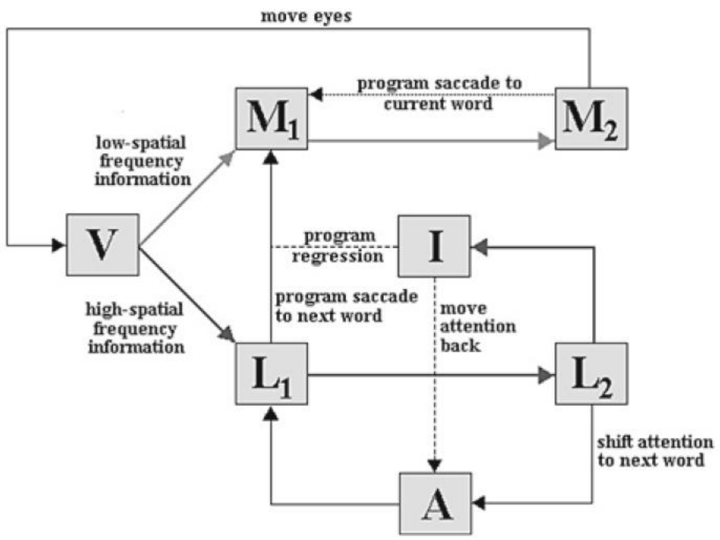
The previous section revealed how complex the reality of eye movement is. To explain all stages and processes the eye goes through when reading, researchers have developed several models of eye movement control in reading. This section sheds light on two of the more popular and widely used models, namely the E-Z Reader 10 Model (Reichle et al., 2009) and its main competitor, the SWIFT II model (Engbert et al., 2005).

1.2.1 E-Z Reader 10 Model

The E-Z Reader 10 model aims to predict how cognitive processes determine the movement of the eyes during reading. It was devised by Reichle et al. (1998) based on the work from Morrison (1984). Since its original development, it has seen many updates (currently version 10). It is based on two core assumptions. Firstly, it assumes that attention is allocated one word at a time, i.e., serially. Secondly, the completion of certain stages of processing on one word is the trigger to program or execute a saccade to move to the next word, thus lexical processing is directly linked to the moving of the eyes (Reichle et al., 2006). The following paragraphs provide a more detailed explanation of the different stages and processes involved in reading according to the E-Z Reader model. For a more in-depth overview of the model, see Reichle et al. (2003) for E-Z Reader 7, Reichle et al. (2006) for E-Z Reader 9, Reichle et al. (2009) for E-Z Reader 10 and Reichle (2021) for the latest update of the model.

The first assumption of the E-Z Reader model is that the processing of a word happens in two stages, namely an initial visual processing stage followed by a lexical processing stage. The visual processing stage can already start before a word is fixated and can thus be pre-attentive.

The rate of visual processing is dependent on two factors: Firstly, a minimum of 50ms is required as it takes about 50ms for information on the retina to be transmitted to the brain (Foxe & Simpson, 2002). The second factor is visual acuity. As was mentioned before, the parafoveal view of readers can extend to about 15 characters to the right or left (depending on the reading direction of the language) and about 3-4 characters in the opposite direction of reading (Häikiö et al., 2009; McConkie & Rayner, 1975). This parafoveal preview allows visual processing to occur. The closer the character is to the center of the fixation, the higher the rate of visual processing. This also implies that visual processing of a word is accelerated when the word is fixated near its center, a consistent find in previous research (Reichle et al., 2006). The early visual processing stage (denoted by V in Figure 2 below) provides the brain with low-spatial frequency information, such as word length, shape and boundaries between words and sentences (i.e., whitespaces, capital letters, punctuation), necessary for selecting saccade targets and programming saccades and with high-spatial frequency information, consisting of key features necessary for letter/word identification in the subsequent lexical processing stage.



V = Early processing stage

L_1 = Lexical familiarity check

L_2 = Completion of lexical access

A = Attention shift

I = Textual integration

M_1 = Labile saccadic program

M_2 = Non-labile saccadic program

Figure 2: The E-Z Reader model (Reichle et al., 2009)

The moment attention is allocated to a word, lexical processing starts. Lexical processing is divided into two stages. The first stage is called the “familiarity check” (denoted with L1 in Figure 2), whereas the second stage is referred to as the “completion of lexical access” (denoted with L2 in Figure 2). During the first stage, it is assumed that there is no full lexical access. The reader simply identifies the orthographic and/or phonological form of the word. The duration of this first stage is influenced by the normative frequency of a word in printed text and its predictability within the sentence or, in some cases, the previous sentence. If the reader succeeds in predicting the word, the duration of this first stage is 0. This has previously been observed in studies (Reichle et al., 2009) where readers skip short, highly predictable words. After a word is recognized by the reader, the second stage of lexical processing commences. In this stage, full lexical access is achieved and the word’s semantic information is retrieved. The duration of this second stage of lexical processing is a fixed proportion of the duration of the first stage, and thus also influenced by the frequency and predictability of a word. Excluding visual acuity effects and including the minimum of 50ms needed for visual processing, the E-Z Reader model predicts the time to identify words (that were not predicted by the reader) to be between 151ms and 233ms.

After the two stages of lexical processing, attention can be shifted to the next word with a mean duration of 50ms (denoted by A in Figure 2). A saccade would also be expected to happen. But before this can happen, a saccade has to be programmed. Saccadic programming happens in two stages: a labile stage (denoted with M1 in Figure 2) and a non-labile stage (denoted with M2 in Figure 2). When the first stage of lexical processing ends, the programming of a new saccade starts. The program enters into its labile stage, which is still subject to cancelling. During the labile stage, the oculomotor system is first prepared and engaged to begin programming the saccade. If the program is cancelled in this first stage, the time to prepare the oculomotor system is not lost and is transferred to the new program. Then, using the low-frequency spatial information retrieved during visual processing, the coordinates of a spatial target are converted into a saccadic distance to calculate the muscle force needed to move the eyes to the target location. If the program is cancelled during this second sub-stage, the time spent is lost as the destination of the new saccadic program requires a new calculation of necessary muscle force. In the second, non-labile stage of saccadic programming, the movement of the eyes is ordered. When the processing system enters this second stage, the saccade can no longer be cancelled or modified. The saccade will therefore be executed, regardless of whether its target location is still desired or not. The respective times required to complete these two stages of saccadic programming have means of 100ms and 25ms according to the E-Z Reader model. If the previously mentioned eye-mind lag of 50ms is taken into account, the minimal time required to program a target saccade and execute the eye movements is predicted to be 175ms (the mean saccadic latency) (Reichle et al., 2006).

Because of the difference between the time required to program and execute a saccade and the time required for lexical processing, it is possible that word identification has already completed before the eyes are moved to the next word. When this happens, lexical processing of the next word can already start based on the information retrieved from the parafovea, meaning saccadic programming and lexical processing can happen simultaneously. The time that can be allotted to this parafoveal processing is highly dependent on the difficulty of the word, which affects the duration of lexical processing. This creates the spillover effect mentioned in the Section 1.1.2 of this chapter (Findelsberger et al., 2019). If a word is difficult to process, it will also affect the time required to process the next word as less time could be spent on the parafoveal processing of that word.

One final thing to consider regarding saccades, is that the eyes sometimes fail to land on the targeted position. When reading a language, there is a preferred saccade length, which in English is seven character spaces. If the target location is more or less distant than the preferred saccade length, it is possible that the eye undershoots or overshoots the target location. This is the systematic error of the oculomotor system. Aside from the systematic error, the oculomotor system is also subject to random error. When the target location of a saccade does not allow for efficient lexical processing, which can be due to the previously mentioned error or failure to find an efficient target location for a saccade, the programming of a refixation saccade can be initiated. This tends to happen when the saccade target is near the end or beginning of a word, which leads to poorer visual acuity of the entire word and concludes in inefficient lexical processing. Because of this, word length has a significant effect on the chances of a refixation (i.e., longer words receive more refixations on average) (Vergilino & Beauvillain, 2000).The programming of a saccade to the current word (i.e., refixation saccade) can only happen if no new forward-moving saccadic program has been initiated.

So far, the E-Z Reader model has explained the processes in reading, but has not yet included any measure of higher-level language processing or integration of words into syntactical structures or into discourse. E-Z Reader 10 (Reichle et al., 2009), however, attempts to integrate post-lexical processing and integration as well (denoted with I in Figure 2). In this case, I is a placeholder for future, more elaborate theories on high-level, post-lexical processing and does not go into detail. It does, however, predict a duration of this post-lexical processing with a mean of 25ms and makes some additional assumptions about the process. Firstly, the predictability of a word, which has an influence on the time required to lexically process the word, is only available after the previous word has been processed and integrated. Secondly, if one fails to integrate a certain word, it results in comprehension difficulty. Failure to integrate a word also prevents the next word from being integrated successfully. This failure can then lead to an immediate pause or regression if it happens early or can lead to the eyes and attention being directed back to the point of difficulty, constituting a regressive saccade. These regressive saccades take more time to program than progressive saccades with the labile stage of saccadic programming for regressions taking 30ms longer. This creates three different scenarios: In Scenario 1, no difficulties occur: Lexical processing of one word completes so attention is shifted to the next word. The saccadic programming completes and a saccade to the next word is executed. During the lexical processing of the next word, integration of the first word is completed. The reading process is not interrupted. In Scenario 2, the reader fails to integrate the first word before the lexical processing of the next word is completed. Because of this integration failure, the attention shifts back to the previous word. The forward saccade program is cancelled and a new regressive saccade is programmed. In Scenario 3, the reader fails to integrate the first word before the labile stage of saccadic programming to the next word is completed. This labile stage is therefore cancelled. The first stage of lexical processing of the next word is also cancelled and attention shifts back to the first word. Sometimes this results in a regression within the same word or to the word before.

1.2.2 SWIFT II Model

The SWIFT model is another well-known cognitive model of eye-movement control in reading. SWIFT stands for Saccade-generation With Inhibition by Foveal Targets. Arguably the biggest difference between the SWIFT model and the E-Z Reader model is that the SWIFT model is a GAG (Guidance by Attentional Gradients) model, whereas the E-Z Reader model is a SAS (Sequential Attention Shifts) model (Richter et al., 2006). In essence, the SWIFT model proposes parallel mechanisms of attention allocation in reading as opposed to the sequential processes in the E-Z Reader model (see Figure 3 below for a visual representation of serial and parallel processing). The following paragraphs will first elucidate the key principles of the SWIFT model, highlighting similarities and differences with the E-Z Reader model, and finally provide an overview of the SWIFT model. For a more detailed description of the SWIFT model see Engbert et al. (2002), Engbert et al. (2005) and Richter et al. (2006).

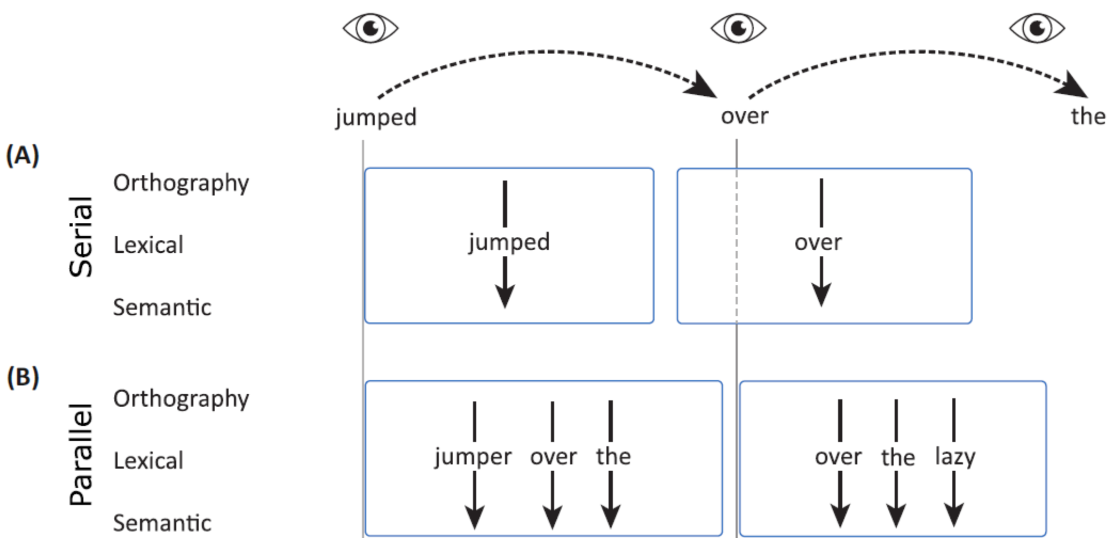


Figure 3: A visual representation of processing mechanisms (Jensen et al., 2021)

There are seven core principles to the SWIFT model. The first one claims spatially distributed processing of an activation field: the SWIFT model adopts a dynamic-field approach in which several words are processed at the same time (parallel processing). Processing rate depends on the distance between the word and the currently fixated position. Essentially, processing speed is limited by visual acuity, with the currently fixated word being processed the fastest and processing speeds decreasing moving horizontally to the right or left. Depending on the reading direction, processing speeds drop faster for words that are behind the current fixation than for words that are in front. As in the E-Z Reader model, processing speeds are also dependent on word frequency and word predictability. However, the SWIFT model defines word difficulty based on word frequency alone and modulates processing speeds based on word predictability. This would be more in line with the study of Rayner et al. (2004), which demonstrates that predictability effects were larger for low-frequency words than for high-frequency words.

The other six principles are concerned with saccades and saccadic programming. We will list them in this paragraph first and go into more detail in the subsequent paragraphs: 2) Separate pathways for saccade timing (i.e., when) and saccade target selection (i.e., where); 3) Random saccade generation with time-delayed foveal inhibition; 4) Two-stage saccade programming with labile and non-labile stages; 5) Systematic and random errors in saccade lengths; 6) Error correction of mislocated fixations; and 7) Modulation of saccade latency by saccade length.

The second principle separates the questions of when to start programming a saccade and what to target in the next saccade. This ties into the third principle, which claims saccade generation is an autonomous process, independent of lexical processing. Essentially, a random timer continuously runs in the background and decides the pace of saccades. This timer relates to the individual reader's reading speed. What this also means is that fixation durations are dependent on this random timer as well. However, to account for lexical processing difficulty (i.e., longer fixations on more complex words), this random timer is influenced by a lexical decision circuit. If processing difficulty occurs, the timer is slowed down to allow for more time during lexical processing. The SWIFT model calls this foveal inhibition. Because saccade programming is faster than lexical processing, it is possible that there is a delay in foveal inhibition. This can cause lag effects in processing. The concept of this random timer constitutes a second big difference between the SWIFT model, which assumes continuous reading is expected to be more of an autonomous process, whereas the E-Z Reader illustrates continuous reading as a multitude of consecutive, active processes.

The fourth principle of the SWIFT model distinguishes two stages in saccadic programming. Similar to the E-Z Reader model, there is a labile stage, subject to cancellation, and a non-labile stage, which cannot be stopped. One difference, however, is that in the E-Z Reader model, a target for a saccade is already picked and required muscle force is calculated. Because of the parallel processing assumption in the SWIFT model, words are competing to be the target of a saccade. The decision of what to fixate next happens when the labile stage of the saccadic program transfers into the non-labile stage. A target is selected randomly, but this is also highly dependent on the lexical activation of the target. The higher the relative activation of a word at the time a saccade target is decided upon, the higher the odds a word is the target for a saccade.

The fifth principle, similar to the E-Z Reader model, accounts for systematic and random errors in saccades. These can lead to undesirable fixation locations. These can in turn lead to immediate programming of a new saccade to correct the error (principle six). With regard to refixations, i.e., intra-word saccades, the probability of a refixation is smallest when a word is fixated in the center, or slightly left of the center to be exact, which is defined as the optimal viewing position. This implies that the chance of a refixation or what can be considered a mislocated fixation is higher when the fixation lands near the boundaries of a word. Because this is immediately followed by the start of a new saccadic program, the duration of a fixation near word boundaries is expected to be shorter. The duration of a fixation before a refixation is in turn modulated by principle 7, which assumes that saccade latency is influenced by saccade amplitude. Shorter saccades are assumed to yield higher saccade latency. Per consequence, the fixation before the shorter saccade is also longer as it must wait for the saccadic program to be completed. In turn, longer saccades have a reduced saccade latency and might cause successor effects, where the word before a long word (that is being parafoveally processed) receives shorter fixation durations.

To sum up, Figure 4 gives an overview of the SWIFT model. A random timer depending on the reading rate of the individual decides when to start programming a new saccade. At the same time, multiple words are processed in parallel with processing rates being higher closer to the foveal fixation. When the labile stage of the saccadic program goes over into the non-labile stage, a target is selected based on the lexical activation of the words currently being processed. The non-labile program orders the saccade to be executed and leads to new words in the parafovea that can be processed. Finally, should any difficulty occur during the processing, the lexical decision circuit sends a signal to the random timer to slow down the reading process (foveal inhibition). Table 1 provides a brief overview of the major differences between the E-Z Reader model and the SWIFT model.

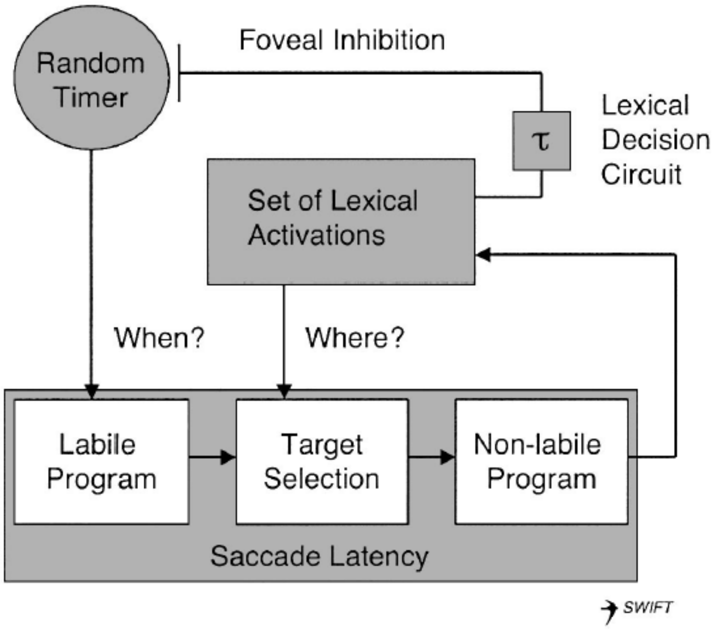


Figure 4: The SWIFT model (Engbert et al., 2005)

Table 1: Differences between E-Z Reader and SWIFT

E-Z Reader	**SWIFT**
Serial processing	Parallel processing
Local decision process for reading speed	Automatic timer for reading speed
Saccade target selection at start of programming	Saccade target selection based on lexical activation

2 The reading of subtitles

This section focuses on the reading of subtitles. Here we take a step away from eye movement control and more complex cognitive models. Instead, the focus of this section lies on establishing a basic understanding of research on audiovisual translation (AVT), and specifically subtitles.

2.1 The presence of subtitles

Audiovisual material is ubiquitous in the current day and age. With 6.259 billion smartphone users worldwide and an estimated 1.72 billion households owning a television in 2021 (Statista, 2022a; Statista, 2022b), the dominance of multimedia is undeniable. It has become so common, its inherent complexity is no longer noticed. At the turn of the 20th century, the early days of cinema, film provided only visual information as sound was not yet included. Silent movies also typically used intertitles, where scenes would be alternated with scenes containing a transcript of the dialogue in the preceding scene, which means that viewers would at any time either be reading or viewing a scene. However, in the 1920s and 1930s, as films with speech slowly started to take over (Gernsbacher, 2015), auditory information joined the mix. With the inclusion of sound and speech in movies, came the exclusion of those with hearing impairments. To allow equal access to the material, captions came into existence. In the 1970s, captions eventually reached television shows (Downey, 2008). Since then, they have only gained more and more ground up to the point where nowadays they are all but mandatory. More attention is being paid to inclusivity and accessibility for all, the corporate world wishes to maximize profits by reaching larger, multilingual audiences with accessible material, and, even legally, measures are being taken to increase general accessibility, e.g., the EU Accessibility Act (Directive 2019/882) and the renewed EU Audiovisual Media Service Directive (Directive 2010/13/EU) in Europe.

With subtitles slowly becoming as omnipresent as audiovisual material itself, the complexity of audiovisual material grows. It generally provides four types of information to be processed in different channels: 1) Non-verbal auditory information, e.g., music, ambient sounds; 2) Verbal auditory information, e.g., conversations, narration, lyrics; 3) Non-verbal visual information, e.g., images, scenes, characters; and 4) Verbal visual information, e.g., writing in film and, more prevalent, subtitles. It is important to mention that subtitles rarely occur in an isolated environment and, consequently, there is much to consider when discussing subtitles. We go into more detail on this multimodality and the aforementioned channels in Section 3 of this chapter.

Before going any further, it is important to mention that subtitles are a type of audiovisual translation (AVT). Despite this thesis focusing on subtitling, limited to two modes, intralingual and interlingual, the thesis would be incomplete if it does not at the very least mention other types of AVT and modes of subtitles. Three types of AVT have been frequently studied in the past decade, namely subtitling (including subtitles for the deaf and hard of hearing (SDH), interlingual or translation subtitles and live subtitles). The other two types of AVT are audio description (AD) and dubbing (also called lip-syncing or voice-over) (Gambier, 2009).

The first form of subtitling is SDH which is sometimes called captioning and is generally intralingual, that is in the same language as the narration. In addition, SDH includes sound descriptions. The second form of subtitling is translation subtitles that tend to be interlingual (i.e., providing a translation into the language of the audience of films that are in a foreign language). Subtitles are mostly pre-recorded and then synchronized with the film. They can, however, also be produced live. Live subtitling requires a trained mediator to either type, stenotype or velotype (quick typing) or respeak the words that are being said into a computer so they can be shown on-screen simultaneously, usually with a short delay, as the speech itself. A format of subtitle creation that is gaining ground fast is automatically generated subtitles, in which speech-to-text software is used to automatically transcribe content. These can in turn be broadcasted live from that software or can be corrected by a professional subtitler real-time to accurately generated live subtitles. While the accuracy of automatically transcribed subtitles was considerably lower than the accuracy of a respeaker a decade ago, certain systems are now capable of producing subtitles that match or even surpass the accuracy of a respeaker (P. Romero-Fresco, personal communication, April 11, 2023). AD is a type of AVT that aims to orally describe visual representations for the vision impaired. This can be done live or can be pre-recorded and included in, for example, a DVD or added to a streaming service as a separate sound track that can be activated. The last common type of AVT studies is dubbing where two forms can be distinguished namely lip-synchronized dubbing where the original dialogue is replaced with a translation spoken by voice actors and synchronized with the lip movements of the on-screen characters; and voice-over where the volume of original soundtrack is reduced and one person speaks the translation without attempts to synchronize the translation with lip movements. These three types have one particular thing in common. They are mainly aimed at accessibility, either for the hearing or vision impaired.

For this thesis, however, we are interested in pre-recorded subtitles. More importantly, subtitles or captions aimed at the general public and not necessarily the hearing or vision impaired. Subtitles can be offered in different languages and different styles. Most commonly, they present the narration in a different language, meaning they are interlingual. This is often in the native language of the audience, meaning they are L1 subtitles. If they present the content in the same language as the narration, they are instead called intralingual subtitles, and if this is not the first language of the audience, they are L2 subtitles. In some countries, including Belgium, Switzerland and China, it happens that subtitles are offered in two languages at the same time to cater to a multilingual audience (e.g., French and Dutch in Belgium). In that case they are called bilingual subtitles. If the user has the possibility to turn subtitles on or off at their own leisure, they are considered open subtitles. Subtitles that are embedded in the content, on the other hand, are called closed subtitles (Díaz Cintas & Remael, 2014).

Regardless of which mode of subtitles is shown on screen, the mere presence of the text on the screen has important implications for the viewing behavior of the audience. Back in the late 1980s and early 1990s, d’Ydewalle and colleagues conducted a number of studies on the reading of subtitles and how this was affected by sound and language proficiency. Their findings gave rise to one of the most important assumptions for research on subtitles, namely that subtitle reading is largely automatic and is not dependent on the presence of sound and does not even require the language to be known by the audience (d'Ydewalle & Gielen, 1992). In a later study by d'Ydewalle and De Bruycker (2007), the subtitle reading of 12 Dutch-speaking adults and eight Dutch-speaking children (age 10-12) was examined using eye tracking. They were shown two movie fragments, one in Swedish with Dutch subtitles and one in Dutch with Swedish subtitles. None of the participants knew Swedish. They found the Swedish subtitles were more often skipped, were fixated less frequently and showed longer latencies and fixations than the Dutch subtitles. The eye movements of the adults were similar to those of children, which were in turn in line with previous research. Despite the foreign language subtitles being skipped more, the mere presence of the subtitles did attract the viewer’s attention and invited the viewer to attempt to read the subtitles, even though they knew they did not understand what was shown on-screen.

While the study by d'Ydewalle and De Bruycker (2007) considered more than just time spent on the subtitles based on fixation indices, AVT studies frequently do not distinguish between mere visual attention and actual reading of subtitles. For long, few eye-tracking studies investigated subtitle processing (Kruger et al., 2014). Despite the research that has been conducted so far (see Section 2.3 of this chapter for references), numerous questions remain unanswered concerning the process of subtitle reading. In 2014, Kruger and Steyn attempted to make the study of subtitle processing easier by developing a measure of subtitle reading. They came up with the Reading Index for Dynamic Texts (RIDT). It quantifies subtitle reading by considering the number of unique fixations a participant has on a subtitle and dividing that by the length of that subtitle in number of average words (characters divided by average word length in characters across all subtitles). It then multiplies that by the average forward saccade length per participant in that subtitle, divided by the average word length of the video. Figure 5 below shows a visual representation of the formula (see Kruger & Steyn, 2014, for a more detailed description of the RIDT). In recent years, however, it has become easier to consider actual reading with the rise of new and more accurate eye-tracking devices, such as the SR EyeLink system, and the subsequent use of these systems in AVT research to obtain word-level eye movement data. Nevertheless, the RIDT can still serve as a rudimentary measure of subtitle reading, especially in studies using older, less accurate eye-tracking devices where the text cannot be analyzed with automatically generated areas of interest.

$$\mathrm{RIDT}_{vps} = \frac{\text{number of unique fixations for } p \text{ in } s}{\text{number of standard words in } s} \times \frac{\text{average forward saccade length for } p \text{ in } s}{\text{standard word length for } v}$$

Figure 5: The reading index for dynamic texts (RIDT) (Kruger & Steyn, 2014)

2.2 Subtitle typology

As subtitles gradually became a common sight in the world of cinema, the need increased for a common framework regarding the production of subtitles. Subtitles take up space on screen and as they are read automatically, they should be carefully crafted so as to not disturb the viewing experience. To ensure the quality of subtitles, scholars like Karamitroglou (1998) and Ivarsson and Caroll (1998) proposed general subtitling guidelines. These guidelines deal with most parts of subtitle production, including spotting, reduction, segmentation, subtitle duration, number of subtitle lines, number of characters per line, etc. While these two guidelines might be more than 20 years old, they are still widely seen as conventional in the profession. Alongside these two guidelines, there are various other similar guidelines, often composed by specific broadcasting companies to fit their own preferences and needs, such as the Norms and Instructions for Open Subtitling of the Belgian public broadcasting company, VRT (VRT, 2023), the BBC Subtitle Guideline (BBC, 2023), the English Timed Text Style Guide from Netflix (Netflix, 2023). It is, however, impossible to make any of these conventions binding as all languages and cultures have their own unique characteristics and idiosyncrasies. Differences in subtitling can be observed across the globe, most of which concern the same core parameters in subtitling. These parameters are of great importance for subtitling (Gottlieb, 2012). In this section, the core parameters are listed, alongside some other considerations regarding subtitle typology. The section only discusses "standard subtitling practice" as many of these parameters or characteristics become less relevant when creative subtitling is employed. Creative subtitles are generally integrated into the media and, in a way, become part of the visual art itself or attempt to enhance viewer's experiences and immersion (Black, 2022; Díaz Cintas & Remael, 2021; Fox, 2016; Kruger et al., 2018).

Before considering one core parameter, the following paragraphs first detail some other characteristics of subtitles. The first is subtitle line length. As mentioned before, subtitles take up space on-screen. To avoid them taking up too much space, the length of a subtitle line is limited. This length is expressed in characters per line (CPL), including blank spaces and typographical signs. Traditionally, the maximum length of a subtitle line was around 37 characters (Díaz Cintas & Remael, 2014, 2021). Nowadays, the length of a subtitle is dictated more by the space available for the subtitle. It depends more on font choice, size of characters and which characters are used (e.g., 'i' vs. 'w') (Díaz Cintas & Remael, 2021). Still, in general media, companies often decide what length to adhere to. Many broadcasting companies, e.g., BBC (2023), adhere to a limit of 37 CPL. In the movie industry, on the other hand, the norm has been between 37-39 CPL for years. Streaming services, like Netflix or Disney+, however, go even higher, allowing up to 42 CPL. In research, this same inconsistency is found: the norm varies from 32 characters, to around 35, 36 and 37 being the maximum (e.g., d'Ydewalle & De Bruycker, 2007; Díaz Cintas, 2003; Karamitroglou, 1998; Kruger et al., 2014).

A second characteristics is the number of subtitle lines. In most cases, this is limited to a maximum of two lines, though three lines are accepted provided there is space for them (BBC, 2023; Ivarsson & Caroll, 1998). It is, however, still important to consider as different reading behavior has been found for one-line, two-line and three-line subtitles (d'Ydewalle & De Bruycker, 2007; Szarkowska & Gerber-Morón, 2019; Szarkowska et al., 2021). Provided it is a language the viewer understands, two-line subtitles were shown to lead to more regular reading, i.e., subtitles were being skipped less frequently, when compared to one-line subtitles. They also led to proportionally more time being spent on the subtitle and fewer regressive eye movements (Szarkowska et al., 2021). Three lines were shown to lead to even more time spent on the subtitle and also led to a higher cognitive load as perceived by the viewer (Szarkowska & Gerber-Morón, 2019). It is unclear whether this implies there is a benefit to using two-line subtitles over one-line subtitles or they just require more time to be read as they tend to be more syntactically complex and contain more words. In any case, these subtitles do take up more space on-screen and the necessity of having to move the eye from one line to the other (i.e., return sweep) generally means more time is needed to read the subtitle fully.

If we combine the number of lines of a subtitle and the characters per line, a total number of characters in the subtitle is obtained. When duration is considered alongside the total number of characters, the result is one of the core parameters of subtitles, namely presentation speed. Presentation speed is also referred to as subtitle speed or presentation rate and is generally expressed in characters per second (CPS) or words per minute (WPM). It should not be confused with reading speed, which is the speed at which a subtitle is read by the viewer. The optimal presentation speed would be the time an average viewer of a particular audience needs to comfortably read a subtitle while also having enough time to look at the image. One of the most known, though possibly slightly outdated, rules regarding subtitle speed is the six-second rule. The six-second rule states that it should be possible to read a full two-line subtitle comfortably in six seconds and that shorter subtitles should be timed proportionally. As was described in previous paragraphs, the ideal length of a subtitle is a topic of debate. In 2003, Díaz Cintas recommended a presentation speed of 12 CPS or about 144 WPM. It is, however, possible that the average comfortable reading speed of viewers has increased over time, especially in subtitling countries, as the subtitles become increasingly more common (Gottlieb, 2012). Nowadays, a presentation speed of 17 CPS is recommended for adult programs in Latin-based languages and 13 CPS for children programs (Díaz Cintas & Remael, 2021). Streaming services have even gone beyond that. Netflix, for example, currently allows a whopping 20 CPS for adult programs in English and 17 CPS for children's programs (Netflix, 2023), much to the discontent of researchers and subtitling professionals. In response to this constant rise, the past years have seen a gradual increase in research on the topic of subtitle presentation speed. Earlier studies showed that viewers were able to keep up with reading speeds up to 20 CPS and that 12 CPS possibly even led to re-reading of the subtitles (Szarkowska & Bogucka, 2019; Szarkowska & Gerber-Morón, 2018). An increase in presentation speed may have a negative impact on the viewing experience. While in the previous studies, the presentation speed was manipulated while the subtitle content remained the same, this would not be the same in an appropriately spotted video with high presentation speeds. In such case, the content in the subtitles that has to be processed increases significantly. This comes at the expense of watching the image instead (Romero-Fresco, 2009). Liao et al. (2021) and Kruger et al. (2022) also studied the effects of reading speed (12, 20 and 28 CPS) on subtitle reading. Contrastingly, they found fewer fixations when the presentation speed increased and, more importantly, found more subtitles were being skipped when the presentation speed was higher rather than lower. Essentially, increasing presentation speed led to viewers starting to increase their reading speed and start to skim and skip the subtitles rather than thoroughly read them. More subtitles are also not read to completion at higher speeds, as exemplified by more skipping at the end of sentences. This shows the importance of presentation speed when investigating subtitle processing and has important implications for the use of higher presentation speeds in standard media.

In order to reduce presentation speed, it is possible to omit or reword content in the subtitles. This is the second core parameter: reduction. It is considered essential to cope with the time-space constraints of subtitles (Gottlieb, 2012). This, however, leads to yet another hot topic in the world of AVT, namely edited subtitles vs. verbatim subtitles (Romero-Fresco, 2009; Szarkowska et al., 2011). Generally, deaf associations demand verbatim subtitles because editing is often seen as a form of censorship, denying the deaf community equal access to media (Romero-Fresco, 2009). Verbatim subtitles, however, tend to lead to very fast subtitle presentation speeds, which, as discussed in the previous paragraph, can lead to problems of its own. So how do verbatim and edited subtitles compare? In one study (Szarkowska et al., 2011), viewers were shown to spend more time on the image than the subtitles when the subtitles were edited. This is to be expected as they contain less information to be processed. In terms of processing, however, it was the verbatim subtitles that read and processed faster than the edited subtitles. In another study by Szarkowska et al. (2016), 44 deaf, 33 hard of hearing and 60 hearing participants were tested in different presentation speed (12 vs 15 CPS) and subtitle style (verbatim vs. edited) conditions. They found participants to go back and forth between the image and the subtitle more often when the subtitles were edited than verbatim. Participants also spent more time on the edited subtitles. It is assumed that more time is spent comparing the image and the edited subtitles because participants know there might be discrepancies between them. Additionally, comprehension was considered but no difference was found between the two conditions. However, when the groups were taken into account, the deaf appeared to have benefited more from the verbatim subtitles. These results, of course, also depend on the degree of editing present in the subtitles. Nevertheless, these findings further fuel the debate of which would be the better type of subtitles and underline the importance of the audience when discussing the matter.

Another key component of subtitles is segmentation. Subtitles cut up the original narration in multiple pieces as each subtitle can contain only a certain amount of information. On top of that, subtitles can also consist of one, two or three lines, meaning this piece of information is spread across these different lines. Thus, a professional subtitler has to decide where to segment the subtitle, both at subtitle level (several subtitles) and at line level (several lines separated by line-breaks). The common rule is that each segment, line or subtitle, should ideally be semantically and syntactically self-contained (Díaz Cintas & Remael, 2014, p. 172; Ivarsson & Caroll, 1998) and "should appear segmented at the highest syntactic nodes possible" (Karamitroglou, 1998, p. 6). Ideally, each subtitle contains only a single sentence. If a sentence is too long, it can either be edited or parsed at the most logical syntactic and semantic place, e.g., not separating verbal phrases, nouns from adjectives, prepositions from their object, etc. If this is not done properly, it is expected to disrupt the reading of the viewer and increase processing loads (Perego, 2008). This assumption was confirmed by a number of studies: In a study by Perego et al. (2010) longer mean fixation durations, i.e., a sign of more extensive language processing, were identified on ill-segmented subtitles; Rajendran et al. (2013) found more fixations, longer fixations and more saccadic crossovers on word-by-word (scrolling) subtitles compared to phrase or sentence-chunked subtitles; and Gerber-Morón et al. (2018) uncovered higher cognitive loads in non-syntactically segmented subtitles when compared to syntactically segmented subtitles. However, both Gerber-Morón et al. (2018) and Rajendran et al. (2013) were unable to confirm any negative impact of segmenting on comprehension, which raises a number of questions regarding the processing difficulties associated with ill-segmented subtitles or word-by-word subtitles.

One final subtitle characteristic and another subtitle-related issue we wish to mention is the subtitle positioning and shot changes. The most common position of subtitles is arguably the bottom of the screen. It does happen that subtitles are moved around. This could be to avoid obscuring other relevant text on screen (e.g., credits) or as part of integrated subtitling. In integrated subtitling, subtitles are placed closer to the speaker or other relevant parts of the scene to increase engagement and reduce the distance the eyes have to travel (Fox, 2016; Fox, 2018; McClarty, 2014). A study by Black (2022) revealed a positive reception of integrated subtitles in children's education without having detrimental effects on processing or comprehension. While integrated subtitles are less relevant for the present thesis, the previous study does highlight the importance of subtitle positioning for educational purposes as well. Of course, when we are talking about education and subtitles, it is mainly important that the subtitles do not cover any key information shown on, for example, slides of a PowerPoint presentation or drawings on a blackboard. Shot changes have also been a frequently visited topic in research on subtitles (e.g., I. Krejtz et al., 2013; Szarkowska et al., 2017). Guidelines recommend subtitles should not carry over shot changes as they would elicit re-reading of the subtitle. Though this is highly relevant for movie subtitles, it plays a marginal role in the topic of subtitles and education (perhaps a change of PowerPoint slide could be considered a shot change).

2.3 The effects of subtitles

Research on AVT only started in the early 1970s (Díaz Cintas & Szarkowska, 2020). Initially, the field consisted of mainly descriptive studies. Only since the turn of the millennium, cognitive and empirical studies started to emerge (Díaz Cintas, 2020). This change partially mimics the development of the field of translation studies. However, as opposed to translation studies, AVT studies focus substantially more on the reception of AVT by the viewer instead of the process of subtitling by the professional (Díaz Cintas & Szarkowska, 2020). The AVT reception studies conducted so far are predominantly concerned with AVT as a tool for accessibility (i.e., AD and SDH) or the benefits of AVT regarding language learning (Díaz Cintas, 2020). This section will not touch upon the numerous studies that have been conducted on AVT and accessibility, but will instead focus on experimental studies that concern the cognitive effects of subtitles and subtitle processing. This includes language learning but the main scope of this thesis is subtitles and cognition. This main focus will be reflected in the discussion of research in this section.

2.3.1 Language learning

A considerable number of empirical studies on subtitling have already been conducted (e.g., Bird & Williams, 2002; Danan, 2004; Markham, 1999; Vanderplank, 1988). As was mentioned before, a common topic in these studies has been language learning. To illustrate the general findings regarding the relation between subtitles and (incidental) language learning, a few studies on the matter are discussed below.

Bairstow and Lavaur (2017) tested the effects of dubbing, L1 subtitles with L2 dialogue and L2 subtitles with L1 dialogue on comprehension and vocabulary acquisition. They recruited 40 students of psychology to watch an approximately 7-minute movie excerpt in either a control condition (L2 dialogue only) or one of the other three aforementioned conditions. No difference in comprehension was found between the subtitled conditions and the dubbed condition, although both were still significantly better than the control condition. Furthermore, vocabulary retention improved more with L2 subtitles, followed by L1 subtitles than it did from just the foreign language film.

Birulés-Muntané and Soto-Faraco (2016) showed 60 Spanish students an episode from Downtown Abbey, either with edited English (L2) subtitles, Spanish (L1) subtitles or no subtitles. They investigated listening proficiency, vocabulary acquisition and comprehension using pre-tests and post-tests. Listening comprehension improved when the episode was watched with L2 subtitles. Their results on vocabulary acquisition remained inconclusive, but with regard to comprehension, the L1 subtitles outperformed the L2 subtitles and the L2 subtitles outperformed the no subtitle condition.

Another group of academics who have done frequent research on subtitle and language learning is Montero Perez and her colleagues. They mostly looked at the reception and effects of different types of captions (full captions, keyword captions, glossed captions) on comprehension and vocabulary acquisition. The studies mentioned here employed Flemish students with Dutch as L1 and had them watch French video. In only two studies, captioning was shown to lead to significantly better comprehension (Montero Perez et al., 2013, 2018). Generally, captions did not necessarily lead to improved recall or comprehension. They did, however, improve vocabulary gains (Montero Perez, 2020; Montero Perez et al., 2014) and/or form recognition (Montero Perez, 2020; Montero Perez et al., 2015; Montero Perez et al., 2018). The gains were strongly dependent on the vocabulary size of the participant, with higher vocabulary size scores leading to more vocabulary gains (Montero Perez, 2020; Montero Perez et al., 2014). Regardless of whether there were or were no effects of captioning, most students felt they benefited from the full captions and reported that they would turn on full captions when possible (Montero Perez, 2020; Montero Perez et al., 2014; Montero Perez et al., 2013).

Although the studies mentioned above are only a fraction of the studies that have been conducted so far, they provide substantial insight into the matter of subtitles and language learning. The key takeaway is that vocabulary acquisition generally does benefit from having subtitles on-screen. Nevertheless, the actual gains are dependent on factors such as the current vocabulary knowledge of the viewer, the language of the subtitles and the language of the dialogue, among others. The results regarding comprehension and subtitles remain inconclusive. This is a pattern that continues in cognitive, empirical studies on subtitles that were conducted outside of the scope of language learning.

2.3.2 Cognitive processing and comprehension

The studies that do not focus on language learning or accessibility are often concerned with cognitive effects of subtitles and subtitle processing. They frequently include the concept of cognitive load and measure the cognitive load alongside comprehension and/or reading/subtitle processing. Cognitive load, in short, is the load imposed on an individual to perform a certain task or process a certain piece of information. The load can be generated by the inherent complexity of the material (i.e., intrinsic load) or any extraneous material (i.e., extraneous load). More detailed information on the concept of cognitive load will be provided in Section 3.1 of this chapter.

One topic in research on subtitles that is currently receiving significant attention is how the language of the subtitles affects the cognitive effects of subtitles and the reading of the subtitles. In 2013, Hefer (2013a, 2013b) showed that L2 speakers of English spent more time reading English subtitles than L1 speakers of English in her sample. Depending on the language of the subtitles, which was Sesotho or English in one of the studies, the total fixation duration, dwell time (i.e., the total time spent in the area of interest, including both fixation and saccade durations) and fixation count on the subtitles were different. This reinforces the assumption that the language of the subtitles and language proficiency of the audience plays a key role in subtitle processing.

Another example of a study that considered global subtitle reading depending on the language of the subtitles and even soundtrack was conducted by Bisson et al. (2014). They used eye tracking to study the reading of standard, reversed and intralingual subtitles. Thirty-six English-speaking participants who had no knowledge of Dutch were included in the eye tracking analysis. They found that the English and Dutch subtitles were fixated more (as measured by total fixation duration and fixation count) and skipped less frequently when the soundtrack was in Dutch compared to when the soundtrack was in English. In other words, participants looked at the subtitles more, regardless of the language of the subtitles, when they did not understand the source language. Mean fixation duration also increased from standard to interlingual to reversed subtitles. Considering the eye movement data on the image, they found a gradual decrease of time spent on the image, going from the no subtitle condition to the English sound-Dutch subtitle condition, on to the Dutch sound-Dutch subtitles, with Dutch sound-English subtitles leading to the least amount of time spent on the image. This is in line with previous findings: subtitles always take up some of the time spent on the image, and the time spent on the subtitles increases as the need for the subtitles to understand the content increases (see also Liao et al., 2022, and Liao et al., 2021, discussed later in this section).

In Chan (2020), subtitle language was examined in a lecture context. Students were found to look at the L2, intralingual subtitles considerably more than the L1, interlingual subtitles. Of the total viewing time, they were found to spend about 24% of the viewing time on L1 subtitles, whereas they would spend almost double (45%) of the total viewing time on the L2 subtitles in the same lecture. They did not find an effect of subtitle language or presence on comprehension or cognitive load.

Liao et al. (2020) compared effects of L1, L2 and bilingual (showing both L1 and L2) subtitles on L2 educational video. While they found no differences between the subtitle conditions with regard to cognitive load or comprehension, they did find the viewers' visual attention to be much more stable when the subtitles were in their native language. This, once again, underlines the importance of language for subtitle processing, even if there is no immediate effect on comprehension or self-reported cognitive load.

Alongside the language of the subtitles, there is a second frequently studied topic, namely the subtitle characteristics mentioned in Section 2.2 of this chapter. We already discussed subtitle typology and how slight changes in these characteristics of subtitles might affect the effects of subtitles. Editing is one of those characteristics. In 2012, Ghia examined the effect of translation strategies on the reading of subtitles, comparing literal and non-literal translated subtitles. No difference was found regarding regression, but the number of fixations and number of deflections (i.e., revisits to the subtitle after prior reading) did differ significantly. Non-literal translations lead to more deflections and fixations, and especially non-literal content words were fixated more frequently.

Chan et al. (2019) studied the effect of automatically generated and corrected subtitles in educational video on comprehension and self-reported cognitive load. They found no differences between the two subtitle conditions or when either was compared to having no subtitles on screen. They attribute this finding to the high and variable subtitle speeds of the two tracks. Because the study did not employ eye tracking, nothing could be said about the reading of the subtitles.

Other characteristics also play a role. Lång et al. (2021) showed that the proportional reading time (time based on the time the subtitle is visible) changed depending on the length of the subtitle in characters and the duration the subtitle was on screen. Subtitles with more characters were generally the recipients of proportionally longer dwell times. This finding adds to Section 2.2 of this chapter and again illustrates how slight changes can alter viewing.

A third recurring topic is the effect of subtitles on comprehension and retention. The results seem to be pointing in one direction. A fair share of studies find positive effects of subtitles on brand recall (e.g., Brasel & Gips, 2014), movie comprehension (e.g., Bairstow, 2012) and, most importantly for this thesis, potential comprehension and retention benefits in an educational context (e.g., Kruger et al., 2014; Kruger & Steyn, 2014; Vulchanova et al., 2015). It is important to know that this is not a consistent find. Some studies (e.g., Chan et al., 2019; van der Zee et al., 2017) find no benefits of the presence of subtitles in education. One important factor might be the fact that, according to Kruger and Steyn (2014), the subtitles have to be read properly before any benefit can come from them.

If reading is vital for subtitles to affect comprehension and cognitive load, it should be considered when related research is conducted. However, how exactly subtitles are read and processed is still relatively unknown. Perego et al. (2010) conducted one of the first studies on cognitive processing of subtitles, employing eye tracking and word and scene recognition tasks. Even when the viewer spent less than 40% of the viewing time on the image, the performance on word and scene recognition tasks remained high. Because performance remained high and no trade-off was found between word and scene recognition performance, their study implied no disadvantage of subtitles and a clear effectiveness of subtitle processing in viewers.

Since then, a number of attempts have been made to investigate the cognitive processing of subtitles more accurately (e.g., Chan et al., 2019; Kruger et al., 2013, 2014; Kruger et al., 2022; Liao et al., 2022; Liao et al., 2021; Perego et al., 2016). Two studies that are of particular interest for this thesis (predominantly for Paper 4 in Chapter 5) are Liao et al. (2021) and Liao et al. (2022). Both studies revolve around the multimodal integrated language framework, a cognitive theory that conceptualizes the processing of subtitled audiovisual content (see Section 3.3.1 of this chapter for more information). The first study (Liao et al., 2021) found viewers to spend less time on the subtitles when other visual material was present. They also found that as subtitle presentation speed increased, viewers skipped more words at the end of sentences, had fewer crossovers between the video and the subtitles and spent proportionally more time on the subtitles. Essentially, reading changed considerably with the introduction of video and a more demanding (faster) version of the subtitles. The second study (Liao et al., 2022) also found time spent on subtitles to decrease when concurrent video was introduced. Furthermore, they found viewers to rely less on subtitles, i.e., fewer fixations, shorter fixation durations, longer saccades) when L2 audio was introduced. Their subtitle reliance reduced even further when the audio was changed to L1. These two studies further highlight how changes to the multimodal environment, e.g., presence of audio or video, can have a significant impact on how and how often subtitles are read. A follow-up question that has remained largely unanswered is how the complexity of the concurrent material then impacts subtitle reading.

This question highlights the need for more research into the reading and processing of subtitles. Furthermore, methodologies are, at times, very different from study to study. This can in turn lead to inconsistent results, which has been the case when it comes to the effects of subtitles and whether they are beneficial or not. More accurate measures and newer systems allow us to conduct more through research into reading and processing of subtitles, which in turn allows us to develop sound methodological frameworks and strong foundations for future research into the cognitive effects of subtitles.

2.4 Research methodology

The field of research, specifically experimental AVT research, is still very much in its early stages. As mentioned in the previous section, methodologies within AVT research differ considerably from study to study. So far, there have been a number of methodology papers that attempted to unify the field by recommending certain approaches and highlighting good practices for experimental AVT research (e.g., Doherty & Kruger, 2018; Kruger, 2016; Kruger & Doherty, 2016; Kruger et al., 2016; Orero et al., 2018). Research design is arguably the biggest hurdle in AVT research and, more specifically, two components of design, namely the overall design of a study and its level of control.

Between-group studies are common in AVT research, both experimental and quasi-experimental. While it can yield strong results, participant variability can potentially skew the results. In order to mitigate these effects, large sample sizes should be used (Mellinger & Hanson, 2017). Limited sample sizes are, however, a frequently occurring problem in AVT research (Díaz Cintas, 2020). It could therefore be interesting to employ within-subject designs (i.e., repeatedly testing the same participants in multiple conditions) or a mix of the two designs. Repeated testing generally implies the use of two or more sets of AVT material. This AVT material should be controlled to avoid risking the material skewing the data.

Control, however, is not an easy feat in AVT research. Because of the highly dynamic and multimodal environment AVT occurs in, it is virtually impossible to design a study that is both ecologically valid and controls for all the confounding variables. Highly controlled experimental environments are therefore rare in AVT research, while limited control in more ecologically valid studies come with a risk of limited generalizability and replicability. Thus, good research practice requires careful participant sampling, use of appropriate measures and meticulous selection and preparation of materials. Some of the previously mentioned methodology papers recommend which characteristics of AVT materials should be mentioned. While this is regularly done by AVT researchers, generally few words are spent on how and why material was selected and how it was prepared. This is a key component for validity and replicability of research. It is especially relevant when two sets of materials (for the within-subject designs that were recommended earlier, for example) are used. In that case, statistical mixed modelling can account for some of the variance of the material, though a certain level of comparability of the material is still key. In an attempt to unify the field and increase overall validity and replicability, this thesis also presents an approach for preparing comparable material for quasi-experimental and experimental AVT research in Papers 1 and 2 in Chapter 4.

3 Multimodal processing

In the previous sections, we first shed some light on the cognitive process of reading and then continued with the reading of subtitles and the effects of subtitles. As was made apparent, subtitles are, under normal circumstances, accompanied by both sound and images. In other words, subtitles are generally part of a multimodal environment. Because of this multimodality, numerous processes occur concurrently (e.g., reading, viewing scenes, listening) and, consequently, affect each other. This section focuses on how processing proceeds in these multimodal environments. The first key topic discussed and one of the most relevant theories for multimodal processing is the cognitive load theory (CLT). After the discussion of this theory, the section addresses some major multimodal processing frameworks and their implications. Then, it discusses ways to measure cognitive load and cognition. It concludes with practical research on instructional design, providing insight into how different styles of lectures and teaching can impact learning.

3.1 Cognitive load theory

The cognitive load theory formulated by John Sweller and colleagues (Sweller, 1988; Sweller et al., 2011; Sweller et al., 1998, 2019) explains how information processing can be affected by a task and consequently affect a person's ability to learn and gain knowledge. It is a key theory to understand multimodal information processing. However, before exploring the theory itself, some information on human cognitive architecture, types of knowledge distinguished by the theory and how to acquire said knowledge is required.

3.1.1 Human cognitive architecture

Human cognitive architecture refers to the components that constitute human cognition and how they interact. The emphasis here tends to lie on the short-term working memory and the long-term memory and how they relate to each other. Another vital aspect to understand human cognitive architecture is knowledge itself. Regarding knowledge, an important distinction should be made between biologically primary knowledge and biologically secondary knowledge, as suggested by Geary (2008) and Geary and Berch (2016).

Geary (2008) explains how, from an evolutionary perspective, humans have evolved to acquire particular knowledge without the need for extra instruction (e.g., speaking and listening), which he classifies as biologically primary knowledge. He assumes this knowledge to be modular, which means that each skill associated with biologically primary knowledge has its own cognitive processes with little to no relation to the cognitive processes for other skills. These skills have evolved in their own time and, while some degree of plasticity is expected, change is limited. The majority of primary skills rely on domain-general knowledge rather than domain-specific knowledge, e.g., basic problem-solving, thinking, learning (Tricot & Sweller, 2013). They cannot be taught and are acquired automatically and unconsciously. However, applying these skills in specific domains has to be actively learned. The knowledge that is acquired to do so is called biologically secondary knowledge.

Biologically secondary knowledge and skills are acquired because they are deemed necessary by society. Two basic examples are reading and writing. Both skills have been used by humans for millennia already, yet explicit instruction is required to be able to read and write. Secondary knowledge is not modular and the processes of learning knowledge in different domains therefore tend to share a common ground (Sweller et al., 2019).

To make clear the distinction between primary and secondary knowledge and how they are closely related, we will explain the skill of reading. Humans learn to see and interpret the world with their eyes on their own. This does not have to be taught or explained and the 'skill' to be able to see is therefore associated with biologically primary knowledge. However, when we are offered a piece of paper with words on it, we are initially unable to comprehend the meaning of it and consider it only as a piece of paper with scribbles on it. We require someone to teach us the meaning of the scribbles explicitly. The first lesson would revolve around recognizing the scribbled as letters/symbols. The second lesson would teach us to see them as meaningful words. Reading is therefore a skill associated with biologically secondary knowledge. By learning this secondary knowledge, we can apply our primary knowledge/primary skill to do something more.

Sweller et al. (2011) underline the importance of this classification for instructional design. As biologically primary knowledge is acquired naturally, any attempt to teach these skills may be futile (Tricot & Sweller, 2013). Biologically secondary knowledge, on the other hand, needs to be taught and cannot be acquired by immersion only. Schools should therefore be designed to optimize the process of teaching secondary knowledge.

To acquire secondary knowledge, it first needs to be processed and stored by human cognition, i.e., skills related to biologically primary knowledge. Two key parts for processing and storing information are the short-term working memory and the long-term memory. In order to better understand the process of handling and storing information in these two key parts of human cognition, Sweller et al. (2019) list five relevant principles of cognitive architecture.

Firstly, there is the information store principle (Sweller et al., 2019). Over the course of life, humans are exposed to unfathomable amounts of information. Consequently, human cognition demands a place to store this information. This place is the long-term memory. The biologically primary function related to this is our innate ability to organize and store the information we receive in the long-term memory.

The second principle is the borrowing and reorganizing principle (Sweller et al., 2019). Humans do not live in isolation and the information stored in our mind mostly comes from others. The ability to share and receive information from others is a biologically primary skill.

If no one provides us with information, we are able to generate information ourselves. This relates to the randomness as genesis principle (Sweller et al., 2019). It is a biologically primary skill of humans to be able to solve problems. However, when no one is there to tell us how to solve a problem, all that can be done is randomly picking an approach and trying to solve a problem, testing the approach for its effectiveness in the process. Effective approaches can be stored and shared with others later.

The narrow limits of change principle relates to the limitations of the working memory (Sweller et al., 2019). The working memory takes care of all that we actively think, see, do, etc. It is basically our consciousness (Sweller et al., 1998). It is, however, subject to certain limitations. For one, the capacity of the working memory is narrow. As is stated in the famously titled paper of Miller (1956), the capacity of the working memory is generally limited to “The magical number seven, plus or minus two”. This is an established limit based on recall tests. Depending on whether the test processes allow rehearsal and grouping of items, and exclude any other distractions, the limit may vary a little bit. For young adults, recent work suggests working memory can hold about 3-5 items or chunks of information (Cowan, 2010). This capacity is fixed for each individual.

The final principle is the environmental organizing and linking principle (Sweller et al., 2019). As opposed to the limited capacity of the working memory, long-term memory appears to be limitless. After novel information is processed by the working memory, the information is organized, embedded and stored with knowledge already stored in the long-term memory. Once it is stored in long-term memory, it can be actively and passively called upon, which in turn leads to appropriate action being taken in known situations. The ability to link new information with already stored information is a biologically primary skill and can therefore not be taught.

3.1.2 Cognitive load theory

As was mentioned before, the cognitive load theory explains how information processing can be affected by a task and consequently affect a person’s ability to learn and gain knowledge. The theory dates back to the late 1980s (Sweller, 1988). Over the years, numerous empirical studies explored human cognition further and the cognitive load theory was adapted and improved in 1998 (Sweller et al., 1998) and 2019 (Sweller et al., 2019). This eventually made it into a well-known and solid theory explaining the relations between cognitive load (the core concept of the theory to be explained in the next paragraphs), instructional design and learning.

The cognitive load theory defines learning as 'schema acquisition'. A schema is "a cognitive construct that permits us to classify multiple elements of information into a single element of information according to the manner in which the multiple elements are used" (Sweller et al., 2011, p. 22). The concept of schema acquisition is based on the idea that new information is processed in the working memory and collected in a schema. Learning occurs when this new schema is formed effectively and transferred to the long-term memory for storage. While new schemas require the learner to make an effort and process the new information consciously, practice and repetition makes the processing of schemas less and less conscious. Eventually, the knowledge of a schema can be accessed and applied effortlessly, and schema automation occurs. While learning new information, the vast number of schemas already stored in the long-term memory can also be accessed and applied to make the processing of the new information easier. In other words, prior knowledge of a subject can reduce the processing load for new information on the same subject and facilitate learning. Prior knowledge thus has an influence on schema acquisition. This is, however, not the only influence. The capacity of the working memory, for example, also varies for each individual. The cognitive load theory focuses on yet another influence, namely instructional design. What new information is taught and how it is presented is key for maximizing knowledge transfer.

To substantiate the effects of instructional design on learning, the cognitive load theory uses the concept of cognitive load. Cognitive load can be considered "a multidimensional construct that represents the load that performing a particular task imposes on the cognitive system of a particular learner" (Paas & van Merriënboer, 1994, p. 1). The theory distinguishes three types of cognitive load: intrinsic load, extraneous load and germane load.

The first type is intrinsic load, which consists of the mental load imposed by the material itself, "the intrinsic nature of the information" (Sweller et al., 2011, p. 57). As intrinsic load essentially originates from what the learner is expected to learn or comprehend, it relies heavily on the learner's expertise and prior knowledge. This conforms with schema acquisition (Paas et al., 2003). It also ties in with element interactivity. Sweller et al. (2019) illustrates this as follows: For someone proficient in English, each word in a text constitutes a single element of information. For a learner, however, some words might be unknown and in that case, each letter is an element of information and they interact with each other to form a word. The higher element interactivity for the learner might make the learner experience the task as more complex and straining than it would for an experienced reader. Intrinsic load can therefore not really be influenced by the teacher, except indirectly. What can be done, however, is lowering the level for a learner and gradually building up their knowledge (scaffolding) until the higher-difficulty tasks require less mental effort as well.

All information that needs to be processed and that does not contribute to the learning goal itself creates extraneous load, the second type of cognitive load. Extraneous load is the load imposed by "the manner in which information is presented or the activities in which learners must engage" (Sweller et al., 2011, p. 57). This can be caused by, for example, external factors, e.g., outside noise in a classroom, music in the background during study, etc. or the design of a course, e.g., supporting/distracting PowerPoint slides. Extraneous load can thus be changed, ideally minimized, by the instructor (Sweller et al., 2019). Element interactivity again plays a role, as effective instruction with minimal extraneous load can decrease element interactivity.

The last type of cognitive load is germane load. Germane load is the load that is required to learn (Sweller et al., 2019). Germane load does not directly contribute to the total load, but instead takes care of the distribution of working memory resources from extraneous to intrinsic information. With the limited capacity of the working memory in mind, if fewer resources are spent on extraneous load, this distribution is more effective and learning can be improved. The line between intrinsic load and germane load, however, is thin. Germane load is thus assumed to have a mere distributive function, not causing its own load (Sweller et al., 2019).

Sweller et al. (2011) suggest that intrinsic cognitive load and extraneous cognitive load are additive. The total cognitive load imposed by a certain task or material thus consists of both loads added together. For learning to take place, the imposed total load should not exceed the total capacity of the working memory. If the total load does exceed the working memory capacity, cognitive overload will occur and the learner will not (fully) succeed in processing the information. The reason for this is that there are insufficient resources left to deal with and process the relevant information, i.e., the intrinsic load. To optimize learning, instructional design must be optimized to reduce extraneous load and maximize the working memory resources that can be devoted to learning and dealing with intrinsic load. While this may initially seem to be rather straightforward, the effects and interactions that can be part of instruction itself are numerous. The next part of this thesis will shed some light on these effects.

3.1.3 Instructional effects

In the past decades, a large number of studies explored the cognitive load theory and the different effects instructional methods could have on cognitive load. This eventually created a body of instructional effects that are key to consider when optimizing the learning experience based on the cognitive load theory. This section lists the instructional effects that are associated with the cognitive load theory as discussed in the work of Sweller and colleagues (Sweller et al., 1998, 2019). While all effects will be mentioned, only four that are considered relevant for this thesis will be discussed in detail, namely the split-attention effect, the redundancy effect, the modality effect and the transient information effect. Mention of these effects will return in Section 3.2 of this chapter. For more information on the other effects, see Sweller et al. (1998) and Sweller et al. (2019).

Before we discuss the four instructional effects relevant for this thesis, we wish to briefly mention the other instructional effects listed in the cognitive load theory. The theory lists four effects that relate to the design of a problem-solving task: (1) the goal-free effect; (2) the worked example effect; (3) the completion problem effect; and (4) the variability effect. The theory also highlights that the expertise of the audience plays a significant role in the experienced cognitive load. It lists three effects that relate to expertise: (1) the element interactivity effect; (2) the expertise reversal effect; and (3) the guidance-fading effect. Beyond expertise and design of a task, the learning approach of learners is also key for cognitive load. The theory sums up three effects that relate to the approach of learning: (1) the self-management effect; (2) the self-explanation effect; and (3) the imagination effect. If the content complexity of the material exceeds the total working memory capacity of a learner, two effects listed by the cognitive load theory might be of help: (1) the isolated elements effect; and (2) the collective working memory effect. Lastly, the cognitive load theory also mentions one last effect related to learning movement tasks (e.g., tying knots, folding paper), which is the human movement effect.

The effects mentioned above are pertinent to the cognitive load theory, but less relevant for the present thesis. There are, however, four effects that are key to understanding cognitive load and how it may be affected in the research conducted here. These are the split-attention effect, the redundancy effect, the modality effect and the transient information effect. The following paragraphs will discuss these in detail.

The first effect that will be discussed here is the split-attention effect. Split attention occurs when learners have to divide their attention between multiple sources of information that are essential to understand content (Ayres & Sweller, 2014). These sources of information can be separated in space (spatial) or in time (temporal). Because the learner is required to split their attention, extraneous cognitive load is increased and a negative impact on knowledge construction can be expected. To avoid the generation of unnecessary load, an attempt should be made to present the different sources of information in an integrated format to minimize extraneous load. There have been numerous studies on the split-attention effect (see Ayres and Sweller, 2014, for an overview), but here we wish to focus on only a select few that provide some key insights relevant for this thesis.

Moreno and Mayer (1999) conducted a study to examine the effect of spatial contiguity and modality on learning. One hundred and thirty-two college students watched an instructional animated video on meteorology in one of three conditions, with on-screen text close to the animation, with on-screen text below the animation or with a concurrent narration and no on-screen text. They found that in terms of spatial contiguity, learning was impaired when the on-screen text was separated from the animation. They also found that the animation with concurrent narration outperformed the one with on-screen text. This is due to the modality effect, which will be explained in one of the next paragraphs. In a second experiment, they tested 127 college students in six conditions, the animated video with on-screen text shown before, at the same time or after the animation and the animated video with narration played before, at the same time or after the animation. Here they found no temporal effect for the narrated animation, but they did find that students who watched the animation with concurrent on-screen text performed worse than those that watched it with the text shown before or after the animation. This is a clear example of the split-attention effect, as learners have to divide their cognitive resources between two sources of information at the same time. Lastly, the narrated conditions once again outperformed the on-screen text conditions. These findings have significant implications for research on subtitled educational video where subtitles essentially provide a written transcript of the narration presented concurrently with the spoken narration and the content of the video (talking head, slides, etc.). If key visual material is shown and subtitles are included, learning could be impaired as learners would have to divide their attention between the multiple sources of information.

The second effect is the redundancy effect (Kalyuga & Sweller, 2014). As explained above, learning is impaired when learners have to split their attention to attend to multiple key sources of information. However, when one of these sources is essentially a copy of the other source (e.g., verbatim subtitles copying the narration) or the information is unnecessarily elaborated, it can be considered redundant. This redundant information that needs to be processed by the learner generates additional load and may thus impair learning. This has important implications for subtitled educational video, in which subtitles will generally be a copy of the narration and can therefore be considered redundant. While there is relatively clear evidence of the redundancy effect in most cases, research on written/spoken text redundancy seems divided. Several studies demonstrate visual information with concurrent narration to yield better learning results compared to visual information and on-screen text (modality effect), but also show this advantage to disappear when both narration and on-screen text are provided (Jamet & Le Bohec, 2007; Kalyuga et al., 2000, 2004; Mayer et al., 2001). However, some studies fail to find a redundancy effect (Craig et al., 2002; Moreno & Mayer, 2002). Moreover, some studies from the field of audiovisual translation specifically, find either no redundancy effect of subtitles or even a reversed effect with subtitles decreasing cognitive load (e.g., Kruger et al., 2013, 2014). This may be because the presence of a redundancy effect seems to depend on a number of factors, such as the length of the text and the complexity of the material (Kalyuga, 2012; Kalyuga & Sweller, 2014).

The third effect, i.e., the modality effect, flows from the redundancy effect. The modality effect states that when information is presented in two modalities, both visual and auditory, knowledge construction is more effective than when presented in only one modality (Low &

Sweller, 2014). The modality effect is related to the dual-coding theory (Paivio, 1986), which assumes that working memory is divided into two cognitive subsystems. One system is specialized in dealing with nonverbal information/imagery, the other is specialized in dealing with verbal information/language. These channels both have a limited capacity, which implies that when only one channel is used, the capacity is limited to that one channel and the capacity of the other channel is not called upon. The modality effect can only occur when split-attention is required, i.e., both sources of information are required for learners to understand the content. If this is not the case, the redundancy effect is expected to generate unnecessary cognitive load instead. While numerous studies confirm the modality effect (e.g., Harskamp et al., 2007; Leahy & Sweller, 2015), research is yet to reach a consensus on whether and when subtitles might decrease cognitive load by presenting information in a different modality than the narration, despite it being verbal, or increase cognitive load as it adds redundant information. The language of the material and the fluency levels of the audience, of course, play a vital role in whether the information is redundant.

The last effect that is highly relevant for the present thesis is the transient information effect. The transient information effect states that cognitive load is increased when information is transient, that is, presented to learners but disappearing again after a moment, e.g., instructional videos (Leahy & Sweller, 2011). Upon further investigation, the transient information effect and its interaction with the modality effect seems to strongly depend on the length of the information (Leahy & Sweller, 2015). Presenting transient information both auditorily and visually is beneficial when the information is short, i.e., only a limited amount of information has to be held in the working memory to understand the transient content. When the chunk of information increases in size, the benefit of using multiple modalities disappears and eventually the modality effect reverses. This has some important implications for the present thesis. Firstly, subtitles can be considered transient content. They are generally short pieces of information and are accompanied by auditory information, meaning they would benefit from the transient information effect. Secondly, in lectures that contain PowerPoint slides or other information on screen, the transient information effect can play a significant role in the advantage such designs bring. In sum, the transient information effect is highly important for developing effective online lectures with subtitles, the central topic of this thesis.

It is nearly impossible to consider one of the aforementioned effects without taking into account the others. The interactions are plentiful and have to be considered when designing effective instructional material. While these effects are specifically mentioned in the cognitive load theory, more effects and principles have been revealed in research. These will be discussed as part of their respective multimodal frameworks in Section 3.2 of this chapter.

3.1.4 Criticism

The cognitive load theory (Sweller, 1988; Sweller et al., 2011; Sweller et al., 1998, 2019) has without a doubt had a significant impact on educational research. The discussion above introduces the theory and highlights relevant aspects of the theory for the present thesis. This means that the discussion is limited, as the theory and all accompanied research could be the topic of a full manuscript in itself. Nevertheless, the present thesis would not be complete without briefly mentioning that the cognitive load theory has also received criticism. One of the most important points is that some of the fundamental concepts or assumptions of the cognitive load theory are nearly impossible to contradict (Gerjets et al., 2009). Every outcome fits within the theory as it can be explained as a consequence of either an increased intrinsic, extraneous or germane load (de Jong, 2010). Furthermore, the distinction between these different loads also gives rise to certain issues. The totality of all aspects of instructional design lead to meaningful learning. The cognitive load theory fails to consider that, for example, misalignment of aspects such as task difficulty and expertise (intrinsic load factors) can also lead to an increase in extraneous load (Schnotz & Kürschner, 2007). As is also shown in our own research (Chapter 6), an instructional environment is highly interactive and complex. The cognitive load theory and its principles can provide some guidance as to what to expect, but cannot provide a definitive answer as to how specific instructional design will impact a certain audience.

3.2 Multimodal processing frameworks

Because of the complexity of human cognitive architecture, it is not an easy task to create a model of all that happens when processing multimodal environments. Nevertheless, some researchers have developed models and frameworks to illustrate and predict multimodal processing. For the sake of clarity and brevity, only three processing models/frameworks will be discussed in this section, the Multicomponent Working Memory of Baddeley and Hitch (1974), the Cognitive Theory of Multimedia Learning of Mayer and Pilegard (2014) and the Multimodal Integrated-Language framework (Liao et al., 2021).

3.2.1 Multicomponent Working Memory (M-WM)

The Multicomponent Working Memory (Baddeley & Hitch, 1974) is one of the earliest frameworks of multimodal processing. Over the years, components were added and the framework changed in response to criticism and controversies around some of its theories (Baddeley, 2000, 2012). Even though the majority of this thesis will focus on the two more recent frameworks that will be mentioned below, the M-WM is seminal and could therefore not be omitted from this thesis.

The M-WM attempts to explain how the working memory, tasked with both temporary storage and processing of information, works and interacts with long-term memory. The model is evidence-based and is kept as simple as possible to avoid making inaccurate assumptions. In its early stages, it assumed the working memory consisted of three different components, the phonological loop, the visuo-spatial sketchpad and the central executive (see Figure 6 below for the early M-WM).

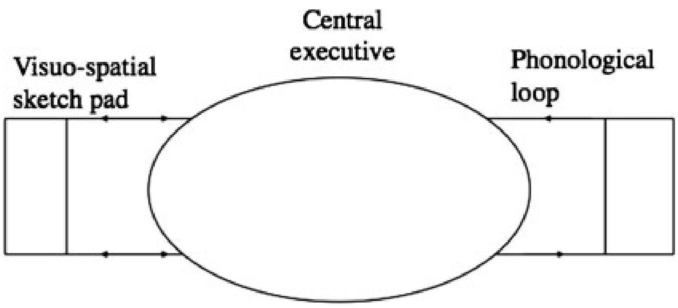


Figure 6: The early Multicomponent Working Memory (Baddeley & Hitch, 1974)

The phonological loop is probably the most researched component of the M-WM, mostly because it is the most accessible and easiest component to be studied. It is considered a short-term storage for information by vocal or subvocal rehearsal (Baddeley, 2012). The concept of the phonological loop came into existence after Baddeley's research on the similarity effect. If participants had to repeat a sequence of words, one sequence being acoustically similar words (e.g., cat, rat, bat) and one sequence being acoustically dissimilar words (e.g., big, huge, cow), they were able to correctly repeat the similar sequence significantly more when tested shortly after having heard the sequence. This effect reversed when they had to repeat the sequence after a delay. This led to the assumption that there was a short-term phonological loop, capable of short-term storage and benefiting from similarly sounding words (i.e., phonological similarity), and a separate longer-term semantic storage. This short-term storage by rehearsal was shown to be limited in time, not in number of items. In order for written verbal information to be registered, it needs to be subvocalized. However, as was shown in further research, auditory verbal information immediately gains access to the phonological loop and can thus interfere with the subvocalization, leading to a drop in performance. In later versions of the M-WM, long-term memory was explored as well. A direct link was revealed between the phonological loop and long-term memory as the phonological loop was shown to play a significant role in early language and vocabulary acquisition.

Another component of the M-WM is the visuo-spatial sketchpad. It is responsible for the short-term retention of visual and spatial information. In other words, it allows us to briefly remember an image, a location, directions, etc. Visual and spatial information was shown to be processed differently than verbal information, as verbal suppression had a more negative effect when participants had to rehearse verbal information as opposed to visual information. The capacity of the visuo-spatial sketchpad was also revealed to be limited, since task performance declined as more information had to be retained (for an overview on research, see Baddeley, 2012).

The last component is the central executive. It was initially assumed to be able to do all things the other two components could not do. Baddeley and Hitch (1974) called it a homunculus, a little man in the head. The model was frequently criticized for taking this approach. However, Baddeley (2012) later explains that the homunculus approach was aimed to show how much still needed to be explored and explained, and was not meant as an explanation itself. Later the central executive became more concrete. Based on research, the central executive would have to be capable of doing four things: (1) to focus attention; (2) to divide attention between two targets; (3) to switch between tasks and thus control attention; and (4) to interact with the long-term memory to store information. The central executive was assumed to not have any storage capacity, but research had shown that humans were capable of integrating information from the other two components (i.e., the phonological loop and the visuo-spatial sketchpad) and holding on to it. This undermined the assumption that the central executive had no storage capabilities and thus a fourth component was added to the model, the episodic buffer.

The episodic buffer explained the working memory's capacity to be able to hold integrated chunks of information containing both phonological, visual and spatial information. It is a buffer store between the various components of the working memory but also between the working memory and the long-term memory. The addition of the episodic buffer made the M-WM what it is today (Figure 7). However, there are still a number of issues with the M-WM, such as the limited links between the working memory and the long-term memory. See Baddeley (2012) for more information about Baddeley's currently adjusted view of the working memory and long-term memory, and a speculative model of the flow of information for the working memory.

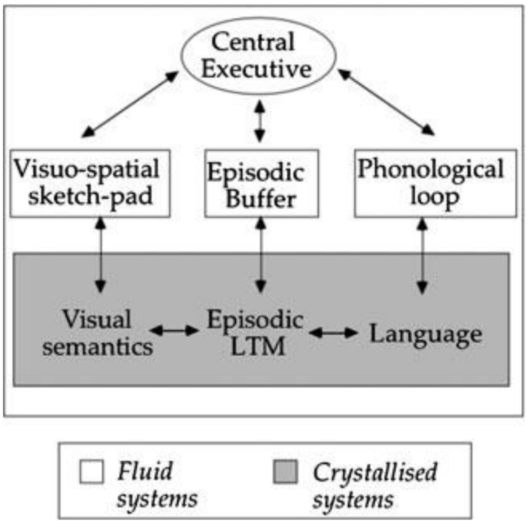


Figure 7: The updated Multicomponent Working Memory (Baddeley, 2000)

3.2.2 Cognitive theory of multimedia learning (CTML)

A second important framework of multimodal processing is Mayer's Cognitive Theory of Multimedia Learning (Mayer & Pilegard, 2014). It is a vital theory for this thesis as it can be used to explain the effects of subtitles on cognitive load, and consequently learning, in lectures. The theory is based on three main assumptions, namely the dual-channel assumption, the limited-capacity assumption and the active processing assumption.

The first assumption is the dual-channel assumption, a key feature of Paivio's dual-coding theory (1986). This states that each human has two channels to process information, one for visually/spatially represented material and one for auditorily/verbally represented material. This is crucial for multimedia learning as it implies that when a learner is confronted with visual information, this information would enter only one channel and be processed in this channel. It is, however, possible for a learner to convert the representation of the information to be processed in the other channel: For example, an image is initially processed in the visual channel, but an experienced viewer may be able to verbally describe the image in the mind, so it can be processed in the auditory channel.

The second assumption of the CTML is the limited-capacity assumption, an essential aspect of the cognitive load theory (Sweller et al., 1998, 2019). It states that only a limited amount of information can be processed in each channel at one time. The capacity of the working memory, and the information that can be held in memory, varies for each individual. As has already been mentioned in earlier parts of this thesis as well, on average five to seven chunks of information can be held in memory. If the amount of information presented exceeds the working memory capacity, the learner is forced to decide which information to pay attention to, how much attention should be paid to connecting the chunks of information and how much effort should be put into integrating the new information with previous knowledge.

The third and last assumption in the CTML is the active processing assumption, which plays a pivotal role in Mayer's earlier theory of active learning (Mayer & Moreno, 2003). The assumption claims that a learner needs to actively invest in cognitively processing new information to learn and construct a mental image of this information. In other words, a learner has to pay attention to be able to select and comprehend relevant material, organize the new information in coherent structures and integrate it with each other and relevant prior knowledge.

With these assumptions in mind, the CTML attempts to explain the stages of information processing. It distinguishes five cognitive processes as core components of active learning: Selecting words, selecting images, organizing words, organizing images, and integrating information. The processes of selecting, organizing and integrating information take place consecutively and constitute the bridges between three distinct memory stores. Figure 8 below shows the learning process as explained in the CTML. First, information is presented and copies of the incoming words and/or images are held in the sensory memory for a very brief period of time. The working memory selects relevant words and images, keeping in mind its limited capacity, and organizes these words and images in coherent cognitive structures. Then, the working memory integrates these cognitive structures with each other and with the relevant prior knowledge provided by the unlimited long-term memory. This constitutes information processing.

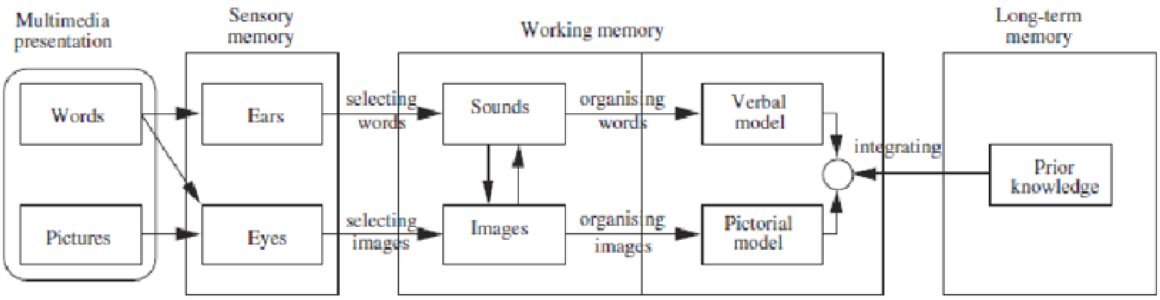


Figure 8: The Cognitive Theory of Multimedia Learning (Mayer & Pilegard, 2014)

As mentioned before, learning and processing requires the learner to actively put in cognitive effort. This implies that there are demands on the cognitive capacity of the learner. Similar to the cognitive load theory, Mayer (2014a) distinguishes three types of processing: (1) essential processing (i.e., processing of intrinsic load); (2) extraneous processing (i.e., processing of extraneous load); and (3) generative processing (i.e., allocation of germane resources).

Essential processing can be equated with the processing of intrinsic load from the cognitive load theory. It refers to the processing of essential material in the working memory. Looking at Figure 8, essential processing is involved with selecting words and images and organizing these words and images into coherent cognitive structures. Extraneous processing is analogous to the processing of extraneous load from the cognitive load theory. It constitutes the processing of material that is not related or relevant to the instructional goal. Examples of extraneous processing can be the processing of a particularly ill-designed lecture (the processing of the design itself) or simply background noise. Generative processing is similar to the process of allocating germane resources explained in the cognitive load theory. In Figure 8, generative processing is both the organizing of words and images into coherent cognitive structures as well as the integration of these structures with each other and prior knowledge from the long-term memory. It is key for actual learning and is closely linked to the active effort a learner puts into learning.

The cognitive capacity of a learner decides what the learner can or cannot process. What is or what needs to be processed is a sum of the three separate processing demands mentioned above. If the demand exceeds the total capacity, cognitive overload occurs. The CTML classifies the types of overload depending on what caused the overload. If the essential processing demand or intrinsic load is too high, essential overload occurs. An exceeding extraneous processing demand or extraneous load causes extraneous overload and a lack of generative processing effort constitutes a generative underutilization. To optimize instructional design, situations of overload or underutilization should be avoided. The CTML lists 15 principles that have a significant influence on distinct processing demands and should thus be considered for instructional design. Table 2 below shows each of these principles and what type of processing demand they affect. The following paragraphs will shed light on each of these principles separated by processing demand.

Table 2: CTML principles to manage processing (Mayer, 2014a)

Essential processing	Extraneous processing	Generative processing
- Modality - Segmenting - Pre-training	- Coherence - Signaling - Redundancy - Spatial contiguity - Temporal contiguity	- Multimedia - Personalization - Voice - Embodiment - Guided discovery - Self-explanation - Drawing

When attempting to manage the essential processing or intrinsic load in instructional design, three principles should be considered, namely the modality principle, the segmenting principle and the pre-training principle.

The modality principle states that presenting information in both the visual and the auditory modality is more effective in terms of learning than presenting the same information in only one of the modes (Low & Sweller, 2014). The modality principle is based on the limited-capacity assumption (borrowed from Paivio's dual-coding theory; see Paivio, 1986). The capacity of the working memory is limited, but if both channels of the working memory are addressed, one can effectively 'expand' the working memory capacity and minimize intrinsic load. For more information on the modality principle, see Section 3.1.3 Instructional effects.

The segmenting principle explains that in order to maximize learning profits, it is more efficient to segment the learning material into logical segments, ideally learner-paced (Mayer & Pilegard, 2014). When the learning material is too much for a learner to process, either because it is too much information or the learner is not familiar with the topic, it is not possible to reduce the amount of material as it is vital for the learning experience. Segmenting the material reduces the load and can allow the learner to first complete the required cognitive processes and integrate the new information before going on to the next segment.

A final method to manage essential processing is described in the pre-training principle. A learner might not be able to process certain learning material, but it is possible to equip the learner with the necessary knowledge to do so beforehand (Mayer & Pilegard, 2014). In a simple example: It might be difficult to teach someone how the complete engine of a car works, so to avoid cognitive overload, it would be better to first teach them the function of every component separately. Equipped with that knowledge, learners will be better equipped to tackle the essential processing load that comes with the more complex chunk of information.

While essential processing is harder to manage as it is inherent to the material, extraneous processing can be more easily managed. Instructional design should theoretically always be optimized to minimize necessary extraneous processing and leave more capacity for essential processing. The CTML lists five principles that are directly related to the reduction of extraneous processing and thus the optimization of instructional material, namely the coherence principle, the signaling principle, the redundancy principle, and the spatial and the temporal contiguity principle.

The coherence principle envelops what is written in the previous paragraph: Learners benefit more from multimedia messages if extraneous material is minimized (Mayer & Fiorella, 2014). To illustrate: A lecture is accompanied by a PowerPoint presentation and to avoid having too much white space on screen, the lecturer has included some related, but essentially irrelevant images on the PowerPoint slides. In such case, it would be better to exclude those images and reduce the content on the slides to the bare essentials.

The signaling principle, also called the cueing principle, describes that learning improves when the multimedia includes cues or signals that guide the learner's attention to what is important (van Gog, 2014). This effectively alleviates the load on the learner by assisting in the process of selecting information. Cueing is mostly done by highlighting, using arrows, underlining, or anything that makes the information visually salient.

The redundancy principle has already been mentioned in Section 3.1.3 of this chapter. It is a key principle for this thesis and states that in order for learning material to be optimized, all redundant information should be omitted (Kalyuga & Sweller, 2014). Examples of redundant information are overly elaborate information (in line with the coherence principle, it is better to be concise) or copying information from one stream to another stream of information. For more information on the redundancy principle, see Section 3.1.3 of this chapter.

The spatial contiguity principle offers another way to manage extraneous processing. It explains how learning is improved when related words and images are presented in close proximity on a screen or page rather than far away from each other (Mayer & Fiorella, 2014). The principle finds its origin in eye movement studies. It takes time and effort to move around the eyes across a screen. If relevant material is presented spatially close, the time and effort to find and move the eyes to the corresponding images or words is reduced. The effort, in this case, relates to extraneous processing.

The temporal contiguity is the last principle that explains how to reduce extraneous load. It adds a second dimension to the spatial contiguity principle by stating that learners benefit from corresponding information (e.g., a narration and animation) being presented simultaneously rather than consecutively (Mayer & Fiorella, 2014). The aforementioned modality principle already showed how learning benefits from presenting information in two modalities, visually and auditorily. If this information is presented at the same time, it enters both channels and is processed in the working memory at the same time. It therefore improves integration and removes the need to hold on to one piece of information before being able to integrate it with another piece of information.

The CTML provides seven principles that can be considered to foster generative processing. These seven principles are the multimedia principle, the personalization principle, the voice principle, the embodiment principle, the guided discovery principle, the self-explanation principle and the drawing principle. Mayer considers some of these principles to be more advanced as they relate more to what to consider when designing the instruction of large amounts of information (e.g., during long-term course design) rather than just a brief multimedia instruction. These are therefore also less relevant for the current thesis, but will still be mentioned as they are part of the framework of the CTML.

The multimedia principle is arguably the most vital principle for the CTML, as otherwise there would be no desire to use multimedia instruction and, consequently, the CTML in the first place. It states that learning from words and pictures is better than learning from words alone (Butcher, 2014). It should not be confused with the modality principle, which concerns the combination of visual and auditory information. The multimedia principle concerns words and pictures, which can both be offered visually as well.

The following three principles are key for this thesis since they concern the presence of the lecturer in multimedia instruction. The first one is the personalization principle, which demonstrates that learning is improved when verbal information is in an informal, conversational style rather than purely formal (Mayer, 2014b). This can be done by directly addressing the learner during an instructional video and avoiding third-person constructions. It actively engages the learner, which in turn increases the cognitive processing and results in improved learning.

The second principle concerning the presence of the lecturer is the voice principle. Much like the personalization principle, the voice principle aims to engage the learner more actively. It illustrates that learners benefit from words spoken in a human voice rather than a machine voice (Mayer, 2014b). To some extent, a human voice simulates a real-life social environment. As listening is the expected behavior in a social environment, the learner is more inclined to do so. Maybe when artificial intelligence has reached new heights in the future, the results will be different.

The last principle that involves the presence of the lecturer in multimedia instruction is the embodiment principle. Once more, it presents the benefits of simulating a normal social environment. It states that when an on-screen agent is shown on screen, learning improves when it exhibits humanlike behavior, e.g., gestures, eye contact, facial expressions (Mayer, 2014b). As these are essential features of human social interaction, they are necessary when simulating and stimulating social interaction and cognitive activation.

The last three principles that foster generative processes are a little more advanced. They require more extensive effort from the instructor to design appropriate and effective instructional material. Because they are less relevant for the current thesis, these principles and their definitions are only briefly discussed in this paragraph. The first principle is the guided discovery principle. It describes that for most students, it is more beneficial to design instructional material in such a way that they are guided to eventually discover or conclude the key parts of the content themselves instead of just receiving all material at once and having to figure it out by themselves. The self-explanation principle is similar. It states that it is better to encourage students to think about certain content which eventually leads to a deeper understanding, rather than revealing all details at once. The generative drawing principle is separated from these other two. It describes how students benefit from having to draw an image of the content they were required to learn. By having to think of ways to draw information, a deeper understanding of the content is often acquired.

There are a few more principles that relate to the CTML, namely the feedback principle, the multiple representation principle, the worked examples principle and the image principle. The feedback principle states that explanatory feedback works better than just providing corrective feedback (Johnson & Priest, 2014). The multiple representation principle describes how learning improves when leaners are exposed to a variety of representations (e.g., images, graphs, illustrations, animations, verbal discussions, narrations) concerning the same information rather than just one (Ainsworth, 2014). In the current age of digitalization, this has become more of an unavoidable fact of instruction rather than something that has to be actively considered in instruction. The worked examples principle was also mentioned in Section 3.1.3 of this chapter and describes how providing full solutions to be studied beforehand can aid in learning how to solve a similar problem alone (Sweller et al., 2019).

These first three principles are less relevant for the present thesis, unlike the last principle, the image principle. The image principle is that, in a multimedia presentation, having the speaker's image on screen does not necessarily lead to a better learning experience than when it is not on screen (Mayer, 2014b). Studies that have been conducted on the on-screen presence of the instructor, however, seem to be somewhat inconclusive. A decrease, no change and an increase in performance have all been found in previous studies. This will be discussed in more detail in Section 3.3 of this chapter.

3.2.1 Multimodal integrated-language framework (MIL)

The multimodal integrated-language framework (Liao et al., 2021) is possibly the most recent framework that attempts to explain the mental processes that take place during multimodal processing. It is particularly interesting for this thesis as it specifically focuses and was developed based on studies of subtitle reading. The framework draws on the E-Z Reader model, which has been discussed in Section 1.2.1 of this chapter.

In Figure 9, a schematic overview of the MIL framework can be found. As it assumes a serial attention allocation, it is impossible to attend to multiple features on screen simultaneously. However, once an item has been identified (i.e., feature binding) and its location determined (i.e., location indexing), it can be held in the working memory to both track the object during the video in peripheral vision and aid in processing new information. It aids in processing new information as an image on-screen is assumed to facilitate the identification of corresponding words/objects. Like in the example shown in Figure 9, if a polar bear is seen on screen, the word "polar bear" can be expected in the subtitles as well. The framework thus highlights the benefit of using multiple modalities to improve information processing. In addition, the ability to be able to engage in dual tasking (i.e., processing new words and tracking identified objects) was further confirmed as their research showed a consistent benefit to comprehension when video was presented on screen alongside subtitles, even at high subtitle presentation speeds (Liao et al., 2021).

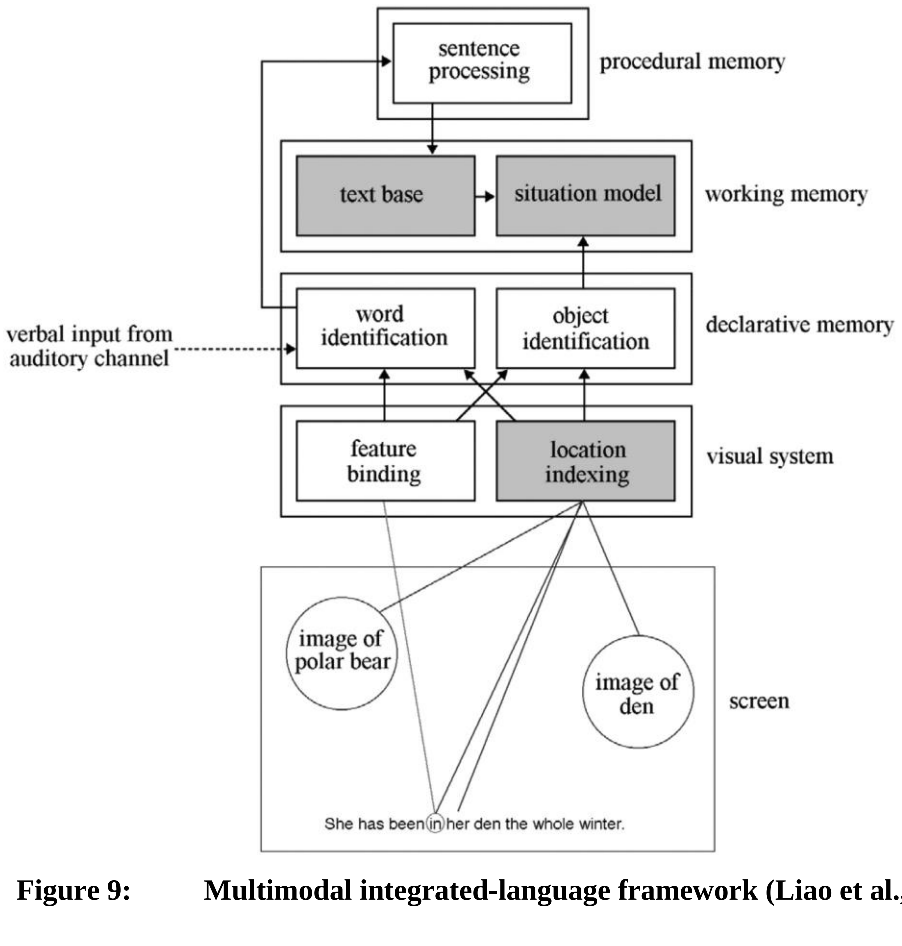


Figure 9: Multimodal integrated-language framework (Liao et al., 2021)

3.3 Measuring cognition

Considering the importance of cognitive load and understanding cognitive processing for instruction, their measurement has a significant role for research. In Section 1 of this chapter, it was already discussed how and why the eyes move during reading. This has also made clear that some movements and decisions are influenced by the complexity of the processing task. Measuring eye movements tells us something about ongoing cognitive processes and, consequently, can also tell us how certain features of the task at hand have an impact on the subject. Eye tracking is, however, only one of the many measures of cognitive processing. In research, four types of measures are being distinguished: subjective/self-report measures, performance measures, physiological measures and behavioral measures (F. Chen et al., 2016). The current section will discuss these measures, focusing mostly on self-report measures and physiological measures, specifically eye tracking.

3.3.1 Subjective/self-report measures

Subjective cognitive load measures require a participant to complete a questionnaire, reporting on their own perceived cognitive load. The main drawback of these measures is that they are offline, i.e., the participants only report on their cognitive load after the task was completed so they do not provide any insight into how cognitive load changed over the course of the task or as a result of specific elements of the task. Nevertheless, these measures have been shown to be reliable and, as learners were shown to be able to distinguish between the different types of cognitive load, can also shed light on the different components of cognitive load (Ayres, 2018). Additionally, they are easy to administer and only take up a short amount of time.

There are a large number of psychometric questionnaires available. Some are unidimensional, measuring total cognitive load only, and some multidimensional, distinguishing between different components of cognitive load (F. Chen et al., 2016). For this thesis, there are two multidimensional measures we wish to discuss in a little more detail, the NASA Task load Index (NASA-TLX) (Hart & Staveland, 1988) and the cognitive load scale from Leppink and van den Heuvel (2015).

The NASA-TLX (Figure 10) is a frequently used multidimensional questionnaire, which attempts to give a broad evaluation of cognitive load. It uses six 7-point scales with high, medium and low increments, meaning each scale has 21 gradations. The six scales explore six components, namely mental demand, physical demand, temporal demand (i.e., the pace of the task), performance (i.e., success in completing the task), effort (i.e., experienced difficulty to complete the task), and frustration (i.e., annoyance, irritation, stress, etc.). Each individual component has a weight attached to it; physical demand has no weight, so the NASA-TLX actually has only five scales. The rating on each scale combined with the weight provides a score. The mean of these scores then provides a mean workload score. While the NASA-TLX is a frequently used psychometric measure, it is unclear to what extent work load and cognitive load are the same (Leppink et al., 2013). Furthermore, because it uses the previously mentioned six scales, the scores cannot be tied to the standard components of cognitive load, i.e., intrinsic, extraneous or germane load.

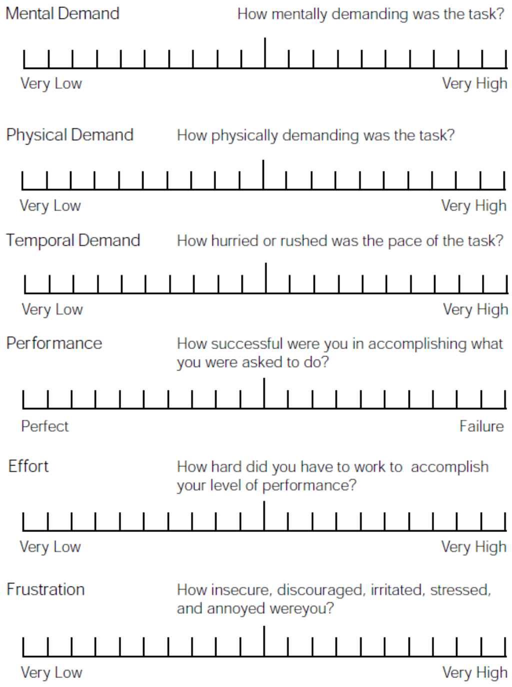


Figure 10: The NASA-TLX (Hart & Staveland, 1988)

An instrument that does focus on cognitive load specifically is the psychometric questionnaire from (Leppink & van den Heuvel, 2015). It was initially developed as a 10-item on a 10-point Likert scale questionnaire (Leppink et al., 2013). The first three items concerned intrinsic load, the following three concerned extraneous load and the last four items concerned germane load. However, after reconsideration, germane load was recategorized as a subtype of intrinsic load. The questionnaire (Figure 11) was adapted to then include 8-items on 10-point Likert scales, of which the first four concerned intrinsic load and the last four concerned extraneous load (Leppink & van den Heuvel, 2015). The advantage of this questionnaire over others is that it was shown to accurately measure cognitive load and distinguish between intrinsic and extraneous load. One limitation of this questionnaire based on personal experience is that questions are rather specific and can at times be misinterpreted. For example, questions about the explanations and instructions can in an experimental setting be interpreted as referring to possible instructions given at the start of an experiment or instructions given during an experiment. Adaptation for specific uses would be recommended.

All of the following eight questions refer to the activity that just finished. Please take your time to read each of the questions carefully and respond to each of the questions on the presented scale from 0 to 10, in which '0' indicates not at all the case and '10' indicates completely the case:

0 1 2 3 4 5 6 7 8 9 10

[1] The content of this activity was very complex

[2] The problem/s covered in this activity was/were very complex

[3] In this activity, very complex terms were mentioned

[4] I invested a very high mental effort in the complexity of this activity

[5] The explanations and instructions in this activity were very unclear

[6] The explanations and instructions in this activity were full of unclear language

[7] The explanations and instructions in this activity were, in terms of learning, very ineffective

[8] I invested a very high mental effort in unclear and ineffective explanations and instructions in this activity

Figure 11: Psychometric questionnaire from Leppink and van den Heuvel (2015)

3.3.2 Performance measures

Performance-based measures provide another way to measure cognitive load or processing during a task. As opposed to subjective measures, they give a more objective insight into the matter, that is, the learner does not have to self-evaluate or provide their personal opinion on the matter. Moreover, they are generally online, meaning they provide real-time data on cognitive load. Performance measures can, however, be intrusive.

Arguably the most frequently used performance measure is a dual-task measure (F. Chen et al., 2016). A dual-task measure requires a participant to perform a secondary task while performing a primary task, such as solving a problem. The idea behind the dual-task paradigm is the following: The secondary task requires cognitive resources, as does the primary task, which takes priority. If the cognitive load imposed by the primary task increases, less resources can be devoted to the secondary task, which tends to result in an observable decrease in performance on the secondary task. Performance in the secondary task can thus be an indication of cognitive load imposed by the primary task. While a dual-task measure can accurately reflect the cognitive load of a primary task, the mere introduction of a dual-task measure has considerable implications to the ecological validity of the experiment. In a real-life setting, one would likely try to avoid any distractions and solely focus on complex tasks, i.e., avoid multitasking.

3.3.3 Physiological measures

The human body constantly reacts to changes in cognitive load and acts on cognitive processing. Instead of introducing additional tasks, it is possible to observe physiological changes of a participant to examine cognitive processing. Some examples of physiological measures are heart rate and changes in heart rate, brain activity measured by electrocardiography (ECG) or electroencephalography (EEG), galvanic skin response (GSR) and, the main focus of this section, eye activity. The main advantage of physiological measures is that they also provide continuous data and are very sensitive to changes (F. Chen et al., 2016).

Section 1.2 of the current chapter made clear that fixations are an important part of reading. The E-Z Reader model ties fixations directly to lexical processing and the SWIFT model claims processing rate increases closer to the fixation, being at its maximum on the point of fixation itself. These models do not make these statements out of the blue but are well-founded in previous research, for example, the eye-mind hypothesis from Just and Carpenter (1980).

The eye-mind hypothesis states that the mind attends to where the eye is fixated (Just & Carpenter, 1980). Fixations are therefore assumed to be directly correlated to processing. However, as we have seen in the E-Z Reader model (Reichle et al., 2009) and SWIFT model (Engbert et al., 2005), this is not completely true. While a fixation does indicate processing, some information can still be retrieved from the parafoveal regions. This information can also be processed and thus the mind does not necessarily attend to the fixated information alone. Additionally, as exemplified in the E-Z Reader model and the study by Anderson et al. (2004), the mind can still attend to information after it was fixated or encoded. This, again, shows a discrepancy between the fixation and the ongoing cognitive processes.

Regardless, a fixation is still an important indicator of ongoing cognitive processing. A key aspect here is the duration of the fixation, which is dependent on a large number of factors. Research has consistently shown that longer fixations are generally associated with difficulty in processing. In reading, longer fixations are observed for more complex sentence structures (Frazier & Rayner, 1982), lower frequency or unpredictable words (see aforementioned models of eye movement), among others. In scene viewing, fixation durations increase as, for example, the scene becomes less discernible (Loftus et al., 1992) or has more colors (Henderson &

Hollingworth, 1998). Fixation durations can also change depending on the task that is required. Comparing fixation durations between different tasks or at different stages in a task can therefore provide insight into the experienced processing difficulty of the viewer. However, if it is used as a measure of processing difficulty, care should be taken to not compare too different viewing tasks. Fixation durations during reading, for example, tend to be shorter than fixation durations while viewing a scene without text (Kruger & Doherty, 2016). This does not necessarily mean that the viewing of the scene is more straining than the process of reading. Fixation durations simply change depending on the task as well.

Beyond fixation durations, there are many more basic eye tracking measures that can be used. A number of examples are: dwell time, i.e., the time a viewer has spent in a particular area of interest; fixation counts, i.e., the total number of fixations; the time to first fixation, i.e., the time it takes for a viewer to look at an area of interest from the moment the area of interest is shown on screen; or fixation sequence, i.e., the order in which viewers fixate on-screen components. These can be used to assess a wide variety of aspects, e.g., how salient on-screen agents may be, how long it takes for viewers to move their eyes or how long it takes for viewers to fully process something on screen. These lists are not exhaustive, as we merely wish to briefly show that eye tracking is a tool that produces rich data which can be employed for various aims and goals.

Another way of measuring cognitive processing, or more specifically processing difficulty, using eye movement data is pupillometry. Pupillometry considers the diameter of the pupil (pupil dilation) and how it changes. The first study into pupil dilation and processing can be traced back to the 1960s (Hess & Polt, 1964). In this study, it was shown that pupil dilation increases as a task becomes more demanding. Since then, the eye trackers used in research have improved considerably, becoming more accurate, easier to use and cheaper to acquire. Consequently, an increase of research into pupil dilation and processing demands or research using pupil dilation to measure demands can be observed since the dawn of the new millennium (for an overview of recent research, see van der Wel and van Steenbergen, 2018). The drawback of pupillometry is that the pupil is also sensitive to changes in lighting and the camera angle of the eye tracker (Duchowski et al., 2018). In an attempt to make pupillometry a viable option for measuring cognitive processing when watching video, Duchowski et al. (2018) developed the Index of Pupillary Activity (IPA).

The IPA is partially based on the Index of Cognitive Activity (ICA), a largely unpublished pupillometry measure used in eye trackers. It first removes data from the 200ms before and after a blink and then applies a formula to account for pupil unrest (also known as pupillary hippus), i.e., the constant fluctuations of a pupil diameter that do not relate to cognitive load, light changes, etc. For detailed information on the measure, see the paper of Duchowski et al. (2018). The IPA was, however, shown to still be restricted by the eye-to-camera angle (off-axis pupil distortion) when it was used in tasks that required the eye to move freely around the screen. It was thus adapted and improved by including low frequency and high frequency ratios of pupil oscillation. In essence, the ratio modulates pupil constriction and pupil dilation executed by the nervous system unrelated to task difficulty. The Low/High Index of Pupillary Activity (LHIPA) (Duchowski et al., 2020) was shown to be more reliable than the IPA, especially in tasks that required the eyes to move around the screen (Duchowski et al., 2020; Rodziewicz-Cybulska et al., 2022). These measures are still relatively new and require some further testing.

A last, interesting new eye movement measure of cognitive load is microsaccades. Earlier, in Section 1.1.2 of this chapter, it was already explained that the eyes move constantly because of tremor and drift and this slight movement is consequently corrected by microsaccades. In 2014, Siegenthaler et al. showed that the rate and magnitude of these microsaccades change depending on the task difficulty of a non-visual task. In 2018, K. Krejtz et al. also found a relation between microsaccadic magnitude and task difficulty. Microsaccadic magnitude was significantly higher for difficult tasks and lower for easier tasks. The rate of microsaccades did not yield any results in this case. Though task difficulty only explained 16% of the variance in microsaccadic magnitudes, the measure has somewhat promising results as it is not affected by ambient light or axis distortion like most pupillometric measures. Further testing, however, is required.

3.3.4 Behavioral measures

The last type of measures that can provide some insight into cognitive processing are behavioral measures. These can be gestures or facial expressions made during a task, mouse and keyboard activity on a computerized task, gaze patterns, etc. Ark et al. (1999), for example, asked six participants to sit behind a computer and act out emotions using their facial expressions for five minutes. While each participant acted out each emotion, physiological measures (galvanic skin response, heart rate and skin temperature) changed based on the emotion. Additionally, the somatic activity, measured by the movement of the mouse, also changed depending on the emotion that was being acted out. This is one example of research that shows how facial expressions or basic movements relate to emotion and possibly impact cognition, subsequently cognitive load. Behavioral measures can thus provide some insight into the process of task completion. The biggest drawback of these measures is that their accuracy is debatable and the observations leave room for interpretation.

3.4 Multimodal processing in online education

It is, of course, one thing to theorize and test certain effects that relate to cognition and education, but an entirely different matter to put it into practice in its highly active, multimodal environment. This section focuses on the research that has been conducted on instructional design, outside of the scope of cognition. Instructional design is "the systematic and reflective process of translating principles of learning and instruction into plans for instructional materials, activities, information resources, and evaluation" (Smith & Ragan, 2004, p. 4). Research on instructional design is often conducted in more ecologically valid environments and thus can provide some much-needed insight into how all processes and effects interact. Before delving further into this research, it may be necessary to illustrate what this ecologically valid teaching environment is.

The practice of teaching has changed considerably in the past decade. Two things in particular are of great interest to this thesis, namely the gradual takeover of English as the global language of instruction and the increasing role of multimedia, joined by the rise in online teaching.

3.4.1 English as a medium of instruction (EMI)

With on average about 8,000 commercial planes in the air at any given time of day, distances have become marginal for the world. People studying on the other side of the globe are no longer exceptional. There are, however, still 7,100 languages in the world. So how do these international students get access to education across the world? The response of educational institutions is rather simple. With approximately one in seven people speaking English (Ethnologue, 2022), they started offering educational programs taught in English. The past decades have seen a considerable rise in English as a medium of instruction (EMI) being used at European higher education institutions (Dearden, 2014; Wächter & Maiworm, 2008, 2014). The Nordic countries, in addition to the Netherlands, seem to be the frontrunners in terms of offering English-taught programs. About 30% of study programs in the Netherlands are provided in English (Wächter & Maiworm, 2014). Even though the Netherlands and Belgium are neighbors and even share native languages, the presence of EMI differs significantly. In

Belgium, a comparatively meager 7.5% of study programs are offered in English. The difference here lies in policymaking and the perception of EMI. The Netherlands actively promotes the use of EMI through policy and institutional changes as it is seen as a method of internationalization (Breetvelt, 2018; Wilkinson, 2013). Contrarily, the Northern part of Belgium, Flanders, restricts the use of EMI through policy (Vlaamse Onderwijsraad, 2017) because of the fear for any drawbacks it may have on its national students and the delicate historical position of the native language. EMI is, however, assumed to yield certain benefits.

Higher education institutions can decide to introduce EMI for a wide variety of reasons. The more common driving forces for the introduction of EMI, however, are the following (Galloway et al., 2017): To increase global competitiveness through internationalization, to boost the enrolment and the consequent income gained from foreign students, to enhance international mobility or to improve employability and international competences of graduates. Of course, the use of EMI has also gained popularity due to the fact that English has received a considerable status as the language of science and academic publications (Al Zumor, 2019).

Another important driving force for using EMI in tertiary education is that it is assumed to increase the students' English language proficiency, making them more competitive on the job market. In that aspect, EMI can be considered a form of Content and Language Integrated Learning (CLIL). This assumption, however, seems to be based mostly based on self-reports from students, who report benefits regarding language acquisition with no downsides to comprehension. Empirical results on actual language acquisition through use of EMI remain inconclusive (Pérez-Cãnado, 2012; Vidal & Jarvis, 2018).

This begs the question whether EMI indeed has benefits other than making education accessible to a larger, international audience. Many fear the introduction of EMI would negatively impact the quality of the education and the comprehension of domestic students (Dafouz & Cmacho-Miñano, 2016). It is still a matter of contention whether this fear is just as some studies indeed report increased comprehension difficulty for EMI lectures when empirically tested (e.g., Airey & Linder, 2006; Hellekjær, 2010), whereas others reveal no impact of EMI on student learning (Dafouz & Cmacho-Miñano, 2016).

To cope with the potential downsides of EMI for domestic students, subtitles might be of help. While it can be considered rather difficult to integrate subtitles in live classroom lectures, some European universities already do so (e.g., Karlsruhe Institute of Technology in Germany). This can be done using automatic speech recognition as a speech-to-text tool. Another way to integrate subtitles in education, would be to do so in recorded lectures. These subtitles do not have to be generated live, but can be prepared in advance. It may even be possible that students start subtitling these lectures in the same way some may do so for their favorite series, i.e., fan-subbing. Admittedly, the latter may be a utopian thought. Nevertheless, online education would offer more and better opportunities for the introduction of subtitles.

3.4.2 Online teaching

In the past, subtitles in online education might have seemed far-fetched due to the minimal use of online lectures. Since the turn of the millennium, however, a considerable shift in education could be observed. Online teaching has flourished. Educational content on the internet (e.g., web lectures, MOOCs, educational videos) is growing and becoming more readily available to the general public. This steady growth received a sudden boost when the COVID-19 pandemic happened, during which all higher education institutions were suddenly required to switch to online teaching. Even though face-to-face education has returned in most institutions, online education, or a hybrid form (face-to-face courses that are streamed or recorded), is clearly here to stay. With the current omnipresence of online education and the more straightforward introduction of subtitles in online lectures, subtitled education in the near future seems very plausible.

One important aspect to consider, however, is the fact that online education also comes in many more shapes and sizes than regular classroom lectures. If subtitles were to be integrated, it is important to consider for which lectures it would be advisable to do so. The combination of lecture styles and subtitles has so far remained relatively underexplored in research. The first step to deepen our understanding of the matter, would be to look at what styles of online lectures occur frequently. Some common styles are:

- A lecture capture: A recording of a classroom lecture.
- A talking head lecture: A recording of the upper body of the lecturer teaching the course (sometimes considered a lecture capture).
- A voice-over lecture: An animated video or (PowerPoint) presentation with only the voice of the lecturer playing in the background.
- A picture-in-picture lecture: An animated video or (PowerPoint) presentation with a video feed of the lecturer showing on screen as well.
- An overlay lecture: An animated video or (PowerPoint) presentation with an integrated video of the lecturer. In contrast to the picture-in-picture lecture, the image of the lecturer in the overlay lecture can move around on screen and has no background. As such, the lecturer seemingly stands or walks in front of a digital blackboard.
- A Khan-style lecture: A lecture showing a digital blackboard or tablet that the lecturer uses to draw or write on. This can be done using screencast (not showing the hands of the lecturer) or using a top-down camera (showing the hands and gestures of the lecturer).

Despite there being a considerable number of lecture styles, there were no guidelines on how to present or what to consider when producing video lectures ten years ago (Ilioudi et al., 2013) and we are not aware of any published guidelines since. Research on the advantages and disadvantages of each style of lecture is also rather limited (C.-M. Chen & Wu, 2015). Still there have been some studies that empirically compared lecture styles as discussed in the following paragraphs.

Because the majority of online lectures is part of a full online course (e.g., as a MOOC), it can be interesting to first look at what students prefer in these lectures. Two studies that have investigated course attrition and viewing patterns for complete online courses were conducted by Bhat et al. (2015) and Guo et al. (2014). Bhat et al. (2015) looked into two frequently used modes on Coursera, a platform that offers MOOCs. MOOCs generally have high attrition because they require students to follow courses autonomously while also offering little to no interaction with an instructor. This feeling of isolation can lead to a decrease in motivation, eventually ending in the student no longer putting effort into following the course. Bhat et al. (2015) studied two styles of lectures, a picture-in-picture style lecture and an overlay lecture. They looked at engagement, motivation and navigation, quantified by a number of variables on online activity (e.g., viewing time, certificate-earner proportion, discussion forum activity). They found learners generally preferred the overlay-styled lectures. The overlay-styled lectures lead to more time being spent on watching the lectures and more lectures being watched overall. This effect is assumed to be because of the proximity of the lecturer to the slide content, the larger size of the lecturer and, potentially, the integrated view of a real instructor. More research, preferably qualitative work, would be required to really assess why an integrated view leads to more active response.

Guo et al. (2014) looked at students' activity on edX, another MOOC platform. They took into account the style of the video the students watched, which could be an ordinary lecture, a tutorial (walkthrough) or any other content (making up only 11% of the videos watched). They also looked at the speed setting students used when watching the lectures, which is something that seems to be regularly forgotten in similar research. Overall, they found shorter videos (i.e., below nine minutes, though ideally below three minutes) and videos alternating between a talking head and slides to be more engaging. They also found that pre-produced lectures (i.e., not just recordings of classroom lectures) led to more engagement, as did the videos where the instructor spoke faster rather than slower. With regards to tutorials, they found Khan-styled tutorials (i.e., an instructor writing on a tablet) to be most effective. The engagement also differed significantly depending on whether it was a lecture or a tutorial. This study has some important implications as it shows that there is no simple answer when it comes to effectiveness of lecture styles, which also depends on speech rate and the type of teaching, among others.

The presence of the instructor is a recurring topic in research on online lecture styles. Going back to one of the cognitive principles, the image principle, it is assumed that the presence of the lecturer (a talking head) does not necessarily improve learning and can thus be excluded. While some lecturers may find this disheartening, research on the matter may provide some comfort as the presence of the lecture has been found to affect students, both positively and negatively.

In 2008, Homer et al. conducted a study showing a 20-minute video on child development to 26 undergraduate students either including both a PowerPoint presentation and a video of the lecturer or only a PowerPoint presentation. They used questionnaires to assess social presence, cognitive load and knowledge transfer. They only found a significant increase in cognitive load for the students who watched the combined lecture, thus supporting the idea that the image of the instructor does more harm than good.

In contrast, a study by Lyons et al. (2012) found that the presence of the instructor was generally appreciated by students. They tested 158 students after they followed an online course consisting of 13 lectures. The students received all lectures either with the image of the instructor in the top left corner or without the image of the instructor. They also considered technological efficacy as a factor of the participants. They found the image of the instructor to lead to higher perceived learning and a heightened feeling of interactivity. However, perceived learning and the overall perception of the lectures containing the image decreased as students exhibited lower technological efficacy.

Another study on social presence in online lectures was done by Kizilcec et al. (2014). They had 22 graduate students watch seven video segments, each time three or four segments with or without a video of the instructor. They used eye tracking, response questionnaires and a recall test to examine the effects of social presence. Participants were found to prefer the video with the instructor included. When the face was present, they also spent considerable time (41%) on the face of the instructor. Recall on the other hand was not affected by the presence or absence of the instructor's face.

A slightly different study was conducted by Lee (2014). They studied the effects of online PowerPoint presentations with a humanlike animated character, with a monster-like animated character and without any social presence on 176 undergraduate students. The students were divided into three groups, each assigned to one condition, and had to follow five classes in the respective lecture style. They used questionnaires to assess social perceptions, arousal, pleasure, flow, motivation and learning performance. They found significantly more emotional response and higher learning outcomes in the slide presentation when compared to the presentations that included an animated character. They assume this might be because learners are familiar with that method and associate it more with conventional learning. This could indicate the importance of habit when it comes to the results of instructional design on emotional response and learning.

In 2016, Korving et al. examined the effect of visibility of the lecturer in online lectures on 91 students. The students had to register online and were required to watch three online lectures of approximately 7:30 minutes in three conditions: in one lecture, the video of the lecturer was large, in another it was small and in the third condition, the video of the lecturer was omitted. Using questionnaires, enjoyment, perception of the lecturer and content, interest and relevance were measured, alongside the participants' habitual radio-listening, TV-watching, visiting of elective lectures and attention during online and offline lectures. They found inconclusive results on the relation between visibility of the lecturer and reported attention, as it depended on whether it was the first or second lecture the students watched. The results actually hint at too much visibility negatively impacting reported attention. Following their inconclusive results, they highlight that the visibility of the lecturer might become more important over time as attention gradually decreases and it is easier to pay attention to the lecturer than to pay attention to the slides.

Wang and Antonenko (2017) showed 36 undergraduate students four videos on geometry and algebra: two difficulties, each in two styles (one picture-in-picture, thus including the instructor, and one without an instructor). Using an SR EyeLink 1000+, they measured the participants' eye movements during the viewing. They also assessed perceived learning, satisfaction, cognitive effort, perception of instructor presence and actual learning using self-report questionnaires and tests. They found the presence of the instructor to improve satisfaction and perceived learning in both difficulties. The instructor also received a considerable amount of visual attention (26% of total dwell time), especially when the video was considered to be easier. The instructor's presence also improved recall, but only in the easy condition. They assume the increase in performance is tied to the nonverbal cues that come with the presence of the instructor.

A last study examining social presence in lectures was done by Lackman et al. (2021). They had 26 students between the ages of 20 and 40 watch one 14-minute psychology lecture, either in a rich infographic style, continuously showing images, graphics and text with a narration, or as a video recording of a class lecture showing just the professor and some students. Using electroencephalography (EEG), electrodermal activity (EDA), facial expression recognition and self-assessment questionnaires, they measured emotional and cognitive engagement. With a comprehension test, they measured learning performance. With regard to emotional engagement, they found the infographic lecture to maintain arousal significantly better, but the recorded lecture was found to be more engaging and enjoyable to watch. For cognitive engagement, they did find a positive influence of the infographic lecture on attention. On top of that, the infographic lecture caused the subjects to answer more difficult comprehension questions correctly, meaning it led to an increase in learning performance. They also found a relation between emotional and cognitive engagement and learning performance. In conclusion, these different lecture styles lead to different types of engagement. While more engagement is associated with more learning gains, the arguably more informative infographic does outperform a simple recorded lecture, albeit only marginally.

In sum, these studies generally find that the presence of an instructor on screen has no actual benefit to knowledge transfer, at times increases cognitive load slightly, but does tend to be appreciated by students and perceived as beneficial to their comprehension and retention. Another recurrent finding is that the image of the lecturer also receives a considerable amount of visual attention, often taking up almost half of the total viewing time. However, Louwerse et al. (2009) tested whether this means the image of the lecturer distracts the student from important content or whether they are only looked at when time allows it or are relevant. They examined eye movements of students watching multimedia instruction with animated instructors. They recruited 12 undergraduate students and had them complete an interactive computerized tutorial session on computer literacy. They found students to look at the animated agent in a similar way as they would look at a human conversational partner. In a second study they investigated the eye movements of seven undergraduate students when multiple animated characters were shown on screen. This, again, showed similar eye movement patterns as would be observed in real-life settings. Because both studies showed students only paid attention to the agents when they were relevant, Louwerse et al. (2009) conclude that the mere presence of the agent is unlikely to affect cognitive load, even though they can sometimes be considered redundant.

In the present thesis, we will compare three lecture types in particular, a talking head lecture, a voice-over lecture with a PowerPoint presentation and a picture-in-picture lecture (PowerPoint presentation + a talking head). Subtitles are added to these lectures to see how lecture style and subtitles interact and affect the student. Few studies have compared these three formats empirically and even fewer have studied subtitles in relation to them.

One study that compared these three formats was done by C.-M. Chen and Wu (2015). They recruited 37 students and had them watch three lecture videos in three different styles: (1) a recorded lecture in a classroom setting (slightly different from a talking head lecture); (2) a voice-over PowerPoint presentation; and (3) a picture-in-picture style lecture, showing the PowerPoint presentation and integrating the lecturer's image on the slides. Using rudimentary EEG, heart rate variability (HRV) measures, self-reported cognitive load and a comprehension test, they measured attention, emotion, cognitive load and learning performance. They found sustained attention to be higher for the voice-over lecture than the picture-in-picture lecture. It also outperformed the lecture capture, though this difference was not significant. They found no effect on emotion, but did find an effect on cognitive load, with cognitive load being higher for the voice-over lecture. As expected, they also found learning performance to be higher for the other two lecture styles.

3.4.3 Lecture styles and subtitles

Two important studies that considered lecture style (or complexity) and subtitles were done by van der Zee et al. (2017) and Chan (2020). Van der Zee et al. (2017) studied the effects of subtitles, lecture complexity and language proficiency on knowledge transfer and cognitive load in asynchronous lectures. As part of the study, 125 students watched four videos: two lectures with high-visual textual complexity, one with and one without subtitles, and two lectures with low complexity, with and without subtitles. Though they did not find a significant effect of the subtitles, they did find an effect of lecture complexity and language proficiency on knowledge transfer. The study was, however, limited to questionnaires and tests only. It would be interesting to see how the eyes move in these different environments and how they relate to the learning experience.

Chan (2020) conducted two studies investigating cognitive load and comprehension in recorded lectures with intralingual and interlingual subtitles. In the first study, 40 L1 English and 63 L1 Chinese students had to watch recorded lectures (Khan-style) in a classroom environment. Five students also wore eye-tracking glasses. They had to watch these lectures either without subtitles, with English (intralingual) subtitles or with Chinese (interlingual) subtitles. No difference was found for cognitive load depending on the subtitle presence or language. A significant interaction was found between the first language of the participant and the language of the subtitles, for which, surprisingly, L1 English speakers reported significantly higher cognitive load reading English subtitles than L2 English speakers reading the English subtitles.

Chan assumes the underlying reason to be that they are not used to reading their native language in this context, making the subtitles more distracting than beneficial. In addition, no difference was found for comprehension. Only when subtitles were shown on screen that were not in a language that the viewer could understand, comprehension was negatively affected. In the second study, 70 Chinese L1 participants watched the same lectures, again without, with English or with Chinese subtitles in a laboratory environment. Chan again found no difference for cognitive load. Comprehension, on the other hand, improved in the second study for students that had access to L1 subtitles. With regard to eye movements, significantly different eye movement patterns could be observed depending on the language of the subtitles. L1 Chinese speakers fixated more and longer on L2 subtitles than on L1 subtitles. In all cases, the reading patterns of L1 subtitles were similar to those in static reading. It only changed for L2 subtitles, which required longer fixation durations to process all information properly.

4 Conclusion

To conclude, a large number of studies have already examined the process of reading, studied audiovisual translation (AVT), or more specifically subtitles, and explored the design of lectures. Still, a number of knowledge gaps are present in the state-of-the-art. This thesis identifies three gaps it wishes to address.

Firstly, the process of reading static texts has been extensively studied. Consequently, there are well-founded models of eye movement control in reading that explain how, why and where the eyes move when reading. However, subtitles are distinctly different from static texts. They are dynamic in nature and regularly accompanied by audio and other visual information. So far, few studies have closely examined the process of reading subtitles with high-accuracy eye trackers. This could, however, provide valuable insight into the processing of subtitles and how subtitles and other content on screen impact one another.

Secondly, a number of studies have already investigated the effects of subtitles on comprehension and cognitive load. However, because of differing methodologies and the highly reactive and complex environment of subtitled audiovisual material, the results are inconclusive. This knowledge gap leads to two goals in this thesis. Firstly, this thesis wishes to add to a future methodologically sound foundation in AVT research by documenting and recommending an approach to prepare quasi-experimental and experimental AVT research. Using more unified and streamlined approaches in research could provide more insight into the actual effects of subtitles and what to consider. Secondly, this thesis examines comprehension and cognitive load in multiple settings of subtitled education, which could provide more insight into the interactions and effects of subtitles in this context.

Lastly, educational research has explored the design of online lectures and how, for example, the presence of the instructor or the use of a PowerPoint impacts the learner. Still, there are no guidelines present to aid in creating online lectures and what should be considered when producing online lectures. This thesis does not wish to create such a guideline, but instead wishes to contribute to the state-of-the-art by comparing three commonly used online lecture designs and studying their effect on comprehension and cognitive load. Additionally, few have so far considered both lecture design and subtitles in a single study. Considering how strongly subtitles and other visual content interact, this is rather surprising. This thesis intends to deepen our understanding of subtitles and lecture design by considering both at once in a number of studies conducted over a four-year period.

5 References

Ainsworth, S. (2014). The multiple representation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 464–486). New York, NY: Cambridge University Press.

Airey, J., & Linder, C. (2006). Language and the experience of learning university physics in Sweden. *European Journal of Physics*, *27*(3), 553–560.

Al Zumor, A. Q. (2019). Challenges of using EMI in teaching and learning of university scientific disciplines: Student voice. *International Journal of Language Education, 3*(1), 74–90. doi:10.26858/ijole.v1i1.7510

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. *Psychological Science, 15*(4), 225–231. doi:10.1111/j.0956-7976.2004.00656.x

Ark, W., Dryer, C. D., & Lu, D. J. (1999). *The emotion mouse.* Paper presented at the 8th International Conference on Human-Computer Interaction, London, England.

Ayres, P. (2018). Subjective measures of cognitive load: What can they reliably measure? In R. Z. Zheng (Ed.), *Cognitive load measurement and application: A theoretical framework for meaningful research and practice* (pp. 9–29). New York, NY: Routledge.

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Vol. 2, pp. 206–226). Cambridge, England: Cambridge University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *TRENDS in Cognitive Sciences, 4*, 417–423.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89): Elsevier.

Bairstow, D. (2012). *Rôles des sous-titres dans la compréhension et la mémorisation de films.* (Doctoral Thesis). Paul Valéry University of Montpellier,

Bairstow, D., & Lavaur, J.-M. (2017). Sous-titrage, compréhension de films et acquisition de vocabulaire. *Psychologie Française, 62*(3), 249–261. doi:10.1016/j.psfr.2015.05.001

BBC. (2023, March 23). The BBC subtitle guideline. https://www.bbc.co.uk/accessibility/forproducts/guides/subtitles/

Bhat, S., Chinprutthiwong, P., & Perry, M. (2015, June 26–29). *Seeing the instructor in two video styles: Preferences and patterns.* Paper presented at the 8th International Conference on Educational Data Mining, Madrid, Spain.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*(4), 509–533. doi:10.1017/S0142716402004022

Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PLoS One, 11*(6). doi:10.1371/journal.pone.0158409

Bisson, M.-J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics, 35*(2), 399–418. doi:10.1017/s0142716412000434

Black, S. (2022). Could integrated subtitles benefit young viewers? Children's reception of standard and integrated subtitles: A mixed methods approach using eye tracking. *Perspectives, 30*(3), 503–519. doi:10.1080/0907676X.2020.1849324

Brasel, A. S., & Gips, J. (2014). Enhancing television advertising: Same-language subtitles can improve brand recall, verbal memory, and behavioral intent. *Journal of the Academy of Marketing Science, 42*(3), 322–336. doi:10.1007/s11747-013-0358-1

Breetvelt, I. (2018). English-medium instruction in Dutch higher education: a policy reconstruction and impact study [Special Issue]. *L1-Educational Studies in Language and Literature*, 18, 1–24. doi:10.17239/L1ESLL-2018.18.03.10

Butcher, K. R. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 174–205). New York, NY: Cambridge University Press.

Chan, W. S. (2020). *An investigation of subtitles as learning support in university education* [Unpublished doctoral dissertation]. Macquarie University, Sydney, Australia.

Chan, W. S., Kruger, J.-L., & Doherty, S. (2019). Comparing the impact of automatically generated and corrected subtitles on cognitive load and learning in a first- and second-language educational context. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*(18), 237–272.
Chen, C.-M., & Wu, C.-H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education, 80*, 108–121.
Chen, F., Zhou, J., Wang, J., Kun, Y., Arshad, S. Z., Khawaji, A., & Conway, D. (2016). *Robust multimodal cognitive load measurement*. Bern, Switzerland: Springer.
Clifton Jr., C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). Amsterdam, Netherlands: Elsevier.
Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*(1), 51–57. doi:10.1177/0963721409359277
Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94*(2), 428–434. doi:10.1037//0022-0663.94.2.428
d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist, 12*(3), 196–205. doi:10.1027/1016-9040.12.3.196
d'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In K. Rayner (Ed.), *Eye movements and visual cognition* (pp. 415–427). New York, NY: Springer.
Dafouz, E., & Cmacho-Miñano, M. M. (2016). Exploring the impact of English-medium instruction on university student academic achievement: The case of accounting. *English for Specific Pruposes, 44*, 57–67. doi:10.1016/j.esp.2016.06.001
Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta, 49*(1), 67–77. doi:10.7202/009021ar
de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*, 105–134. doi:10.1007/s11251-009-9110-0
Dearden, J. (2014). *English as a medium of instruction: A growing global phenomenon*. British Council
Díaz Cintas, J. (2003). *Teoría y práctica de la subtitulación: Inglés-Español [Theory and practice of subtitling: English-Spanish]*. Barcelona, Spain: Ariel.
Díaz Cintas, J. (2020). Audiovisual translation. In E. Angelone, M. Ehrensberger-Dow, & G. Massey (Eds.), *The Bloomsbury Companion to Language Industry Studies* (pp. 209–230). London, England: Bloomsbury Publishing Plc.
Díaz Cintas, J., & Remael, A. (2014). *Audiovisual Translation: Subtitling*. Manchester, England: St. Jerome.
Díaz Cintas, J., & Remael, A. (2021). *Subtitling: Concepts and practices* (Vol. 1). New York, NY: Routledge.
Díaz Cintas, J., & Szarkowska, A. (2020). Introduction: Experimental research in audiovisual translation - Cognition, reception, production. *The Journal of Specialised Translation, 33*, 3–16.
Doherty, S., & Kruger, J.-L. (2018). The development of eye tracking in empirical research on subtitling and captioning. In T. Dwyer, C. Perkins, S. Redmond, & J. Sita (Eds.), *Seeing into screens: Eye tracking and the moving image* (pp. 46–64). London, England: Bloomsbury.
Downey, G. J. (2008). *Closed captioning subtitling, stenography, and the digital convergence of text with television*. Baltimore, MD: John Hopkins University Press.
Duchowski, A. T., Krejtz, K., Gehrer, N. A., Bafna, T., & Bækgaard, P. (2020). *The low/high index of pupillary activity*. Paper presented at the CHI Conference on Human Factors in Computing Systems, Honolulu, HI.
Duchowski, A. T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., . . . Giannopoulus, I. (2018). *The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation*. Paper presented at the CHI Conference on Human Factors in Computing Systems, Montréal, Canada.
Engbert, R., Longtin, A., & Reinhold, K. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42*, 621–636.
Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*(4), 777–813. doi:10.1037/0033-295X.112.4.777
Ethnologue. (2022, November 9). Welcome the 25th edition. https://www.ethnologue.com/ethnoblog/welcome-25th-edition/

Findelsberger, E., Hutzler, F., & Hawelka, S. (2019). Spill the load: Mixed evidence for a foveal load effect, reliable evidence for a spillover effect in eye-movement control during reading. *Attention, Perception, & Psychophisics, 81*, 1442–1453.

Fox, W. (2016). Integrated titles: An improved viewing experience? In S. Hansen-Schirra & S. Grucza (Eds.), *Eyetracking and applied linguistics* (pp. 5–30). Berlin, Germany: Language Science Press.

Fox, W. (2018). *Can integrated titles improve the viewing experience? Investigating the impact of subtitling on the reception and enjoyment of film using eye tracking and questionnaire data*. Berlin, Germany: Language Science Press.

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans: A framework for defining "early" visual processing. *Experimental Brain Research, 142*(1), 139–150.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210. doi:0010-0285/82/020178-33 $05.00/0

Frey, A., & Bosse, M.-L. (2018). Perceptual span, visual span, and visual attention span: Three potential ways to quantify limits on visual processing during reading. *Visual Cognition, 26*(6), 412–429. doi:10.1080/13506285.2018.1472163

Galloway, N., Kriukow, J., & Numajiri, T. (2017). *Internationalisation, higher education and the growing demand for English: An investigation into the English medium of instruction (EMI) movement in China and Japan*. London, England: British Council.

Gambier, Y. (2009). Challenges in research on audiovisual translation. In A. Pym & A. Perekrestenko (Eds.), *Translation Research Projects 2* (pp. 17–25). Tarragona, Spain: Intercultural Studies Group.

Geary, D. C. (2008). An Evolutionarily Informed Education Science. *Educational Psychologist, 43*(4), 179–195. doi:10.1080/00461520802392133

Geary, D. C., & Berch, D. B. (2016). Evolution and children's cognitive and academic development. In D. C. Geary & D. B. Berch (Eds.), *Evolutionary Perspectives on Child Development and Education* (pp. 217–249). Bern, Switzerland: Springer.

Gerber-Morón, O., Szarkowska, A., & Woll, B. (2018). The impact of text segmentation on subtitle reading. *Journal of Eye Movement Research, 11*(4). doi:10.16910/jemr.11.4.2

Gerjets, P., Scheiter, K., & Cierniak, G. (2009). The scientific value of cognitive load theory: A research agenda based on the structuralist view of theories. *Educational Psychology Review, 21*, 43–54. doi:10.1007/s10648-008-9096-1

Gernsbacher, M. A. (2015). Video captions benefit everyone. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 195–202. doi:10.1177/2372732215602130

Ghia, E. (2012). The impact of translation strategies on subtitle reading. In E. Perego (Ed.), *Eye tracking in audiovisual translation* (pp. 157–182). Rome, Italy: Aracne.

Gottlieb, H. (2012). Subtitles: Readable dialogue? In E. Perego (Ed.), *Eye tracking in audiovisual translation* (pp. 37–82). Rome, Italy: Aracne.

Guo, P. J., Kim, J., & Rubin, R. (2014). *How video production affects student engagement: An empirical study of MOOC videos.* Paper presented at the ACM conference on learning, Atlanta, Georgia.

Häikiö, T., Bertram, R., Hyönä, J., & Niemi, P. (2009). Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. *Journal of Experimental Child Psychology, 102*, 167–181.

Harskamp, E. G., Mayer, R. E., & Suhre, C. (2007). Does the modality principle for multimedia learning apply to science classrooms? *Learning and Instruction, 17*, 465–477. doi:10.1016/j.learninstruc.2007.09.010

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139–183.

Hefer, E. (2013a). Reading first and second language subtitles: Sesotho viewers reading in Sesotho and English. *Southern African Linguistics and Applied Language Studies, 31*(3), 359–373. doi:10.2989/16073614.2013.837610

Hefer, E. (2013b). Reading second language subtitles: A case study of Afrikaans viewers reading in Afrikaans and English. *Perspectives, 21*(1), 22–41. doi:10.1080/0907676x.2012.722652

Hellekjær, G. O. (2010). Lecture comprehension in English-medium higher education. *HERMES - Journal of Language and Communication in Business, 23*(45), 11–34. doi:10.7146/hjlcb.v23i45.97343

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception* (pp. 269–293). Amsterdam, The Netherlands: Elsevier Science Ltd.

Henderson, J. M., Weeks Jr., P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology, 25*(1), 210–228. doi:0096-I523/99/S3.00

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143*(3611), 1190–1192. doi:10.1126/science.143.3611.1190

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior, 24*(3), 786–797.

Ilioudi, C., Giannakos, M. N., & Chorianopoulos, K. (2013, April 8 – April 9). *Investigating differences among the commonly used video lecture styles.* Paper presented at the Workshop on Analytics on Video-based learning, Leuven, Belgium.

Inhoff, A. W., Kim, A., & Radach, R. (2019). Regressions during reading. *Vision, 3*(35). doi:10.3390/vision3030035

Ivarsson, J., & Caroll, M. (1998). *Subtitling*: TransEdit.

Jamet, E., & Le Bohec, O. (2007). The effect of redundant text in multimedia instruction. *Contemporary Educational Psychology, 32*, 588–598.

Jensen, O., Pan, Y., Frisson, S., & Wang, L. (2021). An oscillatory pipelining mechanism supporting previewing during visual exploration and reading. *TRENDS in Cognitive Sciences, 25*(12), 1033–1044. doi:10.1016/j.tics.2021.08.008

Jorden, T. R., McGowan, V. A., Kurtev, S., & Paterson, K. B. (2016). A futher look at postview effects in reading: An eye-movements study of influences from the left of fixation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(2), 296–307. doi:10.1037/xlm0000157

Johnson, C. I., & Priest, H. A. (2014). The feedback principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 449–463). New York, NY: Cambridge University Press.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354.

Kalyuga, S. (2012). Instructional benefits of spoken words: A review of cognitive load factors. *Educational Research Review, 7*(2), 145–159. doi:10.1016/j.edurev.2011.12.002

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*(1), 126–136. doi:10.1037//0022-0663.92.1.126

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors, 46*(3), 567–581. doi:10.1518/hfes.46.3.567.50405

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Vol. 2, pp. 247–262). Cambridge, England: Cambridge University Press.

Karamitroglou, F. (1998). A proposed set of subtitling standards in Europe. *Translation Journal, 2*(2).

Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). *Showing face in video instruction: Effects on information retention, visual attention, and affect.* Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Toronto, Canada.

Kolb, H. (2005, January 31, 2012). Simple anatomy of the retina. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK11533/

Korving, H., Hernández, M., & de Groot, E. (2016). Look at me and pay attention! A study on the relation between visibility and attention in weblectures. *Computers & Education, 94*, 151–161.

Krejtz, I., Szarkowska, A., & Krejtz, K. (2013). The effects of shot changes on eye movements in subtitling. *Journal of Eye Movement Research, 6*(5), 1–12. doi:10.16910/jemr.6.5.3

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLoS One, 13*(9).

Krekelberg, B. (2010). Saccadic suppression. *Current Biology, 20*(5), 228–229.

Kruger, J.-L. (2016). Psycholinguists and audiovisual translation. *Target, 28*(2), 276–287. doi:10.1075/target.28.2.08kru

Kruger, J.-L., & Doherty, S. (2016). Measuring cognitive load in the presence of educational video: Towards a multimodal methodology. *Australasian Journal of Educational Technology, 32*(6), 19–31. doi:10.14742/ajet.3084

Kruger, J.-L., Doherty, S., Fox, W., & de Lissa, P. (2018). Multimodal measurement of cognitive load during subtitle processing: Same-language subtitles for foreign-language viewers. In I. Lacruz & R. Jääskeläinen (Eds.), *Innovation and expansion in translation process research* (pp. 267–294). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Kruger, J.-L., Hefer, E., & Matthew, G. (2013). *Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts.* Paper presented at the 2013 Conference on Eye Tracking South Africa, Cape Town, South Africa.

Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research*, *7*(5), 1–15. doi:10.16910/jemr.7.5.4

Kruger, J.-L., Soto-Sanfiel, M. T., Doherty, S., & Ibrahim, R. (2016). Towards a cognitive audiovisual translatology. In R. M. Martín (Ed.), *Reembedding Translation Process Research* (pp. 171–193). Amsterdam, The Netherlands: John Benjamins.

Kruger, J.-L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly, 49*(1), 105–120. doi:10.1002/rrq.59

Kruger, J.-L., Wisniewska, N., & Liao, S. (2022). Why subtitle speed matters: Evidence from word skipping and rereading. *Applied Psycholinguistics, 43*, 211–236.

Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly, 45*(2), 230–251. doi:10.1598/RRQ.45.2.4

Lackman, S., Léger, P.-M., Charland, P., Aubé, C., & Talbot, J. (2021). The influence of video format on engagement and performance in online learning. *Brain Sciences, 11*(2).

Lång, J., Vrzakova, H., & Mehtätalo, L. (2021). Modelling gaze behaviour in subtitle processing: The effect of structural and lexical properties. *Journal of Audiovisual Translation, 4*(1), 71–95. doi:10.47476/jat.v4i1.2021.104

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25*, 943–951. doi:10.1002/acp.1787

Leahy, W., & Sweller, J. (2015). Cognitive load theory and the effects of transient information on the modality effect. *Instructional Science, 44*, 107–123. doi:10.1007/s11251-015-9362-9

Lee, Y.-H., Hsiao, C., & Ho, C.-H. (2014). The effects of various multimedia instructional materials on students' learning responses and outcomes: A comparative experimental study. *Computers in Human Behavior, 40*, 119–132.

Leppink, J., Paas, F. G. W. C., Van der Vleuten, C. P. M., van Gog, T., & van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*(4), 1058–1072. doi:10.3758/s13428-013-0334-1

Leppink, J., & van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education, 4*(3), 119–127. doi:10.1007/s40037-015-0192-x

Liao, S., Kruger, J.-L., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of Specialised Translation*(33), 70–98.

Liao, S., Yu, L., Kruger, J.-L., & Reichle, E. D. (2022). The impact of audio on the reading of intralingual versus interlingual subtitles: Evidence from eye movements. *Applied Psycholinguistics*, *43*(1), 237–269. doi:10.1017/S0142716421000527

Liao, S., Yu, L., Reichle, E. D., & Kruger, J.-L. (2021). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading, 25*(5), 417–435.

Lobier, M., Dubois, M., & Sylviane, V. (2013). The role of visual processing speed in reading speed development. *PLoS One, 8*(4).

Loftus, G. R., Kaufman, L., Nishimoto, T., & Ruthruff, E. (1992). Effects of visual degradation on eye-fixation durations, perceptual processing, and long-term visual memory. In K. Rayner (Ed.), *Eye Movements and Visual Cognition* (pp. 203–226). New York, NY: Springer.

Louwerse, M. M., Graesser, A. C., McNamara, D. S., & Lu, S. (2009). Embodied conversational agents as conversational partners. *Applied Cognitive Psychology, 23*, 1244–1255.

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Vol. 2, pp. 227–246). Cambridge, England: Cambridge University Press.

Lyons, A., Reysen, S., & Pierce, L. (2012). Video lecture format, student technological efficacy, and social presence in online courses. *Computers in Human Behavior, 28*, 181–186.

Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals, 32*(3), 321–328. doi:10.1111/j.1944-9720.1999.tb01344.x

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81*(12), 899–917.

Mayer, R. E. (2014a). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43–71). New York, NY: Cambridge University Press.

Mayer, R. E. (2014b). Principles based on social cues in multimedia learning: Personalization, Voice, Image, and Embodiment principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 345–368). New York, NY: Cambridge University Press.

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 279–315). New York, NY: Cambridge University Press.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187–198. doi:10.1037//0022-0663.93.1.18

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52. doi:10.1207/S15326985EP3801_6

Mayer, R. E., & Pilegard, C. (2014). Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 316–344). New York, NY: Cambridge University Press.

McClarty, R. (2014). In support of creative subtitling: Contemporary context and theoretical framework. *Perspectives, 22*(3), 592–606. doi:10.1080/0907676X.2013.842258

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*(6), 578–586.

Mellinger, C. D., & Hanson, T. A. (2017). *Quantitative research methods in translation and interpreting studies*. New York, NY: Routledge.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97. doi:10.1037/h0043158

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the selective reanalysis hypothesis. *Journal of Memory and Language, 59*, 266–293. doi:10.1016/j.jml.2008.06.002

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, 1–25. doi:10.1017/s0272263119000706

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology, 18*(1), 118–141.

Montero Perez, M., Peters, E., & Desmet, P. (2013). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL, 26*(1), 21–43. doi:10.1017/s0958344013000256

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal, 99*(2), 308–328. doi:10.1111/modl.12215

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning, 31*(1–2), 1–26. doi:10.1080/09588221.2017.1375960

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358–368.

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*(1), 156–163. doi:10.1037//0022-0663.94.1.156

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology, 10*(5), 667–682. doi:10.1037/0096-1523.10.5.667

Netflix. (2023, March 23). English Timed Text Style Guide. https://partnerhelp.netflixstudios.com/hc/en-us/articles/217350977-English-Timed-Text-Style-Guide

Orero, P., Doherty, S., Kruger, J.-L., Matamala, A., Pedersen, J., Perego, E., . . . Szarkowska, A. (2018). Conducting experimental research in audiovisual translation (AVT): A position paper. *Journal of Specialised Translation, 30*, 105–126.

Paas, F. G. W. C., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63–71. doi:10.1207/s15326985ep3801_8

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122–133. doi:10.1037/0022-0663.86.1.122

Paivio, A. (1986). *Mental Representations: A dual coding approach*. Oxford, England: Oxford University Press.

Perego, E. (2008). Subtitles and line-breaks: Towards improved readability. In D. Chiaro, C. Heiss, & C. Bucaria (Eds.), *Between text and image: Updating research in screen translation* (pp. 211–223). Amsterdam, The Netherlands: John Benjamins.

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology, 13*(3), 243–272. doi:10.1080/15213269.2010.502873

Perego, E., Orrego-Carmona, D., & Bottiroli, S. (2016). An empirical take on the dubbing vs. subtitling debate: An eye movement study. *Lingue e Linguaggi, 19*, 255–274. doi:10.1285/i22390359v19p255

Pérez-Cãnado, M. L. (2012). CLIL research in Europe: Past, present, and future. *International Journal of Bilingual Education and Bilingualism, 14*(3), 315–341.

Pollatsek, A., Raney, G. E., Lagasse, L., & Rayner, K. (1993). The use of information below fixation in reading and visual search. *Canadian Journal of Experimental Psychology, 47*(2), 179–200.

Radach, R., & Kennedy, A. (2013). Eye movements in reading: Some theoretical context. *The Quarterly Journal of Experimental Psychology, 66*(3), 429–452. doi:10.1080/17470218.2012.750676

Rajendran, D. J., Duchowski, A. T., Orero, P., Martínez, J., & Romero-Fresco, P. (2013). Effects of text chunking on subtitling: A quantitative and qualitative examination. *Perspectives, 21*(1), 5–21. doi:10.1080/0907676X.2012.722651

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology, 41*, 211-236 doi:10.1016/0022-0965(86)90037-8

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology, 30*(4), 20–732

Rayner, K., Pollatsek, A., Ashby, J., & Clifton Jr., C. (2012). *Psychology of reading*. New York, NY: Psychology Press.

Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin & Review, 17*(6), 834-839. doi:10.3758/PBR.17.6.834

Reichle, E. D. (Ed.) (2021). *Computational models of reading: A handbook*. New York, NY: Oxford University Press.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125–157. doi:10.1037/0033-295X.105.1.125

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research, 7*(1), 4–22.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*, 445–526.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16*(1), 1–21.

Richter, E. M., Engbert, R., & Kliegl, R. (2006). Current advances in SWIFT. *Cognitive Systems Research, 7*, 23–33.
Rodziewicz-Cybulska, A., Krejtz, K., Duchowski, A. T., & Krejtz, I. (2022). *Measuring cognitive effort with pupillary activity and fixational eye movements when reading: Longitudinal comparison of children with and without primary music education*. Paper presented at the Symposium on Eye Tracking Research and Applications, Seattle, WA.
Romero-Fresco, P. (2009). More haste less speed - Edited versus verbatim respoken subtitles. *VIAL: Vigo International Journal of Applied Linguistics, 6*, 109–133.
Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*, 469–508. doi:10.1007/s10648-007-9053-4
Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophisics, 74*, 5–35. doi:10.3758/s13414-011-0219-2
Siegenthaler, E., Costela, F. M., McCamy, M. B., Di Stasi, L. L., Otero-Millan, J., Sonderegger, A., . . . Martinez-Conde, S. (2014). Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *European Journal of Neuroscience, 39*(2), 287–294.
Smith, P. L., & Ragan, T. J. (2004). *Instructional Design* (Vol. 3). NY, New York: Wiley.
Statista. (2022a, August 29). Global: Number of smarthphone users 2013-2028. https://www.statista.com/forecasts/1143723/smartphone-users-in-the-world
Statista. (2022b, August 29). Number of TV households worldwide 2026. https://www.statista.com/statistics/268695/number-of-tv-households-worldwide/
Sturt, P., & Kwon, N. (2018). Processing information during regressions: An application of the reverse boundary-change paradigm. *Frontiers in Psychology, 9*(1630). doi:10.3389/fpsyg.2018.01630
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.
Sweller, J., Kalyuga, S., & Ayres, P. (2011). *Cognitive load theory* (1 ed.). New York, NY: Springer.
Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. doi:1040-726X/98/0900-0251
Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292. doi:10.1007/s10648-019-09465-5
Swets, B., Desmet, T., Clifton, C. J., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition, 6*(1), 201–216. doi:10.3758/MC.36.1.201
Szarkowska, A., & Bogucka, L. (2019). Six-second rule revisited: An eye-tracking study on the impact of speech rate and language proficiency on subtitle reading. *Translation, Cognition & Behavior, 2*(1), 101–124. doi:10.1075/tcb.00022.sza
Szarkowska, A., & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: Evidence from eye movements. *PLoS One, 13*(6).
Szarkowska, A., & Gerber-Morón, O. (2019). Two or three lines: A mixed-methods study on subtitle processing and preferences. *Perspectives, 27*(1), 144–164.
Szarkowska, A., Krejtz, I., Klyswejko, Z., & Wieczorek, A. (2011). Verbatim, standard, or edited? Reading patterns of different captioning styles among deaf, hard of hearing and hearing viewers. *American Annals of the Deaf, 156*(4), 363–378. doi:10.1353/aad.2011.0039
Szarkowska, A., Krejtz, I., & Krejtz, K. (2017). Do shot changes really induce the rereading of subtitles? In J. Díaz Cintas & K. Nikoli (Eds.), *Fast-forwarding with audiovisual translation* (pp. 61–79). Bristol, England: Multilingual Matters.
Szarkowska, A., Krejtz, I., Pilipczuk, O., Dutka, L., & Kruger, J.-L. (2016). The effects of text editing and subtitle presentation rate on the comprehension and reading patterns of interlingual and intralingual subtitles among deaf, hard of hearing and hearing viewers. *Across Languages and Cultures, 17*(2), 183–204.
Szarkowska, A., Silva, B., & Orrego-Carmona, D. (2021). Effects of subtitle speed on proportional reading time: Re-analysing subtitle reading data with mixed effects models. *Translation, Cognition & Behavior, 4*(2), 305–330.
Tricot, A., & Sweller, J. (2013). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review, 26*, 265–283. doi:10.1007/s10648-013-9243-1

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review, 25*, 2005–2015. doi:10.3758/s13423-018-1432-y

van der Zee, T., Admiraal, W., Paas, F., Saab, N., & Giesbers, B. (2017). Effects of subtitles, complexity, and language proficiency on learning from education videos. *Journal of Media Psychology, 29*(1), 18–30.

van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 263–278). New York, NY: Cambridge University Press.

Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT Journal, 42*(4), 272–281. doi:10.1093/elt/42.4.272

Vergilino, D., & Beauvillain, C. (2000). The planning of refixation saccades in reading. *Vision Research, 40*, 3527–3538.

Vidal, K., & Jarvis, S. (2018). Effects of English-medium instruction on Spanish students' proficiency and lexical diversity in English. *Language Teaching Research*, 1–20. doi:10.1177/1362168818817945

VRT. (2023, March 23). Normen en instructies voor open ondertiteling. https://www.vrt.be/content/dam/vrtbe/over-de-vrt/externe-mederwerkers/normen_en_instructies_openOT.pdf

Vlaamse Onderwijsraad. (2017). *Advies taalbeleid in het Vlaamse hoger onderwijs*. https://assets.vlor.be/www.vlor.be/import/rho-rho-adv-1617-004_0.pdf

Vulchanova, M., Aurstad, L. M. G., Kvitnes, I. E. N., & Eshuis, H. (2015). As naturalistic as it gets: Subtitles in the English classroom in Norway. *Frontiers in Psychology, 5*(1510). doi:10.3389/fpsyg.2014.01510

Wächter, B., & Maiworm, F. (2008). *English-taught programmes in European higher education: The picture in 2007*. Bonn, Germany: Lemmens.

Wächter, B., & Maiworm, F. (2014). *English-taught programmes in European higher education: The state of play in 2014*. Bonn, Germany: Lemmens.

Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior, 71*, 79–89.

Wilkinson, R. (2013). English-medium instruction at a Dutch university: Challenges and pitfalls. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.), *English-medium instruction at universities: Global challenges* (pp. 3–24). Bristol, England: Multilingual Matters.

Wolverton, G. S., & Zola, D. (1983). The temporal characteristics of visual information extraction during reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 41–51). New York, NY: Academic Press.

Part II: Methodology and preparation

Chapter 3 – Main methodology

Chapter 3 elaborates on the methodology of this project. It consists of two separate sections. The first section provides a short introduction of the project and its main research questions. The second section gives an overview of the core materials used in the study and details the studies and their respective methodologies conducted to answer the research questions. As was made clear in the Section 2.4 of Chapter 2, there are some inconsistencies when it comes to methodologies in AVT research. A considerable part of this thesis was devoted to laying the groundworks for a solid research approach for future quasi-experimental and experimental AVT research, as well as strengthening the foundations of the experiments conducted for the present thesis. This means that although this chapter details the methodologies briefly, the methodologies (mainly materials) were altered over the course of the PhD project. Chapter 4 will shed more light on how this was done.

1 Research aims

This research project focuses on the use of subtitles in asynchronous online lectures taught in English. In broad terms, it aims to shed light on how the presence of subtitles and different languages of subtitles in different lecture styles interact and affect: (1) the viewing, reading and processing of a lecture; (2) the perceived cognitive load; and (3) the comprehension of the lecture. The project studies three lecture styles in particular, a talking head lecture, a voice-over PowerPoint lecture and a composite lecture (picture-in-picture lecture). Language proficiency (i.e., different levels of proficiencies among L2 speakers, but also a comparison of L2 speakers with L1 speakers) is taken into account throughout the research project.

The project is inspired by two challenges that many higher education institutions (HEIs) face today. Firstly, these institutions have to provide equal educational access to increasingly multilingual and multicultural student populations. A regularly suggested answer to this challenge is the switch to English as a medium of instruction (EMI). While this would offer more students access, it may negatively impact students and lecturers with limited proficiency in English. Another possible answer to this challenge would be subtitles. Subtitles could be offered in the necessary languages and still allow institutions to stick to the national language. They could also be combined with EMI lectures and offer support in English, the national language or any other foreign language. As was illustrated in the conclusion of Chapter 2, however, so far relatively little is known about subtitles in education. The second challenge is the offering of online and recorded lectures. The past decade has seen an increasing demand for online teaching. For a brief moment, online teaching even became obligatory in light of the COVID-19 pandemic. A considerable amount of research has focused on the effects of different types of online lectures, yet there are still no specific guidelines on what to take into account when producing online lectures. Even more so, the knowledge of subtitles in online lectures, a combination that is arguably more conceivable than subtitles in live classrooms, is limited to only a handful of studies. This research project thus wishes to advance the state-of-the-art by studying subtitles in asynchronous, online education and contribute to the specific knowledge gaps mentioned in the conclusion of Chapter 2.

Each of these knowledge gaps relate to one or more research questions formulated in this thesis. These research questions are listed below. For the sake of brevity, only major questions are included in the list below. As it would not be possible to address these knowledge gaps in a single study, a total of six experiments with a total of more than 250 participants were conducted over the course of four years. Each of these experiments, and consequent papers, relate to one or more research questions and hypotheses. To provide structure for the reader, Table 3 at the end of this chapter provides an overview of the specific research questions each individual paper addresses.

The overarching research question of the project was the following:

> *What is the effect of subtitle presence, subtitle language, and students' language proficiency levels on cognitive load, comprehension, viewing/reading behavior and processing in different styles of asynchronous, online lectures.*

This question can be broken down into multiple research questions:

> *RQ.1: What is the effect of lecture styles and subtitles on perceived cognitive load?*
>
> *RQ.2: What is the impact of different styles of subtitled lectures on comprehension?*
>
> *RQ.3: How is attention allocation affected by lecture styles and subtitles?*
>
> *RQ.4: How is reading affected by lecture styles and subtitles?*
>
> *RQ.5: What is the opinion of students on different styles of subtitled online lectures?*

2 Materials and methodologies

As mentioned before, methodologies and materials changed over the course of this thesis. Still, there are a number of tools for measurements and research materials that consistently remained at the core of the project. These were three recorded philosophy lectures in three lecture styles with premade subtitle tracks for English and Dutch, the psychometric questionnaire from Leppink and van den Heuvel (2015) and a biographical survey. For the sake of clarity, these core materials are discussed in a little more detail below. Each experiment also employed a comprehension test, but this test was adapted throughout the project and will thus not be discussed in this section (see Chapter 4). Other materials used were eye tracking systems (SMI or SR EyeLink), interviews, post-hoc surveys, language tests, working memory assessments and reading exercises. For specific information on these materials, we refer to the methodology section of each article.

2.1 Lectures

At the heart of this research project are three recorded EMI lectures on philosophy. We decided to produce three lectures as this would allow us to employ within-subject/repeated measures designs in which we could compare lecture styles and subtitle conditions. The lectures were carefully created starting from writing three similar texts to recording the texts being read aloud by a professional lecturer and comparing and analyzing the product. The specific preparation and subsequent analysis are also briefly discussed in Paper 1 in Chapter 4.

2.1.1 Texts

We chose philosophy as a topic for the lectures as it is a common course subject (be it a mandatory or optional course) at faculties of arts or linguistics at universities. It was especially representative of a course at one of the partner institutions of this thesis, namely the University of Antwerp, as it is an optional course for students studying applied linguistics. Philosopher and lecturer prof. dr. Frank Albers, who taught philosophy at the University of Antwerp at the time, was asked to write three equally difficult and similarly themed texts in English to be used as philosophy lectures. The lectures had to be suitable to be taught to undergraduate students. Professor Albers chose to write about three philosophers, namely Thomas Piketty, Jean-Jacques Rousseau and Alexis de Tocqueville, and their view on the topic (in)equality as, according to him, "(in)equality is one of the most pressing issues of our day, and these thinkers from the 18th, 19th and 21st century have/had important things to say about this vexed issue, things that are still relevant today" (F. Albers, personal communication, May 8, 2018). All three texts[1] can be found in the Appendix at the end of this thesis. Each of the texts focused on one of the aforementioned philosophers and the theme of (in)equality in one of their popular philosophical works:

- Thomas Piketty wrote 'Capital in the Twenty-First Century' in 2013. As is mentioned in the texts of the lecture: "The book sharply criticizes what is felt by many as a threat and a curse in an era of globalization, namely growing inequality." It focuses on (in)equality, but mainly from an economic point of view. There are, however, possible obstacles that need to be acknowledged with regard to this work as theme for the lecture in particular. As opposed to the other two works, this is fairly recent. Because of this, there is an increased chance that the book might be known or even read by some. Additionally, the situation in the book is still similar to today's situation. Prior knowledge and current relevance might have potential to significantly influence the results, so caution should be exercised.
- Jean-Jacques Rousseau wrote 'Discourse on the Origins and Foundation of Inequality among Mankind', commonly known as 'The Second Discourse', in 1755. The topic as stated in the texts of the lecture follows "a radical hypothesis: people in the eighteenth century are miserable, society is corrupted and decadent – and the root of it all is inequality." This work approaches (in)equality from a societal and economic perspective.
- Alexis de Tocqueville published the two volumes of his work 'Democracy in America' in 1835 and 1840. In it, he takes a look at (in)equality and its aspects in the rising democracy of the United States at the time. As opposed to the other two works, de Tocqueville's book sheds light on (in)equality from a political point of view. It needs to be acknowledged that, although all works focus on (in)equality, this work might be an outlier. Due to the political approach of this work, especially focused on democracy, the lecture might be perceived as easier than the others due to being more relatable. Nevertheless, we believe, given the way the lecture is written, this different subtheme will not significantly influence the experiments. The main theme, i.e., (in)equality, remains the same.

2.1.2 Recorded lectures

The next step consisted of recording the lectures. Once again, it was crucial that all three videos were comparable. The recording took place in a recording studio at the University of Antwerp. Professor dr. Frank Albers of the University of Antwerp was asked to read aloud all three texts while being recorded. It is important to note that these lectures, as opposed to actual lectures, are not entirely spontaneous as the texts were written out in advance and merely had to be read aloud. Notably, the professor does deviate a little from the texts itself. These deviations, however, are minute and not considered large enough to influence the matter. The differences also mainly concern quotes and citations as the professor has a tendency to read aloud quotation marks or approach a quote by adding "according to [Name]" or similar wordings. With regard to the lectures[1], we distinguish two different categories in terms of comparability: sound and image. The two are discussed in the following paragraphs in order.

[1] Although the experimental order was completely counterbalanced in each experiment, the order in which the lectures are discussed in this thesis will, for consistency, always be Piketty first (referred to as P1), Rousseau second (referred to as R2) and de Tocqueville last (abbreviated to Tocqueville and referred to as T3)

In terms of sound, we consider both the narration, e.g., reading speed in words per second (WPS), intonation, articulation, as well as additional sounds, like coughing, stuttering or exterior noise. The professor has a relatively constant intonation, neither monotonous nor ‘singing-like’, and clear articulation in all three lectures. As he merely had to read the written lectures out loud, he never stutters and rarely has to correct himself. Aside from these presentation characteristics, it is also essential that the speech rate matches for all three videos. Table 1 shows the general data for every video. As can be seen, every video is approximately seven minutes long. Each video starts with a six-second slide stating “[Philosopher’s name] on (in)equality”, meaning the actual length of the text reading is six seconds shorter. Table 1 also shows the slightly altered word counts, including the pronounced quotation marks and slight variations. As can be seen in the second to last column, the average words spoken per second (WPS) for each video is approximately the same. To ensure the speech rate remained constant, the videos were segmented in 30 second intervals, measuring the average WPS every 30 seconds. The last column shows the overall average of these 30 second intervals, which are near 1.95 WPS. In conclusion, it can be said that the lectures should be sufficiently similar to be used in the within-subjects design.

Table 1: General video data

		Length (Start-End = Actual Length)	Words	Overall WPS	Interval WPS
1	**Piketty**	7:21 (0:06 – 7:14 = 7:08)	817	1.90888	1.95000
2	**Rousseau**	7:08 (0:06 – 7:06 = 7:00)	829	1.96912	2.03571
3	**Tocqueville**	7:24 (0:06 – 7:24 = 7:18)	797	1.81963	1.91905

Image was less straightforward. Visually, there were the lectures in three different styles and the subtitles in two different languages. To create comparable visual material, the different components were developed in steps. The specific consecutive steps briefly discussed here are the recording of the talking head lecture, the production of the subtitle tracks and the creation of the PowerPoint presentations.

As a first step, we recorded a talking head lecture of professor dr. Frank Albers (both audio and video). Each lecture was recorded on the same day, meaning professor dr. Frank Albers looks the same and wears the same attire in every lecture. The background is plain black in each talking head lecture as well. Furthermore, supportive hand gestures are kept to a minimum at all times. This makes for three similar talking head lectures with minimal distractions. While this could arguably have an impact on the ecological validity of these lectures, they were still deemed representative of an average lecture with the benefit of having minimal other possible influences.

The second step was the production of the subtitle tracks. This was a thorough process to ensure comparability across different-language and same-language subtitle tracks. Initially, the English subtitle tracks for each lecture was created following a set of predetermined subtitling rules. Based on these tracks alongside the source lectures, the Dutch subtitle tracks were produced. The English subtitle tracks were then slightly edited to be more comparable to the Dutch subtitle tracks. For a more detailed picture of the process, we refer to Paper 2 in Chapter 4.

The last step in the preparation of visual material was the creation of PowerPoint presentations. The PowerPoints were created based on the already recorded lectures and original texts, meaning we designed the PowerPoints without the collaboration of professor dr. Frank Albers. The English PowerPoint template of the University of Antwerp was employed as a base design to make it representative of a real lecture. Calibri Bold was used as title font and Calibri was used as text font in each PowerPoint. We looked for logical breaks in the text and decided, based on the text analyses, to allow seven slides in each presentation. The first three slides were similar for all three lectures, namely a first six-second slide with the title of the lecture; a second slide with an image of the philosopher and some personalia (two bullets); and a third slide with the title of the book discussed in the lecture and some information (two bullets and one quote). The fourth and fifth slide of each presentation consisted of a slide with a chart or table and a regular slide with four bullets. Depending on the lecture, the order of these two slide designs changed. The sixth slide in each presentation contained a number of bullets and an arrow to indicate either causality, contrast or conclusion. The seventh and final slide in each presentation only had bullets and finished with a quote. The specific design of each PowerPoint was research-driven (Bateman et al., 2017; Carter, 2013; Evergreen, 2018; Kosslyn et al., 2012; Raybould, 2015) to ensure high-quality PowerPoints. We also made sure none of the PowerPoints offered an unfair advantage over another towards the comprehension test. An overview of the slide type, word count, duration and resulting WPS can be found in Table 2 below.

Table 2: PowerPoint slide data

		Piketty	**Rousseau**	**Tocqueville**
Slide 1	Type:	Title	Title	Title
	Word Count:	4	5	5
	Duration (seconds):	5	5	5
	WPS:	0.8	1	1
Slide 2	Type:	Image + Bullets	Image + Bullets	Image + Bullets
	Word Count:	10	17	8
	Duration (seconds):	31	32	31
	WPS:	0.323	0.531	0.258
Slide 3	Type:	Bullets	Bullets	Bullets
	Word Count:	36	44	45
	Duration (seconds):	76	72	75
	WPS:	0.474	0.611	0.6
Slide 4	Type:	Table	Bullets	Bullets
	Word Count:	36	23	36
	Duration (seconds):	92	55	100
	WPS:	0.391	0.418	0.36
Slide 5	Type:	Bullets	Table	Table
	Word Count:	26	25	21
	Duration (seconds):	67	63	67
	WPS:	0.388	0.397	0.313
Slide 6	Type:	Arrow + Bullets	Arrow + Bullets	Arrow + Bullets
	Word Count:	28	32	18
	Duration (seconds):	96	122	81
	WPS:	0.292	0.263	0.222
Slide 7	Type:	Bullets	Bullets	Bullets
	Word Count:	56	33	52
	Duration (seconds):	74	79	86
	WPS:	0.757	0.418	0.605

2.2 Psychometric questionnaire

This project employed the validated psychometric questionnaire of Leppink and van den Heuvel (2015). This measure was already discussed in Section 3.3 of Chapter 2. This specific measure was chosen as it had been shown to distinguish extraneous and intrinsic cognitive load. The questionnaire consisted of eight statements which needed to be filled in after each viewing of a lecture. These statements corresponded to the participants' perceptions of the mental effort that the lectures required. The first four statements concerned the difficulty and effort needed to understand the content, i.e., the intrinsic load. The last four concerned the effort associated with the presentation of the content, i.e., the extraneous load. All participants were required to answer each statement on a 10-point Likert scale going from 'not at all the case' to 'absolutely the case'. The use of this questionnaire could give us insight into perceived intrinsic and extraneous loads for each lecture, vital when we want to examine the effects of the different subtitle conditions and lecture styles in the experiments. We did not adapt the questionnaire to cater to our specific experimental conditions as to not affect the validity of the questionnaire. We would, however, recommend future studies to do so because the wording in the questionnaire can lead to misunderstandings and potentially influence the results if it is not used carefully.

2.3 Biographical survey

At the start of every experiment, a biographical survey was used to examine the participants' backgrounds. This survey contained a variety of questions that were mainly aimed at ensuring the participants in each experiment were similar. Throughout the project, the biographical survey remained largely the same, with only minor changes and a translated version for the study conducted with a native English-speaking audience in Australia. The biographical survey asked questions about the participants' sex, year of birth, native language, highest obtained degree, language of instruction in primary and secondary school, use of subtitles at home, experience, prior knowledge and interest in philosophy, and prior knowledge of the philosophers and the topic of the lecture (i.e., (in)equality). Additionally, the biographical survey also asked the participants to rate their own reading and listening skills for English on the Common European Framework of Reference for Languages (CEFR) based on the reference level descriptions (COE, 2020). For the experiments conducted in Belgium, the survey also asked whether participants were studying or had studied English at a higher education institution and how frequently they used English at home, at work, in social contexts or any other places/contexts, which most frequently were when watching series or playing games. We refer to the Appendix at the end of this thesis for a full version of the biographical survey.

2.4 Methodologies

This thesis is composed of five published/to be published papers. Each of these papers discuss one or more experiments conducted as part of this project. The methodology of these experiments was adapted to fit the aim of the study. For purposes of clarity, the following paragraphs briefly discuss the papers, their methodologies and which research questions they attempt to answer (see Table 3 for an overview of the research questions).

The first paper discusses two experiments, both of which employed a repeated measures design. Participants watched all three lectures in a talking style only and without subtitles. These experiments were aimed at preparing the material and analyzing comparability and thus do not answer any research questions but serve as a contribution to the methodologies within AVT research.

The second paper also discusses two experiments with repeated measures designs. In the first experiment, participants watched all three lectures in a talking head style with English subtitles. In the second experiment, all lectures were watched with Dutch subtitles. These experiments were also aimed at preparing the material, specifically the subtitles, and thus serve as a contribution to the methodologies within AVT research.

The third paper discusses the quantitative component of a major experiment conducted at the University of Antwerp. It employed a 2 x 3 mixed-methods design (2 styles of lectures – between-group: talking head, voice-over PowerPoint presentation x 3 subtitle conditions –

within-subject: no subtitles, intralingual/L2 English subtitles, interlingual/L1 Dutch subtitles). It aimed to answer research questions about the effect of the presence and language of subtitles, and the effect of lecture styles on attention allocation, perceived cognitive load and comprehension.

The fourth paper discusses the qualitative component of the same experiment conducted at the University of Antwerp. The data of this experiment were split across two papers due to the sheer amount of data gathered in the study. It employed the same 2 x 3 mixed-methods design, but aimed at answering research questions about the students' perceptions of subtitles in lectures and different lecture styles. These data were gathered in a post-hoc interview.

The fifth and last paper discusses the data from a second major experiment conducted at Macquarie University. It employed a three-factorial repeated measures design, which compared a talking head lecture, a PowerPoint presentation lecture and a composite/picture-in-picture lecture, each of which had English subtitles. Using a more accurate eye-tracker, it aimed to answer research questions about reading and viewing behavior, alongside the similar questions in Paper 3 excluding the specific questions about the effects of subtitle presence and language. In addition, it also serves as a comparison between L1 and L2 speakers of English, although this is not directly addressed in the paper.

Table 3: Overview of papers and research questions

	Paper 1 Chapter 4	**Paper 2 Chapter 4**	**Paper 3 Chapter 5**	**Paper 4 Chapter 5**	**Paper 5 Chapter 6**
Methodology contribution	X	X			
RQ.1 – Cognitive load			X		X
RQ.2 – Comprehension			X		X
RQ.3 – Attention allocation			X		
RQ.4 – Reading					X
RQ.5 – Opinions				X	

3 References

Bateman, J. A., Wildfeuer, J., & Hiippala, T. (2017). *Multimodality: Foundation, research and analysis: A problem-oriented introduction*. Berlin, Germany: De Gruyter/Mouton.

Carter, M. (2013). *Designing science presentations: A visual guide to figures, papers, slides, posters, and more*. London, England: Academic Press.

COE. (2020). Global scale – Table 1 (CEFR): Common reference levels. https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale

Leppink, J., & van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education*, *4*(3), 119–127. doi:10.1007/s40037-015-0192-x

Evergreen, S. D. H. (2018). *Presenting data effectively: Communicating your findings for maximum impact* (2 ed.). London, England: Sage.

Kosslyn, S. M., Kievit, R. A., Russell, A. G., & Shephard, J. M. (2012). PowerPoint presentation flaws and failures: A psychological analysis. *Frontiers in Psychology, 3*. doi:10.3389/fpsyg.2012.00230

Raybould, S. (2015). *Presentation genius: 40 insights from the science of presenting*. London, England: Hodder & Stoughton.

Chapter 4 – Pilot study

The experimental design of the studies conducted as part of the research project recommends comparable materials. Though statistical testing (mixed modelling) can account for random effects arising from the material that was used in a study, the remainder of this thesis will make clear that research can benefit greatly from using comparable materials. In order to attain comparability, (extensive) pretesting is required. This chapter contains two papers that discuss pilot studies conducted as part of the project. A total of four pilot studies have been conducted as part of the project. While, admittedly, this is a little excessive. We still highly recommend conducting at least two small-scale pilot studies to prepare materials and provide some initial insight into the research topic. The first paper in this chapter focuses on the preparation of comparable comprehension tests. The second paper details how comparable subtitles can be produced. Together these two papers bring forward a ten-step approach for preparing experimental AVT research.

Paper 1 – Methodological Preparation of a Within-Subject Audiovisual Cognition, Reception and Perception Study

This paper was published in the Journal of Audiovisual Translation (2022):

Van Hoecke, S. M., Schrijver, I., & Robert, I. S. (2022). Methodological preparation of a within-subject audiovisual cognition, reception and perception study. *Journal of Audiovisual Translation*, *5*(1), 94–128.

Formatting has been slightly altered to fit the present thesis.

Abstract

In the past decade, cognitive empirical AVT research has been on the rise. The majority of these studies are between-subject studies, focused on subtitles for the deaf and hard of hearing (SDH). The few experimental studies that are aimed at other audiences tend to have small sample sizes. Within-subject studies are rarely used in experimental AVT cognition, reception and perception research, although they can increase statistical power due to the repeated testing and shed light on the idiosyncratic nature of the matter. This paper pleads for the introduction of complementary within-subject designs by illustrating the contrasts between the within-subject and between-subject research design. Drawing from the broader spectrum of Translation Studies and the case of the Subtitles for Access to Education (S4AE) research project, this paper highlights obstacles in the preparation of a within-subject AVT cognition, reception and perception experiment and proposes a possible approach to prepare similar within-subject AVT studies.

Key words

Audiovisual translation (AVT), research design, within-subject design, methodology, cognition, subtitle reception, subtitle perception

1 Introduction

Audiovisual translation (AVT) has become a booming and multi-faceted research field over the past decades (Díaz Cintas, 2020). The start of the new millennium saw the emergence of cognitive and empirical AVT studies, which tend to focus on subtitles for the deaf and hard of hearing (SDH) and audio description (Díaz Cintas, 2020). Experimental research into the reception of AVT for other audiences and purposes other than language learning remains scarce (Díaz Cintas, 2020; Díaz Cintas & Szarkowska, 2020). However, as Díaz Cintas and Szarkowska (2020) point out, there is a need for such experimental research as it not only allows us to test new practices, but also enables us to verify old assumptions and theories. This research could "feed back straight into professional practices and processes" (Díaz Cintas, 2020, p. 222). These scholars also underline the importance of sound methodologies, replicability and reproducibility in said research.

Adhering to the aforementioned importance of methodological transparency, replicability and reproducibility, the aim of this paper is to present the methodological preparation of a large-scale, within-subject (repeated measures) study into the reception and perception of and cognitive load posed by subtitles, the so-called Subtitles for Access to Education (S4AE) project. This article follows in the footsteps of a number of publications that lay out possible methods and methodologies or recommend certain approaches for experimental AVT reception research (e.g., Doherty & Kruger, 2018; Kruger et al., 2016; Kruger et al., 2015). Another important precursor is the position paper by Orero et al. (2018), which can be used as a solid guideline for research as it lists many previously conducted AVT studies, proposes numerous measurement tools and recommends various approaches and research designs. One design, however, receives relatively little attention in these publications, namely the within-subject design. What is more, within-subject designs appear to be scant in AVT cognition, reception and perception studies (for brevity purposes: AVTCRP studies) as a whole, with exceptions such as Jensema et al. (2000), Tsaousi (2016), Montero Perez (2019) and Liao et al. (2020). Slightly more frequent is the use of mixed designs, including both within-subject and between-subject components, e.g., Orrego-Carmona (2015), Gerber-Morón and Szarkowska (2018) and Szarkowska and Gerber-Morón (2018, 2019). These are, however, also limited in number. This article aims to shed light on the advantages and drawbacks of a within-subject design and the possible challenges that arise when preparing such a study.

This paper is structured as follows: Section 2 elucidates the contrast between within-subject and between-subject designs, based on literature sourced from the broader field of Translation Studies. In Section 3, the design of the S4AE project and the methodological preparations are explained in detail. The paper concludes with some methodological recommendations for future within-subject studies as well as a discussion of some limitations in our study.

2 Designs in Experimental AV Cognition, Reception and Perception Studies

The design of any experimental study is determined according to the main research question. Balling and Hvelplund (2015) classify three types of research design: (a) an independent (or between) groups design, comparing two groups; (b) a within-subject (repeated measures) design, examining the same group in various conditions; and (c) a functional relations design, focusing on relations between variables rather than participants' behavior in various conditions. Combinations of these designs, mixed designs, are also possible. In this paper, we will mainly focus on the repeated measures design, contrasting its characteristics with the between-subject design. We chose this focus as we expect most readers to be familiar with between-subject designs, but not necessarily with within-subject designs, especially given the scarcity of such designs in experimental AVT research. For the basis of this paper, we draw from both research in AVT as well as from the broader field of Translation Studies.

A between-group (or between-subject) design is commonplace in AVTCRP studies. It tests different participants in various conditions or in one condition. There are numerous ways to plan a between-subject design, ranging from using a test group and control group in a regular and doctored condition (e.g., Bisson et al., 2014; Kruger & Steyn, 2014; Montero Perez, 2020; Szarkowska et al., 2011) to testing of participants by comparing conditions without control groups (e.g., Moreno & Mayer, 2002; Perego et al., 2010; Vulchanova et al., 2015). In contrast, a within-subject (or repeated measures) design is an experimental design in which the same participants are tested a number of times. Again, the specifics may vary depending on the research goal. Researchers can, for example, test the same participants in multiple conditions to examine how varying situations influence the participants (e.g., S4AE project, see Section 3.1) or they can compare before and after data in one condition (e.g., Montero Perez, 2019). All tests may take place in one session (e.g., pilot tests of the S4AE project) or may span over a longer period of time to assess developments (e.g., Moreno et al., 2011).

These designs have various contrasting advantages and disadvantages. The largest advantage of a within-subject design is the mitigation of variability due to the same participants being used for each condition (Mellinger & Hanson, 2017, p. 137). As a result of this lowered variability, the number of participants required to make reliable conclusions is smaller as well, which may be interesting for participant recruiting and possible recruiting costs as well. Between-subject designs are limited in their ability to account for differences between participants, which reduces statistical power in the case of smaller sample sizes (Mellinger & Hanson, 2017). Díaz Cintas (2020, p. 7) stated that limited sample sizes are a present problem in the few experimental AVTCRP studies that are not focused on SDH. Complementary within-subject designs could therefore be a possible means to increase validity and reliability in experimental AVT research. Though the repeated testing might increase internal validity (i.e., accurate measurement), and reliability (i.e., experimental replicability and reproducibility) to some extent by repeatedly confirming certain findings, revealing patterns or showing consistency, it reduces external validity, i.e., ecological validity, as it is evidently conducted in a more experimental setting compared to a between-subject study (Frey et al., 1991; Saldanha & O'Brien, 2013, p. 33). The mitigation of personal variability can also be of benefit for the idiosyncratic nature of particular research topics, such as perception and cognition, which we expect to be different for every individual. Within-subject designs could filter out any of these undesired individual influences and could, in combination with biographical surveys or participant profiling, also help identify influencing factors. In sum, within-subject designs would be a viable option to strengthen studies with smaller samples and mitigate, and possibly identify, influences resulting from personal differences. These two advantages have already been highlighted by Bernardini in 2001, when she addressed the frequent use of between-subject designs in TAP (Think-Aloud Protocols) based translation process research, often conducted with a very limited number of participants. Another advantage of a within-subject design is that it generally does not require control groups, which reduces the chance of contamination. Contamination occurs when an experimental group (un)intentionally passes on essential information about the experiment to the control group or vice-versa, which may mask the actual effects of what is tested. The reduced chance of contamination in within-subject designs can be considered a substantial advantage. It is nevertheless difficult to estimate how realistic and/or frequent this risk of data contamination is, since there have not been any reports – to our knowledge – in AVT research.

However, a within-subject design also has a number of drawbacks in contrast to a between-subject design. One contrast is the time required to adequately set up and execute an experiment. As Section 3.2 will reveal, it takes considerable effort to prepare a within-subject study compared to a between-subject study. The repeated testing also lengthens the experiment. Another contrast is that due to the extended length, a within-subject experiment becomes more prone to attrition and data loss (Mellinger & Hanson, 2017, pp. 7, 105). In the case of multiple tests at different points in time, participants may simply not be present for the repeated tests. Additionally, multiple tests increase the chances of data being unusable, especially in the case of eye tracking with poor calibration or low tracking ratios. A third drawback of the repeated testing is the influence of certain confounding variables (Charness et al., 2012). Mellinger and Hanson (2017, pp. 7, 105) distinguish three of these variables: (a) fatigue, (b) order effects, and (c) carryover effects. The multitude of tests can be tiresome for participants, which in turn may lead to decreased concentration and/or motivation, especially in later stages of the experiment. The participants' behavior may also be different dependent on the order of the tests. Carryover effects imply that participants learn and improve over the course of an experiment, e.g., by conversing with one another, reading/watching relevant material (outside the experimental design) or becoming familiar with the way of testing, which may result in higher scores in later stages. Evidently, these confounding variables can significantly influence the results of a within-subject study, whereas they are less important in a between-subject study. One common solution is to employ counterbalancing. Nevertheless, Mellinger and Hanson (2018, p. 16) warn these confounding variables may still be present.

3 Project Preparation

3.1 Research Background, Goal and Design

To introduce the S4AE project, we would first like to illustrate its research background. Following modern globalization and migration, higher education institutions (HEIs) face increasingly multilingual and multicultural audiences. To cater to these audiences, many HEIs are starting to use English as a medium of instruction (EMI) (Wächter & Maiworm, 2014). The introduction of EMI, however, may have a negative impact on comprehension, cognitive load and retention for students less proficient in L2 English. Subtitles might help to overcome these language barriers and make EMI lectures more accessible. However, adding subtitles to the classroom implies that students suddenly must process a new source of visual information alongside the already present audiovisual information from the lecturer, the lecture slides, the whiteboard, etc. This increases the amount of information that needs to be processed and might thus be more cognitively demanding for students. Delving into this matter, the S4AE project builds on three previous studies exploring the effects of subtitles on comprehension and cognition in a standard educational context (Chan et al., 2019; Kruger et al., 2014; Kruger & Steyn, 2014) and aims to answer the following question:

> *To what extent do the presence of subtitles (present/not present), the subtitle language (L1/L2), the level of L2 proficiency and the students' prior knowledge influence (1) the (perception of) cognitive load and (2) the comprehension and retention of an L2 English lecture?*

To answer this question, the S4AE project can build on considerable between-subject AVT research[2] focusing on subtitle processing by the viewer (e.g., Bisson et al., 2014; Colm, 2008, 2009; de Linde & Kay, 1999; d'Ydewalle & De Bruycker, 2007; Kruger, 2013; Hefer, 2013a, 2013b; Kruger et al., 2018; Kruger et al., 2013; Moreno, 2017; Perego et al., 2018; Perego et al., 2010; Gerber-Morón et al., 2018; Perego, Laskowska, et al., 2016; Perego, Orrego-Carmona, et al., 2016), in which various approaches are used, e.g., eye tracking,

[2] Due to space constraints, this body of research will only be briefly discussed in terms of approaches and measurement tools.

electroencephalography, functional magnetic resonance imaging, self-report psychometric questionnaires, dual-tasks and recognition tests. Many of these studies also examine the effects of subtitles on comprehension and retention, which is also of interest when researching the use of subtitles in EMI classrooms. The existing research has predominantly focused on two aspects: (a) movie comprehension/retention (e.g., Bairstow, 2012; Bairstow & Lavaur, 2017; Birulés-Muntané & Soto-Faraco, 2016; Caffrey, 2008, 2009; Kruger, 2013; Lavaur & Bairstow, 2011; Szarkowska & Bogucka, 2019) and (b) comprehension/retention in a classroom context (e.g., Bianchi & Ciabattoni, 2008; Danan, 1992, 2004; Díaz-Cintas & Cruz, 2008; Montero Perez, 2020; Montero Perez et al., 2014; Moreno & Mayer, 2002; Vulchanova et al., 2015). These studies mainly examine language learning and as such mostly use vocabulary tests, language proficiency tests, word or scene recognition tests for retention and multiple-choice comprehension tests.

Interestingly, studies into subtitle processing and the effects of subtitles on comprehension and retention in an educational context that is aimed at content and not language learning seem scarce. We know only of the three studies mentioned earlier (Chan et al., 2019; Kruger et al., 2014; Kruger & Steyn, 2014). These use self-report effort, frustration and comprehension questionnaires, comprehension tests and eye tracking. They also distinguish visual attention from actual subtitle reading, using the Reading Index for Dynamic Texts (RIDT) developed by Kruger and Steyn (2014).

Complementing these three (between-subject) studies, the S4AE project revolves around a central within-subject design. However, following the advice of Mellinger and Hanson (2017, pp. 163–164), we extended the initial within-subject design to include between-group independent variables, which in turn allows us to assess the interactions between cognitive load and comprehension, and student L2 proficiency levels and prior knowledge of the subject as well. The inclusion of these variables does not alter the advantages, disadvantages or necessary preparation of a within-subject study that this paper discusses.

In this design, Dutch (Flemish) students will view three different recorded EMI lectures. These lectures will be provided in three conditions: (a) with intralingual (English) subtitles; (b) with interlingual (Dutch) subtitles; (c) with no subtitles. To minimize fatigue, order and carryover effects (Mellinger & Hanson, 2017, p. 105), the order of the lectures and the conditions will be counterbalanced completely. The students will watch the lectures individually in an eye tracking laboratory. Eye tracking will allow us to measure cognitive load and actual subtitle reading using Kruger and Steyn's (2014) RIDT as a complementary tool. After each lecture, the students will fill out an extended version of the psychometric questionnaire on cognitive load validated by Leppink and van den Heuvel (2015) and, subsequently, a comprehension test. Using both a psychometric self-report questionnaire and eye tracking to assess cognitive load allows triangulation of data from objective and subjective measures, as recommended by Orero et al. (2018). One month after the experiment, all participants will complete the same comprehension tests again to measure retention. The scores on the psychometric questionnaires and comprehension tests, as well as the eye tracking data, will be correlated with the students' biographical data, language proficiency and learning preferences, which will be collected one month prior to the experiment.

Although within-subject designs, and mixed designs for that matter, remove personal variability, they may be prone to influences originating from the materials used in the experiment. Therefore, meticulous preparation, preferably including pre-testing, and analysis of the materials is required. The aim of this paper is to show how this may be carried out.

3.2 The Ten Steps

A number of preparatory steps need to be taken to ensure the use of comparable materials in a within-subject AVTCRP study to safeguard the validity of future results. Based on our own experiences, we propose to divide the initial process of preparation into ten distinct steps listed below:

(a) Careful preparation of materials

(b) Lecture content and feature analyses

(c) First pilot study

(d) Reevaluation

(e) Optimization

(f) Second pilot study

(g) Production of comparable subtitles

(h) Subtitles analyses

(i) Third pilot study with subtitles

(j) Finalisation of materials

In the following paragraphs, the first six steps will be explained in detail, integrating relevant research. Each step will generate results which (if applicable) might be carried over and integrated into the next step. Given the limited scope of this article, we will focus exclusively on the preparation of the lectures (which can be considered source texts) and the comprehension tests (steps a–f). The complex production, analyses and testing of comparable interlingual and intralingual subtitles are beyond the scope of this paper and will be published in a future article.

3.2.1 Careful Preparation of Materials

Comparable materials are of the utmost importance for a within-subject design. In the S4AE project we examine the effect of no subtitles, interlingual (Dutch) subtitles and intralingual (English) subtitles. This implies we need three lectures that are comparable in content and language (complexity), length, style, etc. Content-wise, all three lectures focused on philosophy, which was realistic and viable, since optional courses in philosophy are part of the study program of the intended participants. Professor Frank Albers, philosophy lecturer at the University of Antwerp, wrote three comparable lectures on the views on inequality of three renowned philosophers, Thomas Piketty, Jean-Jacques Rousseau and Alexis de Tocqueville. The lecture texts were subsequently analyzed and recorded (see Section 3.2.2).

In addition to the lecture texts, the measurement tools had to be selected and prepared. We used eye tracking and an (existing) psychometric self-report questionnaire to measure cognitive load (Leppink & van den Heuvel, 2015). This validated questionnaire consisted of eight general questions for which each participant had to rate complexity on a scale from 1 to 10, 1 representing low complexity and 10 representing high complexity. The first four questions asked about content complexity and as such provided insight into the overall perceived intrinsic load. The last four concerned instructional complexity and thus provided data on perceived extraneous load. To measure retention, we used a (repeated) comprehension test. This tool had frequently been used successfully in earlier AVTCRP research (e.g., Lavaur & Bairstow, 2011; Montero Perez et al., 2014). We designed the comprehension tests as if they were exams for a philosophy course. All three tests consisted of 12 questions and had equal numbers of multiple-choice questions, input questions, memory questions and insight questions[3]. Finally, we used a biographical survey and would employ additional tests in the main experiment, e.g., proficiency tests aimed at assessing listening and reading competences in both English and Dutch and supplementary surveys, to accurately examine the participants' profiles, proficiency level and prior knowledge.

3.2.2 Lecture Content and Feature Analyses

The lecture texts were first compared in terms of readability to ensure their comparability[4]. To this end, we used the Flesch Reading Ease, the Flesch-Kincaid Grade Level and the New Dale-Chall. The first two calculate readability based on the average sentence length and the average number of syllables per word. The Flesch Reading Ease gives a score out of 100, for which above 90 is considered very easy and below 30 is considered very hard; the Flesch-Kincaid Grade Level indicates the American grade-school level necessary to be able to read the text. Sentence and word length are considered accurate indicators of readability (Smeuninx, 2018), but to include different measures, we also chose to add the New Dale-Chall formula, which calculates readability based on a list of familiar words and the average sentence length and gives a score ranging from 0 to 10 or above corresponding to a grade level (Table 1 reports the grade level). As shown in Table 1, the texts receive very similar scores and are estimated to be difficult texts aimed at twelfth grade (17–18yo) students.

Table 1: Text Readability Scores

	Piketty (P1)	Rousseau (R2)	Tocqueville (T3)
Flesch Reading Ease	43	41	38
Flesch-Kincaid Grade Level	12	12	12
New Dale-Chall	11-12	11-12	11-12

Source: Author's own work.

We then analyzed the texts using Perego et al.'s (2018) construct for film complexity. These researchers distinguish three types of complexity: (a) structural-informative complexity, i.e., number of cuts as a measure of newly introduced information, pace and total number of one and two-line subtitles, (b) linguistic complexity, i.e., total word count, standardized type-token ratio (TTR), words per minute (WPM), total sentence count and average sentence length, and (c) narrative complexity, i.e., number of film locations, number of characters and number of flashbacks. Structural-informative complexity is not relevant at this stage given the absence of subtitles and cuts. Table 2 shows the relevant indices for linguistic complexity, with the word count and standardized TTR being very similar. Sentence count and length vary, but this is deemed less important as these texts will be recorded as lectures (oral texts). WPM/WPS is discussed below (Table 3). Perego et al. (2018) mention chronology and amount of information as key aspects of narrative complexity. After analyzing the texts, we concluded similar information was presented in a comparable order.

[3] We have not released the comprehension tests yet, as they will still be used in various experiments. Please contact the author for a confidential copy if desired.

[4] Although the experimental order was completely counterbalanced, the order in which the lectures are discussed in this article will, for consistency, always be Piketty first (referred to as P1), Rousseau second (R2) and de Tocqueville last (T3).

Table 2: Linguistic Complexity

	Piketty (P1)	Rousseau (R2)	Tocqueville (T3)
Total word count	797	798	800
Standardized TTR	0.517	0.451	0.469
Total sentence count	48	56	64
Average sentence length in words	16.604	14.25	12.5

Source: Author's own work.

The lectures were subsequently recorded in a recording studio using an identical format. In each of the three lecture recordings, Professor Albers is shown against a black background. This talking head format is, of course, a more artificial setting than a normal classroom environment, i.e., lower external validity, but the research project aims to assess the impact of subtitles in a more controlled environment. Additionally, minimizing the effects of the lecturer also reduces extraneous load and increases information transfer following the coherence effect (Mayer & Moreno, 2003). This may enable the students to read and process the subtitles better, which has been shown to correlate directly with performance (Kruger & Steyn, 2014).

Finally, the lecture recordings were analyzed. Each lecture is approximately 7 minutes long. The professor does not use hand gestures nor does he cough, he has a constant intonation, rarely stutters and has a relatively constant facial expression across all three lectures. One notable difference from the lecture texts is that the professor tends to explicitly mention quotation marks or add various expressions for indirect speech to signalize quotes. This results in a slightly different total number of words in the lecture. Table 3 shows the length of each recording, the adjusted word count, the overall speech rate in words per second (WPS) and the mean speech rate across 14 intervals of 30 seconds in WPS. Based on these aspects, our team considered the lecture recordings comparable.

Table 3: Lecture Recordings Indices

	Piketty (P1)	Rousseau (R2)	Tocqueville (T3)
Recording Length (mm:ss)	7:21	7:08	7:25
Adjusted Word Count	833	855	833
Overall Speech Rate (WPS)	1.9463	2.0357	1.9018
Mean Interval Speech Rate (WPS)	1.9500	2.0357	1.9191

Source: Author's own work.

3.2.3 First Pilot Study

To verify the conclusions drawn from step 2, a first pilot study without subtitles was set up and conducted in May 2018 with 75 2nd-year students of the BA in Applied Linguistics from the University of Antwerp. They all completed the biographical survey, self-report cognitive load questionnaires and the comprehension tests. Eye tracking, pre-testing and post-testing were excluded to focus on the materials themselves and to keep data analysis feasible. For the statistical analyses, we have one within-subject variable with three levels, i.e., the three lectures, and two independent between-group variables with two levels: the study of English (i.e., studying English in their BA or not) and prior knowledge of philosophy (i.e., having followed an optional philosophy course taught by the professor featured in the lectures or not)[5]. We consistently use mixed ANOVAs as these can compare the mean differences between the lectures and take into account the two between-group variables. However, it is important to note that these between-group variables only provide rough indications of the students' profiles based on the biographical survey since extensive pre-testing (which will be done in the main

[5] The exact participant distributions for each of these variables and the relevant mean ratings and scores can be found in the Appendix, Tables 5.1, 6.1, 7.1 and 8.1.

study) was foregone at this stage. Consequently, we mainly focus on the within-subject effects for all participants and will only briefly discuss interactions with these between-group variables.

A number of conclusions could be drawn from this experiment[6]:

Firstly, T3 appears to induce significantly lower total load (mean of all questions in the psychometric self-report) than P1 and R2 for all participants (Appendix, Table 5.3). The same can be observed for intrinsic load (mean of questions 1-4; Appendix, Table 6.3). In contrast, no significant main effects were found for extraneous load (mean of questions 5–8; Appendix, Table 7.2). As far as interaction effects are concerned, we observed a significant interaction effect between total load and between-group philosophy variable (Appendix, Table 5.2), and between extraneous load ratings and philosophy (Appendix, Table 7.2). In terms of between-group effects, those studying English show significantly lower total load ratings than those that do not (Appendix, Table 7.3).

For comprehension, we found a significant main effect of the lectures, but no significant interactions when the between-group variables are considered (Appendix, Table 8.2). It was revealed that participants scored significantly lower on R2 than on P1 or T3 (Appendix, Table 8.3). For the between-group effects, those studying philosophy were found to perform better than those that did not (Appendix, Table 8.4).

In this pilot study, we were mainly interested in the differences regardless of groups, which explains why the comprehension results are particularly problematic. These tests need revising since the lack of comparability might not reside in the lectures but in the comprehension questions themselves. In this light, the overall difference in total load, and consequently intrinsic load, between T3 and the other lectures may also be problematic, since it might indeed hint at a difference between the lectures. However, we believe that data noise could be an issue. By data noise we mean the data produced by participants who did not follow the instructions properly[7], e.g., a participant rating all psychometric questions with the same number just to be done with the experiment or always choosing the first multiple-choice answer in the comprehension tests. We did not verify whether the participants actually watched the videos or followed the instructions and were therefore unable to filter this possibly conflicting, inaccurate or meaningless data. Accordingly, we will first focus on the revision of the comprehension tests and implement some sort of participant surveillance.

3.2.4 Reevaluation

Following the results from the first pilot study, all materials were reevaluated in an attempt to pinpoint a possible cause for the differences. Our team of researchers unanimously agreed that, although T3 could be considered slightly easier content-wise due to it being less philosophical and more focused on political rather than monetary (in)equality, the main problem resided in the comprehension tests and the lack of data noise prevention. Consequently, the need for optimization of the comprehension tests arose.

3.2.5 Optimization

We recomposed the tests in view of our within-subject component. We no longer focused on creating tests similar to actual lecture exams, but instead aimed to strengthen comparability

[6] Due to space constraints, extensive reporting and statistics were omitted in this section but can be found in the Appendix, Tables 5–8.
[7] Not to be confused with noise in eye tracking data which refers to data being unusable due to signal loss, inaccuracy of the eye tracker, etc.

between questions for all lectures, including not only main ideas but also secondary details. Due to a lack of research on how to develop comparable within-subject comprehension tests, we devised our own approach. First, all originally used questions were considered, disregarding scores, to establish so-called matches (i.e., comparable questions across the three tests), using a large number of variables such as question type, answer type, question length, answer length, in-text location of the first mention of the answer, in-text repetition of the answer and "hearing guesses" (i.e., the probability of guessing correctly based on listening to the lecture). If no match could be found for a particular question, it was discarded. If a match could be found between two lectures only, we explored the possibility of creating a similar question for the remaining lecture.[8] Consequently, each test contained twelve questions comparable to the questions in the other two tests. Although this may have eliminated undesirable influences from varying degrees of difficulty in the comprehension tests, we expect a possible increase in order and/or carryover effects (Mellinger & Hanson, 2017) and will verify this in statistical analyses. Lastly, we logged mouse activity to check whether participants watched the entire video and monitored participants more closely to prevent inattentive behavior.

3.2.6 Second Pilot Study Without Subtitles

To test the optimized comprehension tests, we conducted a second pilot study without subtitles in March 2019 with 50 2nd-year students of the BA in Applied Linguistics of the University of Antwerp (33 female; 17 male)[9]. The same within-subject (the lectures) and between-group (English and philosophy) variables from the first pilot study were used. The participants filled in the biographical survey first. Then they watched the three lectures, each time followed by filling in the psychometric questionnaire (Leppink & van den Heuvel, 2015) and the respective comprehension test. As in the first pilot study, mixed ANOVAs were used to analyze the data.

The mean total, intrinsic and extraneous load were relatively similar for all three lectures (Appendix, Tables 9.1, 10.1 and 11.1). Additionally, the average scores for the three types of cognitive load for each lecture individually were very similar to the scores from the first pilot study.

We first analyzed the within-subject effects for total load (Appendix, Table 9.1). Maulchy's Test of Sphericity confirmed spherical data, $X^2(2) = 2.879$, $p = 0.237$; a mixed ANOVA only found a significant interaction effect with the English variable, $F(2, 86) = 5.234$, $p = 0.007$, but no significant main effects were found for all participants, $F(2, 86) = 2.808$, $p = 0.066$ (Appendix, Table 9.2). No between-group effects were revealed either (Appendix, Table 9.3).

For intrinsic load (Appendix, Table 10.1), Maulchy's Test of Sphericity revealed a violation of sphericity, $X^2(2) = 9.018$, $p = 0.011$. With a Greenhouse-Geisser correction for non-spherical data, a mixed ANOVA showed no statistically significant main within-subject effect of the lectures on intrinsic load for all participants, $F(1.676, 72.074) = 2.913$, $p = 0.070$, and no interaction effects (Appendix, Table 10.2). Furthermore, no significant between-group effects were found (Appendix, Table 10.3).

Lastly, we looked at extraneous load (Appendix, Table 11.1). After assuming sphericity, as Maulchy's test of sphericity revealed spherical data, $X^2(2) = 5.082$, $p = 0.079$, a mixed ANOVA indicated that the lectures did not significantly differ for all participants in extraneous load either, $F(2, 86) = 1.581$, $p = 0.212$ (Appendix, Table 11.2). Furthermore, it only showed a

[8] Due to space constraints, in-depth explanations of the question categorization were not included in this paper. If necessary, contact the author for more information.

[9] Distributions in Appendix Tables may vary as some scores or ratings were excluded following the data noise filtering.

significant interaction between the extraneous load ratings and the English variable, $F(2, 86) = 6.567$, $p = 0.002$ (Appendix, Table 11.2). We consider this interaction of the English variable not to be problematic, since more extensive testing of the proficiency level will be done for the main experiment and the extraneous load ratings regardless of groups do not differ significantly. In terms of between-group effects, those studying English perceived a significantly lower extraneous load across the videos, $F(1, 43) = 9.414$, $p = 0.004$, $r = 0.47$, than the others. A significantly lower extraneous load was also found for the philosophy students when compared to those who did not follow any philosophy course, $F(1, 43) = 5.535$, $p = 0.023$, $r = 0.36$ (Appendix, Table 11.3). It could be expected that the English students had fewer problems with a course taught in English (extraneous load) given their proficiency level. Experience with philosophy on the other hand was expected to have an effect on content comprehension (intrinsic load) instead of extraneous load, which was not found.

In contrast to the findings of the first pilot study, we no longer found a difference between the cognitive loads of the three lectures for all participants. However, we found two significant interaction effects with the between-group English variable.

We draw two conclusions from these findings. Firstly, the importance of participant surveillance and double-checking mechanisms to verify the viewing and proper answering of the questionnaires and tests cannot be underestimated. The significant main effect of the lectures on total and intrinsic load in the first pilot study disappeared in the second pilot study. The contrast between these findings, with data noise filtering being the only difference, is striking. Secondly, in similar within-subject studies it is key to accurately assess participant profiles, prior knowledge and language proficiency.

With regard to the comprehension scores, R2 again received the lowest mean score and T3 the highest, with P1 scoring in between (Appendix, Table 12.1). However, a mixed ANOVA revealed no significant main within-subject effects for all participants and no interaction effects (Appendix, Table 12.2). The optimization of the comprehension tests for this particular within-subject experiment clearly helped. Both the cognitive load ratings and comprehension test scores indicate that the lectures and the comprehension tests are comparable.

Despite these already promising results, we decided to improve the comprehension test even further to flatten out minor insignificant differences that may still be present between the tests. Based on several guidelines (Demeuse & Henry, 2004; Professional Testing, 2020), advice from statisticians from the University of Antwerp on test item analyses and parts of the Item Response Theory (Baker, 2001), we decided to disregard a number of questions in the test. Three variables were used to decide which questions to disregard: difficulty, discrimination and reliability.

The difficulty score of a question is based on the percentage of examinees having answered that question correctly. Since the threshold for what is considered to be a difficult or an easy question is arbitrary, we adhere to the guidelines of the University of Antwerp: if less than 10% of the participants answer correctly, the question is considered to be difficult, whereas a question is easy when more than 90% answer it correctly. Questions that are too difficult or too easy would no longer be considered in future testing. The discrimination score reveals whether a question is in line with what is assessed, assuming that an examinee with high overall testing scores has a higher chance of answering a question correctly. If a question tends to be answered correctly more often by examinees who obtain lower overall scores, while the better examinees tend to answer that question incorrectly, that question can be considered to not be discriminating and not in line with what is assessed. In our university guidelines, the discrimination score for each question is calculated by deducting the number of correct answers in the worst scoring 25% of the participants from the number of correct answers in the best scoring 25% of the students and dividing that number by the largest of those two numbers. It is advised that questions with discrimination scores lower than 0.20 are disregarded in future testing. Our university guidelines determine reliability/consistency with the Pearson point-biserial correlation coefficient between the question and the total scores and should ideally be equal to or higher than 0.15. Similar to the discrimination score, this variable reveals whether the question is in line with what the entire test wants to assess.

When a question was flagged for two of the three variables, we decided to disregard the question in further analyses. We chose to exempt questions instead of discarding them to maintain an equal number of questions and safeguard the similarities between the comprehension tests for the three lectures. This eventually led to three twelve-question tests, but for P1 the scores of only 10 distinct questions were considered, for R2 10 questions and for T3 11 questions. When we compared the newly weighted average scores of the three tests (Appendix, Table 13.1), we see highly similar scores. After verifying sphericity with Maulchy's test of sphericity, $X^2(2) = 2.503$, $p = 0.286$, a mixed ANOVA again revealed no significant main within-subject effect for all participants, $F(2, 88) = 0.469$, $p = 0.627$, and no interaction effects (Appendix, Table 13.2).

Although this additional enhancement was not required, it clearly strengthened the similarity of the tests in terms of test scores and can be used as another example of adjusting the materials to benefit comparability in within-subject studies.

Since Mellinger and Hanson (2018, p. 16) warned that there might still be order effects despite having counterbalanced orders, we checked whether psychometric ratings, comprehension scores and recoded comprehension scores for each lecture differed when it was watched first compared to when it was watched second or third. No real pattern in cognitive load or comprehension could be detected (see Table 4).

Table 4: Second Pilot Study – Mean Cognitive Load Ratings & Comprehension Scores Based on Order

	Place in series	Piketty (P1)	Rousseau (R2)	Tocqueville (T3)
Intrinsic Load	1	6.2206	5.5694	4.9667
	2	5.8214	6.0294	5.1471
	3	5.3750	5.1667	5.8438
Extraneous Load	1	3.0000	3.4167	2.6000
	2	3.0357	3.1029	3.1324
	3	3.5000	2.9167	3.0156
Total Load	1	4.6103	4.4931	3.7833
	2	4.4286	4.5662	4.1397
	3	4.4375	4.0417	4.4297
Comprehension	1	51.18%	47.22%	46.06%
	2	44.29%	51.76%	55.08%
	3	53.53%	50.00%	48.30%

Source: Author's own work.

4 Conclusions

The use of within-subject designs is rather scarce in the body of research into AVT cognition, reception and perception. The aim of this article is not to plea for within-subject studies to take over the world of AVTCRP research. As Bernardini did in 2001 for TAP-based research, we advocate for more frequent use of within-subject designs in AVT research, conducted alongside between-subject studies. Within-subject designs could give additional insight into the idiosyncratic nature of perception, cognition and comprehension of AVT, and could increase statistical power in studies with limited sample sizes.

However, to safeguard validity, careful preparation and pre-testing of research materials and experimental set-up (preferably using at least two pilot studies) is key. A within-subject design might minimize characteristic influences due to the same participants being tested, but it also has a higher risk of undesirable influences from the materials or experimental setup. In this paper, we proposed a ten-step preparation of a within-subject AVTCRP study, which may guide or inspire future research. Based on the experience gained in the S4AE project, we can conclude it is rather challenging to develop materials and tools that are comparable in content and language (complexity), style, length, etc. We have also demonstrated the necessity to be cautious of initial subjective or intuitive assessment of comparability and to pre-test materials and measurement tools using objective measures. Additionally, we have shown that, instead of creating new materials or refurbishing measurement tools, there are other options to allow for valid and reliable results, for example, by recoding comprehension test scores based on an approach from educational research. Methodological input from the aforementioned field of education, other domains of Translation Studies (e.g., translation process research) or even other fields (e.g., Psychology) may be useful to guide this pre-testing phase.

We acknowledge that the ten-step proposal needs adaptation dependent on specific research goals, as well as further refinement. We acknowledge limitations in our approach, such as potential bias in the initial preparatory steps and the relatively small participant sub-groups, particularly in the second pilot study. However, we hope that this proposal will spark a disciplinary debate on the use of within-subject (or mixed) design in AVT research and the ways in which methodological preparations can be approached.

Acknowledgements

We would like to thank professor Frank Albers (University of Antwerp) for assisting us in the production of the lectures. We would also like to give our thanks to professor emerita Aline Remael for her help setting up the S4AE project. Lastly, we would like to express our gratitude to the statisticians of the University of Antwerp for their guidance regarding the recoding of the comprehension tests.

5 References

Bairstow, D. (2012). *Rôles des sous-titres dans la compréhension et la mémorisation de films* [The role of subtitles in the comprehension and retention of film] [Unpublished doctoral dissertation]. Paul Valéry University of Montpellier, France.

Bairstow, D., & Lavaur, J.-M. (2017). Sous-titrage, compréhension de films et acquisition de vocabulaire [Subtitles, comprehension of film and vocabulary acquisition]. *Psychologie Française, 62*(3), 249–261. https://doi.org/10.1016/j.psfr.2015.05.001

Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation.

Balling, L. W., & Hvelplund, K. T. (2015). Design and statistics in quantitative translation (process) research. *Translation Spaces, 4*(1), 170–187. https://doi.org/10.1075/ts.4.1.08bal

Bernardini, S. (2001). Think-aloud protocols in translation research: Achievements, limits, future prospects. *Target, 13*(2), 241–263.

Bianchi, F., & Ciabattoni, T. (2008). Captions and subtitles in EFL learning: An investigative study in a comprehensive computer environment. In A. Baldry, M. Pavesi, C. T. Torsello, & C. Taylor (Eds.), *From didactas to ecolingua: An ongoing research project on translation and corpus linguistics* (pp. 69–90). Edizonioni Università di Trieste.

Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PLoS One*, *11*(6). https://doi.org/10.1371/journal.pone.0158409

Bisson, M.-J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics, 35*(2), 399–418. https://doi.org/10.1017/s0142716412000434

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization, 81*(1), 1–8. https://doi.org/10.1016/j.jebo.2011.08.009

Colm, C. (2008). Using pupillometric, fixation-based and subjective measures to measure the processing effort experienced when viewing subtitled TV anime with pop-up gloss. *Copenhagen Studies in Language, 36*, 125–144.

Colm, C. (2009). *Relevant abuse? Investigating the effects of an abusive subtitling procedure on the perception of TV anime using eye tracker and questionnaire* [Unpublished doctoral dissertation]. Dublin City University, Ireland.

Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning, 42*, 497–527. https://doi.org/10.1111/j.1467-1770.1992.tb01042.x

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta, 49*(1), 67–77. https://doi.org/10.7202/009021ar

de Linde, Z., & Kay, N. (1999). *The semiotics of subtitling*. St Jerome.

Demeuse, M., & Henry, G. (2004). L'analyse classique d'items [The classic analysis of items]. In M. Demeuse (Ed.), *Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation* [Introduction to theories and methods of measurement in psychological and educational sciences] (pp. 173–186). Institut d'Administration scolaire de l'Université de Mons.

Díaz Cintas, J. (2020). Audiovisual translation. In E. Angelone, M. Ehrensberger-Dow, & G. Massey (Eds.), *The Bloomsbury companion to language industry studies* (pp. 209–230). Bloomsbury Publishing.

Díaz Cintas, J., & Cruz, M. F. (2008). Using subtitled video materials for foreign language instruction. In J. Díaz-Cintas (Ed.), The didactics of audiovisual translation (pp. 201–214). John Benjamins.

Díaz Cintas, J., & Szarkowska, A. (2020). Introduction: Experimental research in audiovisual translation – Cognition, reception, production. The Journal of Specialised Translation, 33, 3–16.

Doherty, S., & Kruger, J.-L. (2018). The development of eye tracking in empirical research on subtitling and captioning. In T. Dwyer, C. Perkins, S. Redmond, & J. Sita (Eds.), Seeing into screens: Eye tracking and the moving image (pp. 46–64). Bloomsbury.

d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. European Psychologist, 12(3), 196–205. https://doi.org/10.1027/1016-9040.12.3.196

Frey, L. R., Botan, C. H., & Kreps, G. L. (1991). Investigating communication: An introduction to research methods. Pearson.

Gerber-Morón, O., & Szarkowska, A. (2018). Line breaks in subtitling: An eye tracking study on viewer preferences. Journal of Eye Movement Research, 11(3). https://doi.org/10.16910/jemr.11.3.2

Gerber-Morón, O., Szarkowska, A., & Woll, B. (2018). The impact of text segmentation on subtitle reading. Journal of Eye Movement Research, 11(4). https://doi.org/10.16910/jemr.11.4.2

Hefer, E. (2013a). Reading first and second language subtitles: Sesotho viewers reading in Sesotho and English. Southern African Linguistics and Applied Language Studies, 31(3), 359–373. https://doi.org/10.2989/16073614.2013.837610

Hefer, E. (2013b). Reading second language subtitles: A case study of Afrikaans viewers reading in Afrikaans and English. Perspectives, 21(1), 22–41. https://doi.org/10.1080/0907676x.2012.722652

Jensema, C. J., El Sharkawy, S., Danturthi, R. S., Burch, R., & Hsu, D. (2000). Eye movement patterns of captioned television viewers. American Annals of the Deaf, 145(3), 275–285. https://doi.org/10.1353/aad.2012.0093

Kruger, J. (2013). Subtitles in the classroom: Balancing the benefits of dual coding with the cost of increased cognitive load. Journal for Language Teaching, 47(1), 55–71. https://doi.org/10.4314/jlt.v47i1.2

Kruger, J.-L., & Doherty, S. (2016). Measuring cognitive load in the presence of educational video: Towards a multimodal methodology. Australasian Journal of Educational Technology, 32(6), 19–31. https://doi.org/10.14742/ajet.3084

Kruger, J.-L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. Reading Research Quarterly, 49(1), 105–120. https://doi.org/10.1002/rrq.59

Kruger, J.-L., Chan, W. S., & Doherty, S. (2019). Comparing the impact of automatically generated and corrected subtitles on cognitive load and learning in a first- and second-language educational context. Linguistica Antverpiensia, New Series: Themes in Translation Studies, (18), 237–272.

Kruger, J.-L., Doherty, S., Fox, W., & de Lissa, P. (2018). Multimodal measurement of cognitive load during subtitle processing: Same-language subtitles for foreign-language viewers. In I. Lacruz & R. Jääskeläinen (Eds.), Innovation and expansion in translation process research (pp. 267–294). John Benjamins.

Kruger, J.-L., Hefer, E., & Matthew, G. (2013, August 29–31). Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts [Conference presentation]. The 2013 Conference on Eye Tracking South Africa, Cape Town, South Africa.

Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. Journal of Eye Movement Research, 7(5), 1–15. https://doi.org/10.16910/jemr.7.5.4

Kruger, J.-L., Soto-Sanfiel, M. T., Doherty, S., & Ibrahim, R. (2016). Towards a cognitive audiovisual translatology. In R. M. Martín (Ed.), Reembedding translation process research (pp. 171–193). John Benjamins.

Kruger, J.-L., Szarkowska, A., & Krejtz, I. (2015). Subtitles on the moving image: An overview of eye tracking studies. Refractory: A Journal of Entertainment Media, 25, 1–14.

Lavaur, J.-M., & Bairstow, D. (2011). Languages on the screen: Is film comprehension related to the viewers' fluency level and to the language in the subtitles? International Journal of Psychology, 46(6), 455–462. https://doi.org/10.1080/00207594.2011.565343

Leppink, J., & van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. Perspectives on Medical Education, 4(3), 119–127. https://doi.org/10.1007/s40037-015-0192-x

Liao, S., Kruger, J.-L., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load and comprehension. The Journal of Specialised Translation, 33, 70–98.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. Educational Psychologist, 38(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Mellinger, C. D., & Hanson, T. A. (2017). Quantitative research methods in translation and interpreting studies. Routledge.

Mellinger, C. D., & Hanson, T. A. (2018). Order effects in the translation process. Translation, Cognition & Behavior, 1(1), 1–20. https://doi.org/10.1075/tcb.00001.mel

Montero Perez, M. (2019). Pre-learning vocabulary before viewing captioned video: An eye tracking study. The Language Learning Journal, 47(4), 460–478. https://doi.org/10.1080/09571736.2019.1638623
Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. Studies in Second Language Acquisition, 1–25. https://doi.org/10.1017/s0272263119000706
Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. Language Learning & Technology, 18(1), 118–141.
Moreno, A. O. (2017). Attention and dual coding theory: An interaction model using subtitles as a paradigm [Unpublished doctoral dissertation]. Autonomous University of Barcelona, Spain.
Moreno, M. R., Otero-Sabogal, R., & Soto, C. (2011). Using web-based training to improve skills among bilingual dual-role staff interpreters. International Journal of Interpreter Eduation, 3, 22–48.
Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. Journal of Educational Psychology, 94(1), 156–163. https://doi.org/10.1037/0022-0663.94.1.156
Orero, P., Doherty, S., Kruger, J.-L., Matamala, A., Pedersen, J., Perego, E., Romero-Fresco, P., Rovira-Esteva, S., Soler-Vilageliu, O., & Szarkowska, A. (2018). Conducting experimental research in audiovisual translation (AVT): A position paper. The Journal of Specialised Translation, 30, 105–126.
Orrego-Carmona, D. (2015). The reception of (non)professional subtitling [Unpublished doctoral dissertation]. University of Rovira I Virgili, Tarragona, Spain.
Perego, E., Del Missier, F., & Stragà, M. (2018). Dubbing vs. subtitling: Complexity matters. Target, 30(1), 137–157.
Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. Media Psychology, 13(3), 243–272. https://doi.org/10.1080/15213269.2010.502873
Perego, E., Laskowska, M., Matamala, A., Remael, A., Robert, I. S., Szarkowska, A., Vilaró, A., & Bottiroli, S. (2016a). Is subtitling equally effective everywhere? A first cross-national study on the reception of interlingually subtitled messages. Across Languages and Cultures, 17(2), 205–229. https://doi.org/10.1556/084.2016.17.2.4
Perego, E., Orrego-Carmona, D., & Bottiroli, S. (2016b). An empirical take on the dubbing vs. subtitling debate: An eye movement study. Lingue e Linguaggi, 19, 255–274. https://doi.org/10.1285/i22390359v19p255
Professional Testing. (2020). Step 9. Conduct the item analysis. http://www.proftesting.com/test_topics/pdfs/steps_9.pdf
Saldanha, G., & O'Brien, S. (2013). Research methodologies in translation studies. St Jerome.
Smeuninx, N. (2018). Dear stakeholder: Exploring the language of sustainability reporting: A closer look at readability, sentiment and perception [Unpublished doctoral dissertation]. Ghent University, Ghent, Belgium. https://biblio.ugent.be/publication/8576556
Szarkowska, A., & Bogucka, L. (2019). Six-second rule revisited: An eye tracking study on the impact of speech rate and language proficiency on subtitle reading. Translation, Cognition & Behavior, 2(1), 101–124. https://doi.org/10.1075/tcb.00022.sza
Szarkowska, A., & Gerber-Morón, O. (2018a). Viewers can keep up with fast subtitles: Evidence from eye movements. PLoS One, 13(6). https://doi.org/ 10.1371/journal.pone.0199331
Szarkowska, A., & Gerber-Morón, O. (2018b). Two or three lines: A mixed-methods study on subtitle processing and preferences. Perspectives, 27(1). https://doi.org/10.1080/0907676X.2018.1520267
Szarkowska, A., Krejtz, I., Klyswejko, Z., & Wieczorek, A. (2011). Verbatim, standard, or edited? Reading patterns of different captioning styles among deaf, hard of hearing and hearing viewers. American Annals of the Deaf, 156(4), 363–378. https://doi.org/10.1353/aad.2011.0039
Tsaousi, A. (2016). Sound effect labelling in subtitling for the deaf and hard-of-hearing: An experimental study on audience reception [Unpublished doctoral dissertation]. Autonomous University of Barcelona, Spain.
Vulchanova, M., Aurstad, L. M. G., Kvitnes, I. E. N., & Eshuis, H. (2015). As naturalistic as it gets: Subtitles in the English classroom in Norway. Frontiers in Psychology, 5(1510). https://doi.org/10.3389/fpsyg.2014.01510

Wächter, B., & Maiworm, F. (2014). English-taught programmes in European higher education: The state of play in 2014. Lemmens.

6 Appendix

Table 5: First Pilot – Total Load

5.1 - First Pilot Study - Descriptive Statistics					
Lecture	**English**	**Philosophy**	**Mean**	**Standard Deviation**	**N**
Piketty (P1)	No	No	5	1.57784	13
		Yes	5.3125	1.55373	10
		Total	5.1359	1.53968	23
	Yes	No	4.1477	1.20307	22
		Yes	4.3625	1.17689	30
		Total	4.2716	1.18112	52
	Total	No	4.4643	1.39543	35
		Yes	4.6	1.32687	40
		Total	4.5367	1.35173	75
Rousseau (R2)	No	No	5.5481	1.54071	13
		Yes	4.4375	0.98116	10
		Total	5.0652	1.41616	23
	Yes	No	4.5455	0.99892	22
		Yes	4.1875	1.54834	30
		Total	4.3389	1.34386	52
	Total	No	4.9179	1.3022	35
		Yes	4.25	1.42015	40
		Total	4.5617	1.39805	75
Tocqueville (T3)	No	No	4.3269	1.18973	13
		Yes	4.325	1.41446	10
		Total	4.3261	1.26117	23
	Yes	No	3.9489	0.96413	22
		Yes	4.0583	1.36881	30
		Total	4.012	1.20463	52
	Total	No	4.0893	1.05265	35
		Yes	4.125	1.36696	40
		Total	4.1083	1.22239	75

5.2 - First Pilot Study - Test of Within-Subject Effects							
Source		**Type III Sum of Squares**	***df***	**Mean Square**	***F***	**Significance**	**Partial Eta Squared**
Total	Sphericity Assumed	11.645	2	5.822	5.87	0.004*	0.076
*Total * English*	Sphericity Assumed	2.622	2	1.311	1.322	0.27	0.018
*Total * Philosophy*	Sphericity Assumed	8.66	2	4.33	4.366	0.014*	0.058
*Total * English * Philosophy*	Sphericity Assumed	1.536	2	0.768	0.774	0.463	0.011
Error (Intrinsic)	Sphericity Assumed	140.845	142	0.992			

5.3 - First Pilot Study - Post-Hoc Pairwise Comparisons (Bonferroni)						
Total		**Mean Difference**	**Standard Error**	**Significance**	**95% Confidence Interval for Difference**	
					Lower Bound	**Upper Bound**
Piketty (P1)	Rousseau (R2)	0.026	0.176	1	-0.407	0.459
	Tocqueville (T3)	0.541	0.156	0.003*	0.158	0.924
Rousseau (R2)	Piketty (P1)	-0.026	0.176	1	-0.459	0.407
	Tocqueville (T3)	0.515	0.199	0.035*	0.027	1.003
Tocqueville (T3)	Piketty (P1)	-0.541	0.156	0.003*	-0.924	-0.158
	Rousseau (R2)	-0.515	0.199	0.035*	-1.003	-0.027

11.4 - First Pilot Study - Test of Between-Subject Effects						
Source	Type III Sum of Squares	*df*	Mean Square	*F*	Significance	Partial Eta Squared
Intercept	1276.476	1	1276.476	1248.379	0.000*	0.946
English	5.947	1	5.947	5.816	0.018*	0.076
Philosophy	0.302	1	0.302	0.295	0.588	0.004
*English * Philosophy*	0.255	1	0.255	0.25	0.619	0.004
Error	72.598	71	1.023			

Source: Author's own work.

Table 6: First Pilot – Intrinsic Load

6.1 - First Pilot Study - Descriptive Statistics					
Lecture	**English**	**Philosophy**	**Mean**	**Standard Deviation**	**N**
Piketty (P1)	No	No	5.8077	1.54837	13
		Yes	6.725	1.41151	10
		Total	6.2065	1.52936	23
	Yes	No	5.4091	1.49114	22
		Yes	5.6	1.20631	30
		Total	5.5192	1.32366	52
	Total	No	5.5571	1.50255	35
		Yes	5.8813	1.33612	40
		Total	5.73	1.41586	75
Rousseau (R2)	No	No	6.2885	1.45361	13
		Yes	5.775	1.62211	10
		Total	6.0652	1.51548	23
	Yes	No	5.75	1.20515	22
		Yes	5.375	1.83212	30
		Total	5.5337	1.59428	52
	Total	No	5.95	1.30863	35
		Yes	5.475	1.7703	40
		Total	5.6967	1.57969	75
Tocqueville (T3)	No	No	5.1346	1.51594	13
		Yes	5.375	1.7129	10
		Total	5.2391	1.57119	23
	Yes	No	5.1364	1.24098	22
		Yes	4.9917	1.59131	30
		Total	5.0529	1.44196	52
	Total	No	5.1357	1.32751	35
		Yes	5.0875	1.60882	40
		Total	5.11	1.47458	75

6.2 - First Pilot Study - Test of Within-Subject Effects							
Source		Type III Sum of Squares	*df*	Mean Square	*F*	Significance	Partial Eta Squared
Intrinsic	Sphericity Assumed	19.639	2	9.819	7.479	0.001*	0.095
*Intrinsic * English*	Sphericity Assumed	2.551	2	1.275	0.971	0.381	0.013
*Intrinsic * Philosophy*	Sphericity Assumed	7.796	2	3.898	2.969	0.055	0.04
*Intrinsic * English * Philosophy*	Sphericity Assumed	1.484	2	0.742	0.565	0.57	0.008
Error (Intrinsic)	Sphericity Assumed	186.439	142	1.313			

6.3 - First Pilot Study - Post-Hoc Pairwise Comparisons (Bonferroni)						
Intrinsic		Mean Difference	Standard Error	Significance	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Piketty (P1)	Rousseau (R2)	0.088	0.189	1	-0.374	0.551
	Tocqueville (T3)	0.726	0.191	0.001*	0.257	1.195
Rousseau (R2)	Piketty (P1)	-0.088	0.189	1	-0.551	0.374
	Tocqueville (T3)	0.638	0.232	0.023*	0.069	1.206
Tocqueville (T3)	Piketty (P1)	-0.726	0.191	0.001*	-1.195	-0.257
	Rousseau (R2)	-0.638	0.232	0.023*	-1.206	-0.069

6.4 - First Pilot Study - Test of Between-Subject Effects						
Source	Type III Sum of Squares	df	Mean Square	F	Significance	Partial Eta Squared
Intercept	1972.032	1	1972.032	1469.735	0.000*	0.954
English	3.514	1	3.514	2.619	0.11	0.036
Philosophy	0.043	1	0.043	0.032	0.858	0
English * Philosophy	0.411	1	0.411	0.307	0.582	0.004
Error	95.265	71	1.342			

Source: Author’s own work.

Table 7: First Pilot – Extraneous Load

7.1 - First Pilot Study - Descriptive Statistics					
Lecture	English	Philosophy	Mean	Standard Deviation	N
Piketty (P1)	No	No	4.1923	2.01854	13
		Yes	3.9	2.322	10
		Total	4.0652	2.10953	23
	Yes	No	2.8864	1.56739	22
		Yes	3.125	1.69526	30
		Total	3.024	1.63094	52
	Total	No	3.3714	1.83449	35
		Yes	3.3188	1.86996	40
		Total	3.3433	1.84115	75
Rousseau (R2)	No	No	4.8077	2.06447	13
		Yes	3.1	1.51932	10
		Total	4.0652	2.00456	23
	Yes	No	3.3409	1.52469	22
		Yes	3	1.90621	30
		Total	3.1442	1.74709	52
	Total	No	3.8857	1.85934	35
		Yes	3.025	1.79904	40
		Total	3.4267	1.86575	75
Tocqueville (T3)	No	No	3.5192	1.58266	13
		Yes	3.275	1.52046	10
		Total	3.413	1.52556	23
	Yes	No	2.7614	1.46685	22
		Yes	3.125	1.92729	30
		Total	2.9712	1.74098	52
	Total	No	3.0429	1.53331	35
		Yes	3.1625	1.81655	40
		Total	3.1067	1.68016	75

7.2 - First Pilot Study - Test of Within-Subject Effects							
Source		Type III Sum of Squares	*df*	Mean Square	*F*	Significance	Partial Eta Squared
Extraneous	Sphericity Assumed	5.872	2	2.936	2.369	0.097	0.032
*Extraneous * English*	Sphericity Assumed	2.704	2	1.352	1.091	0.339	0.015
*Extraneous * Philosophy*	Sphericity Assumed	11.354	2	5.677	4.58	0.012*	0.061
*Extraneous * English * Philosophy*	Sphericity Assumed	1.669	2	0.835	0.673	0.512	0.009
Error (Intrinsic)	Sphericity Assumed	176.009	142	1.24			

7.3 - First Pilot Study - Test of Between-Subject Effects						
Source	Type III Sum of Squares	*df*	Mean Square	*F*	Significance	Partial Eta Squared
Intercept	731.597	1	731.597	320.49	0.000*	0.819
English	9.018	1	9.018	3.95	0.051	0.053
Philosophy	1.708	1	1.708	0.748	0.39	0.01
*English * Philosophy*	2.728	1	2.728	1.195	0.278	0.017
Error	162.075	71	2.283			

Source: Author's own work.

Table 8: First Pilot – Comprehension

8.1 - First Pilot Study - Descriptive Statistics (Absolute Score)					
Lecture	**English**	**Philosophy**	**Mean**	**Standard Deviation**	**N**
Piketty (P1)	No	No	6.08	1.706	13
		Yes	7.1	2.079	10
		Total	6.52	1.904	23
	Yes	No	6.18	1.967	22
		Yes	6.17	2.019	30
		Total	6.17	1.978	52
	Total	No	6.14	1.849	35
		Yes	6.4	2.048	40
		Total	6.28	1.949	75
Rousseau (R2)	No	No	4.08	0.862	13
		Yes	5.1	2.025	10
		Total	4.52	1.534	23
	Yes	No	5.23	1.744	22
		Yes	5.93	2.164	30
		Total	5.63	2.01	52
	Total	No	4.8	1.568	35
		Yes	5.73	2.136	40
		Total	5.29	1.937	75
Tocqueville (T3)	No	No	6	2.198	13
		Yes	6.3	2.83	10
		Total	6.13	2.437	23
	Yes	No	6.86	1.833	22
		Yes	7.6	1.976	30
		Total	7.29	1.934	52
	Total	No	6.54	1.99	35
		Yes	7.28	2.253	40
		Total	6.93	2.152	75

8.2 - First Pilot Study - Test of Within-Subject Effects							
Source		**Type III Sum of Squares**	***df***	**Mean Square**	***F***	**Significance**	**Partial Eta Squared**
Comprehension	Greenhouse-Geisser	90.913	1.707	53.269	12.419	0.000*	0.149
*Comprehension * English*	Greenhouse-Geisser	22.021	1.707	12.903	3.008	0.061	0.041
*Comprehension * Philosophy*	Greenhouse-Geisser	1.305	1.707	0.764	0.178	0.803	0.003
*Comprehension * English * Philosophy*	Greenhouse-Geisser	4.252	1.707	2.492	0.581	0.535	0.008
Error (Intrinsic)	Greenhouse-Geisser	519.755	121.174	4.289			

8.3 - First Pilot Study - Post-Hoc Pairwise Comparisons (Bonferroni)						
Comprehension		**Mean Difference**	**Standard Error**	**Significance**	**95% Confidence Interval for Difference**	
					Lower Bound	**Upper Bound**
Piketty (P1)	Rousseau (R2)	1.297	0.353	0.001*	0.431	2.163
	Tocqueville (T3)	-0.31	0.393	1	-1.274	0.654
Rousseau (R2)	Piketty (P1)	-1.297	0.353	0.001*	-2.163	-0.431
	Tocqueville (T3)	-1.607	0.268	0.000*	-2.263	-0.95
Tocqueville (T3)	Piketty (P1)	0.31	0.393	1	-0.654	1.274
	Rousseau (R2)	1.607	0.268	0.000*	0.95	2.263

8.4 - First Pilot Study - Test of Between-Subject Effects						
Source	Type III Sum of Squares	*df*	Mean Square	*F*	Significance	Partial Eta Squared
Intercept	2291.919	1	2291.919	1559.916	0.000*	0.956
English	4.786	1	4.786	3.258	0.075	0.044
Philosophy	6.187	1	6.187	4.211	0.044*	0.056
*English * Philosophy*	0.367	1	0.367	0.25	0.619	0.004
Error	104.317	71	1.469			

Source: Author’s own work.

Table 9: Second Pilot – Total Load

9.1 - Second Pilot Study - Descriptive Statistics					
Lecture	**English**	**Philosophy**	**Mean**	**Standard Deviation**	**N**
Piketty (P1)	No	No	5.4583	0.98027	9
		Yes	5.0556	2.07017	9
		Total	5.2569	1.5849	18
	Yes	No	4.2989	1.82343	23
		Yes	2.9792	1.11079	6
		Total	4.0259	1.76883	29
	Total	No	4.625	1.69944	32
		Yes	4.225	1.99955	15
		Total	4.4973	1.78849	47
Rousseau (R2)	No	No	4.7639	1.07972	9
		Yes	4.3889	2.46969	9
		Total	4.5764	1.85907	18
	Yes	No	4.538	1.40991	23
		Yes	3.2917	0.32275	6
		Total	4.2802	1.35812	29
	Total	No	4.6016	1.31233	32
		Yes	3.95	1.95759	15
		Total	4.3936	1.55601	47
Tocqueville (T3)	No	No	4.2917	1.29452	9
		Yes	4.3333	2.03869	9
		Total	4.3125	1.65679	18
	Yes	No	4.0978	1.70684	23
		Yes	3.375	0.67082	6
		Total	3.9483	1.56785	29
	Total	No	4.1523	1.5836	32
		Yes	3.95	1.66489	15
		Total	4.0878	1.5946	47

9.2 - Second Pilot Study - Test of Within-Subject Effects							
Source		**Type III Sum of Squares**	***df***	**Mean Square**	***F***	**Significance**	**Partial Eta Squared**
Total	Sphericity Assumed	3.321	2	1.661	2.808	0.066	0.061
Total * English	Sphericity Assumed	6.19	2	3.095	5.234	0.007*	0.109
Total * Philosophy	Sphericity Assumed	1.525	2	0.763	1.29	0.281	0.029
Total * English * Philosophy	Sphericity Assumed	0.028	2	0.014	0.024	0.976	0.001
Error (Intrinsic)	Sphericity Assumed	50.856	86	0.591			

9.3 - Second Pilot Study - Test of Between-Subject Effects						
Source	**Type III Sum of Squares**	***df***	**Mean Square**	***F***	**Significance**	**Partial Eta Squared**
Intercept	665.07	1	665.07	300.181	0.000*	0.875
English	8.382	1	8.382	3.783	0.058	0.081
Philosophy	4.163	1	4.163	1.879	0.178	0.042
English * Philosophy	1.675	1	1.675	0.756	0.389	0.017
Error	95.269	43	2.216			

Source: Author's own work.

Table 10: Second Pilot – Intrinsic Load

10.1 - Second Pilot Study - Descriptive Statistics					
Lecture	English	Philosophy	Mean	Standard Deviation	N
Piketty (P1)	No	No	6.1111	1.36994	9
		Yes	6.1667	1.82431	9
		Total	6.1389	1.5653	18
	Yes	No	5.6957	2.03235	23
		Yes	5.2917	2.19896	6
		Total	5.6121	2.03385	29
	Total	No	5.8125	1.85785	32
		Yes	5.8167	1.95591	15
		Total	5.8138	1.86826	47
Rousseau (R2)	No	No	5.2778	1.38318	9
		Yes	5.5278	2.35001	9
		Total	5.4028	1.87503	18
	Yes	No	5.9457	1.38981	23
		Yes	5.3333	1.21106	6
		Total	5.819	1.35768	29
	Total	No	5.7578	1.39914	32
		Yes	5.45	1.92075	15
		Total	5.6596	1.56943	47
Tocqueville (T3)	No	No	4.7222	1.38318	9
		Yes	5.6667	2.22907	9
		Total	5.1944	1.86405	18
	Yes	No	5.3152	1.84978	23
		Yes	5.4583	1.22899	6
		Total	5.3448	1.72095	29
	Total	No	5.1484	1.73072	32
		Yes	5.5833	1.84116	15
		Total	5.2872	1.7585	47

10.2 - Second Pilot Study - Test of Within-Subject Effects							
Source		**Type III Sum of Squares**	***df***	**Mean Square**	***F***	**Significance**	**Partial Eta Squared**
Intrinsic	Greenhouse-Geisser	5.138	1.676	3.066	2.913	0.07	0.063
Intrinsic * English	Greenhouse-Geisser	4.568	1.676	2.725	2.589	0.091	0.057
Intrinsic * Philosophy	Greenhouse-Geisser	3.211	1.676	1.915	1.82	0.175	0.041
Intrinsic * English * Philosophy	Greenhouse-Geisser	0.218	1.676	0.13	0.124	0.849	0.003
Error (Intrinsic)	Greenhouse-Geisser	75.862	72.074	1.053			

10.3 - Second Pilot Study - Test of Between-Subject Effects						
Source	**Type III Sum of Squares**	***df***	**Mean Square**	***F***	**Significance**	**Partial Eta Squared**
Intercept	1136.857	1	1136.857	447.087	0.000*	0.912
English	0.048	1	0.048	0.019	0.891	0
Philosophy	0.036	1	0.036	0.014	0.905	0
English * Philosophy	1.158	1	1.158	0.456	0.503	0.01
Error	109.341	43	2.543			

Source: Author's own work.

Table 11: Second Pilot – Extraneous Load

11.1 - Second Pilot Study - Descriptive Statistics					
Lecture	**English**	**Philosophy**	**Mean**	**Standard Deviation**	**N**
Piketty (P1)	No	No	4.8056	0.86402	9
		Yes	3.9444	2.84983	9
		Total	4.375	2.09033	18
	Yes	No	2.9022	2.05567	23
		Yes	0.6667	0.6455	6
		Total	2.4397	2.0601	29
	Total	No	3.4375	1.98685	32
		Yes	2.6333	2.74816	15
		Total	3.1809	2.25886	47
Rousseau (R2)	No	No	4.25	1.05327	9
		Yes	3.25	2.9128	9
		Total	3.75	2.18619	18
	Yes	No	3.1304	1.93189	23
		Yes	1.25	0.80623	6
		Total	2.7414	1.91036	29
	Total	No	3.4453	1.78788	32
		Yes	2.45	2.47162	15
		Total	3.1277	2.05751	47
Tocqueville (T3)	No	No	3.8611	1.51096	9
		Yes	3	2.33854	9
		Total	3.4306	1.96065	18
	Yes	No	2.8804	1.9712	23
		Yes	1.2917	0.8279	6
		Total	2.5517	1.89852	29
	Total	No	3.1563	1.88345	32
		Yes	2.3167	2.02984	15
		Total	2.8883	1.94964	47

11.2 - Second Pilot Study - Test of Within-Subject Effects							
Source		**Type III Sum of Squares**	***df***	**Mean Square**	***F***	**Significance**	**Partial Eta Squared**
Extraneous	Sphericity Assumed	1.976	2	0.988	1.581	0.212	0.035
Extraneous * English	Sphericity Assumed	8.208	2	4.104	6.567	0.002*	0.132
Extraneous * Philosophy	Sphericity Assumed	0.501	2	0.251	0.401	0.671	0.009
Extraneous * English * Philosophy	Sphericity Assumed	0.529	2	0.264	0.423	0.657	0.01
Error (Intrinsic)	Sphericity Assumed	53.745	86	0.625			

11.3 - Second Pilot Study - Test of Between-Subject Effects						
Source	**Type III Sum of Squares**	***df***	**Mean Square**	***F***	**Significance**	**Partial Eta Squared**
Intercept	319.001	1	319.001	96.762	0.000*	0.692
English	31.037	1	31.037	9.414	0.004*	0.18
Philosophy	18.249	1	18.249	5.535	0.023*	0.114
English * Philosophy	2.286	1	2.286	0.693	0.41	0.016
Error	141.761	43	3.297			

Source: Author’s own work.

Table 12: Second Pilot – Comprehension

12.1 - Second Pilot Study - Descriptive Statistics (Absolute Score)					
Lecture	**English**	**Philosophy**	**Mean**	**Standard Deviation**	**N**
Piketty (P1)	No	No	5.8	1.874	10
		Yes	5.67	2.598	9
		Total	5.74	2.182	19
	Yes	No	6.3	2.055	23
		Yes	5.5	2.168	6
		Total	6.14	2.065	29
	Total	No	6.15	1.986	33
		Yes	5.6	2.354	15
		Total	5.98	2.099	48
Rousseau (R2)	No	No	5.4	1.647	10
		Yes	6	1.871	9
		Total	5.68	1.734	19
	Yes	No	4.91	2.021	23
		Yes	6.33	2.733	6
		Total	5.21	2.21	29
	Total	No	5.06	1.903	33
		Yes	6.13	2.167	15
		Total	5.4	2.029	48
Tocqueville (T3)	No	No	6.9	1.663	10
		Yes	5.78	1.787	9
		Total	6.37	1.77	19
	Yes	No	6.13	2.096	23
		Yes	6.67	2.503	6
		Total	6.24	2.149	29
	Total	No	6.36	1.981	33
		Yes	6.13	2.066	15
		Total	6.29	1.989	48

12.2 - Second Pilot Study - Test of Within-Subject Effects							
Source		**Type III Sum of Squares**	***df***	**Mean Square**	***F***	**Significance**	**Partial Eta Squared**
Comprehension	Sphericity Assumed	10.483	2	5.241	2.454	0.092	0.053
Comprehension * English	Sphericity Assumed	0.288	2	0.144	0.067	0.935	0.002
Comprehension * Philosophy	Sphericity Assumed	12.396	2	6.198	2.902	0.06	0.062
Comprehension * English * Philosophy	Sphericity Assumed	6.61	1.91	3.461	1.547	0.22	0.034
Error (Intrinsic)	Sphericity Assumed	187.98	88	2.136			

12.3 - Second Pilot Study - Test of Between-Subject Effects						
Source	**Type III Sum of Squares**	***df***	**Mean Square**	***F***	**Significance**	**Partial Eta Squared**
Intercept	1344.354	1	1344.354	477.098	0.000*	0.916
English	0.024	1	0.024	0.009	0.926	0
Philosophy	0.065	1	0.065	0.023	0.88	0.001
English * Philosophy	0.862	1	0.862	0.306	0.583	0.007
Error	123.982	44	2.818			

Source: Author's own work.

Table 13: Second Pilot – Recoded Comprehension

13.1 - Second Pilot Study - Descriptive Statistics (Percentage)

Lecture	English	Philosophy	Mean	Standard Deviation	N
Piketty (P1)	No	No	0.48	0.18738	10
		Yes	0.4778	0.25386	9
		Total	0.4789	0.21494	19
	Yes	No	0.5217	0.2044	23
		Yes	0.4833	0.24833	6
		Total	0.5138	0.20997	29
	Total	No	0.5091	0.19743	33
		Yes	0.48	0.2426	15
		Total	0.5	0.21037	48
Rousseau (R2)	No	No	0.5	0.18257	10
		Yes	0.5667	0.15	9
		Total	0.5316	0.16684	19
	Yes	No	0.4435	0.21068	23
		Yes	0.5833	0.25626	6
		Total	0.4724	0.22344	29
	Total	No	0.4606	0.20146	33
		Yes	0.5733	0.19074	15
		Total	0.4958	0.20312	48
Tocqueville (T3)	No	No	0.5545	0.13853	10
		Yes	0.4545	0.17604	9
		Total	0.5072	0.16125	19
	Yes	No	0.4822	0.18678	23
		Yes	0.5455	0.22268	6
		Total	0.4953	0.19221	29
	Total	No	0.5041	0.17471	33
		Yes	0.4909	0.19376	15
		Total	0.5	0.17889	48

13.2 - Second Pilot Study - Test of Within-Subject Effects

Source		Type III Sum of Squares	*df*	Mean Square	*F*	Significance	Partial Eta Squared
Recoded Comprehension	Sphericity Assumed	0.02	2	0.01	0.469	0.627	0.011
Recoded Comprehension * English	Sphericity Assumed	0.009	2	0.005	0.215	0.807	0.005
Recoded Comprehension * Philosophy	Sphericity Assumed	0.095	2	0.048	2.19	0.118	0.047
Recoded Comprehension * English * Philosophy	Sphericity Assumed	0.047	2	0.024	1.089	0.341	0.024
Error (Intrinsic)	Sphericity Assumed	1.913	88	0.022			

13.3 - Second Pilot Study - Test of Between-Subject Effects

Source	Type III Sum of Squares	*df*	Mean Square	*F*	Significance	Partial Eta Squared
Intercept	9.792	1	9.792	383.85	0.000*	0.897
English	0	1	0	0.007	0.934	0
Philosophy	0.004	1	0.004	0.172	0.68	0.004
English * Philosophy	0.011	1	0.011	0.415	0.523	0.009
Error	1.122	44	0.026			

Source: Author’s own work.

Paper 2 – Preparing and comparing subtitles for quasi-experimental and experimental research in audiovisual translation studies

This paper was published in Translation Spaces (2022):

Van Hoecke, S. M., Schrijver, I., & Robert, I. S. (2022). Preparing and comparing subtitles for quasi-experimental and experimental research in audiovisual translation studies. *Translation Spaces, 11*(1), 113–133.

Formatting has been slightly altered to fit the present thesis.

Abstract

Empirical research on cognitive processing in AVT has been on the rise in recent years. A number of overarching works have recommended more standardized approaches and methodological frameworks to contribute to more streamlined, replicable, reproducible and valid future AVT research. To date, the issue of comparability of research materials (e.g., clips, subtitle tracks, comprehension questionnaires) and, more specifically, how to achieve comparability in quasi-experimental and experimental studies, particularly those involving repeated measures, has received little attention. This paper aims to address this knowledge gap by proposing a common-sense ten-step preparatory process for quasi-experimental and experimental subtitling studies. This preparatory process has previously been used in the S4AE project. The paper will focus on the final four steps, consisting of the preparation and comparison of multiple subtitle tracks. These steps were conceptualized taking into account the present research on subtitle parameters and the obstacles encountered while preparing comparable subtitle tracks.

Key words

audiovisual translation (AVT), subtitling, methodology, cognition, comparability

1 Introduction

The need for subtitling is growing rapidly following the increased importance of overall accessibility, inclusivity and equity, the commercial pressure to reach larger, multilingual and multicultural audiences, and legal measures, such as the recent EU Accessibility Act and the renewed EU Audiovisual Media Service Directive in Europe. Consequently, research into subtitling and audiovisual translation (AVT) in general has never been more relevant. Over the years, studies into AVT has examined a wide range of topics, but there is one relatively new focus that might be of great interest for the sudden surge in practical use of subtitling: empirical research on cognitive processing in AVT, or what Díaz Cintas (2020) calls cognitive and empirical AVT studies. Such empirical research can allow us to effectively test the impact of new practices and verify old assumptions and theories, provided it is scientifically sound, replicable and reproducible. In view of this scientific robustness, Orero et al. (2018) recommend standardized experimental approaches and methodological frameworks, which are, however, still largely missing in the field. There is a body of overarching works that list various methodologies for experimental AVT reception research (e.g., Doherty & Kruger, 2018; Kruger et al., 2016; Kruger et al., 2015; Orero et al., 2018). These position papers refer to many previously conducted AVT studies and recommend approaches, measurement tools and research designs. While these papers can be used as guidelines for more streamlined future AVT research, little attention is devoted to comparability of research materials, e.g., subtitles, video clips or comprehension test questions. The production of comparable materials is not addressed in detail in these papers, with mentions of comparability being limited to "If various fragments are compared, they should be similar in terms of complexity, speech rate, genre, etc. so as not to create confounding variables" (Orero et al., 2018, p. 112). Such quasi-experimental or experimental subtitling studies using several tasks and/or measurements in time can provide valuable insight into AVT, provided they are carefully thought out and meticulously prepared (Van Hoecke et al., 2022a).

This article proposes a common-sense ten-step process to prepare a quasi-experimental or experimental subtitling study that involves multiple conditions, tasks and/or measurements in time, requiring comparable research materials (e.g., subtitle tracks or clips). Step 1 to 6 of this process, which consist of preparing and comparing materials (e.g., comprehension tests, video fragments, etc.), and validating the materials have been previously discussed in Van Hoecke et al. (2022a). In this article, we will focus specifically on steps 7 to 10, which concern the process of preparing multiple comparable interlingual and intralingual subtitle tracks. In these steps, we have given priority to "comparability over quality".

By way of introduction, the article first draws up a theoretical framework in Section 2, which gives an overview of subtitle parameters that are expected to be relevant for the production of similar subtitles. As the ten-step process was developed within the Subtitles for Access to Education (S4AE) project, Section 3 first contextualizes the project and briefly discusses its methodology. Section 4 then continues with the subtitling process, after which the paper concludes with key points from the process in Section 5.

2 Theoretical Framework

To ensure the quality of subtitles, scholars like Karamitroglou (1998) and Ivarsson and Carroll (1998) prescribed general subtitling guidelines for practitioners. Though these works were not based on empirical data or scientific research, they are still widely seen as seminal in the profession. Such guidelines do not, however, address language-specific issues. Consequently, some broadcasting companies, businesses or streaming services can be seen developing their own adapted guidelines (e.g., BBC or Netflix). In view of this article's aim, a number of conventions were selected that are of great importance for subtitling (Gottlieb, 2012) and that have direct implications for comparability between different subtitle tracks, namely (1) reading speed; (2) reduction; (3) segmentation; and (4) linguistic complexity.

The first key component is reading speed, also referred to as subtitle speed or presentation rate. Reading speed, which is generally expressed as characters per second (CPS) or words per minute (WPM), is defined as the time an average viewer of a particular audience needs to comfortably read a full two-line subtitle. A full two-line subtitle depends on the maximum CPL. What this exact maximum is, however, is a matter of contention. D'Ydewalle and his colleagues tend to adhere to a maximum of 32 characters and spaces per line in a number of their studies (d'Ydewalle & De Bruycker, 2007; d'Ydewalle, Van Rensberger, & Pollet, 1987), Karamitroglou (1998, p. 2) mentions "around 35 characters" and Kruger et al. (2014) adhere to a maximum of 37 CPL in one of their studies. It is clear that the maximum CPL varies across studies and, as Díaz Cintas and Remael (2014) mention, this is also the case in the media. Most standard television subtitles have a maximum of 37 CPL, in the movie industry the norm seems to be 40 CPL and there are cases where only 33 or 35 characters are allowed. The maximum CPL may fluctuate slightly, what does appear to be a common convention in subtitling is the so-called six-second rule. The six-second rule states that it should be possible to read a full two-line subtitle comfortably in six seconds and that shorter subtitles should be timed proportionally. Based on the study of Díaz Cintas (2003), the ideal CPL for the six-second rule is 72 CPL, which translates into a reading speed of subtitles of 12 CPS or approximately 144 WPM. This ideal reading speed has, however, started to receive some criticism lately. Gottlieb (2012), for example, states that the average reading speeds increased over time, especially in subtitling countries. Various commercial TV stations and the movie industry already adhere to a reading speed of 14–16 CPS and streaming companies, like Netflix, even going up to a maximum of 20 CPS for adult programs in English. While some studies show viewers are also able to cope with these faster subtitle reading speeds of up to 20 CPS (Szarkowska & Gerber-Morón, 2018; Szarkowska & Bogucka, 2019), others reveal an increase in subtitle reading speed might lead to more words being skipped and the viewer skimming the subtitles instead of actually reading them (Kruger et al., 2021; Liao et al., 2021). The reading speed a particular viewer is capable of is likely to be influenced by the degree of habituation to subtitles and the overall language proficiency as well. With this in mind, it is essential to consider the intended audience during the production of comparable subtitles and the careful consideration of the subtitle reading speed for quasi-experimental and experimental studies. In light of this, Fresno and Sepielak (2020) advise to not only consider the average of the subtitle speed for all subtitles, but also the range of speeds.

The second component one needs to consider when testing or ensuring comparability of subtitle tracks is reduction. To cope with the time-space constraints of subtitling, reduction is considered essential (Gottlieb, 2012). Díaz Cintas and Remael (2014) distinguish two types of reduction: total reduction, i.e., deleting irrelevant information, and partial reduction, i.e., reformulating the message. The amount of reduction in a subtitle also dictates the type of subtitles, namely edited, i.e., content is reduced and simplified, verbatim, i.e., all utterances are included, and standard, i.e., content is slightly edited, subtitles. While the discussion of which is better or more inclusive is very much alive in AVT research (Romero-Fresco, 2009; Szarkowska et al., 2011), it is less relevant for this article. However, what is important to keep in mind is that there are different reading patterns and visual attention distributions for each type. A study by Szarkowska et al. (2011), for example, revealed that viewers spent more time watching the image with edited or standard subtitles than with verbatim subtitles. Verbatim subtitles, on the other hand, were revealed to generally be read faster. The amount of reduction is thus expected to be highly relevant regarding the comparability of different subtitle tracks.

The third component is segmentation. Segmentation takes place on two levels, namely subtitle level, i.e., segmenting over several subtitles, and line level, i.e., segmenting over several lines, also called (subtitle) line-breaks. For both types, the common rule seems to be that each segment, line or subtitle, should ideally be semantically and syntactically self-contained (Díaz Cintas & Remael, 2014, p. 172; Ivarsson & Carroll, 1998) and "should appear segmented at the highest syntactic nodes possible" (Karamitroglou, 1998, p. 6). If a sentence does not fit into a single subtitle line, this sentence should be parsed to see which is the most complete syntactical and semantical part that can fit into a single line. A line-break or, for long sentences, segmentation over multiple subtitles at arbitrary, often less coherent points is expected to disrupt reading and be more challenging for the viewer (Perego, 2008, p. 214). Segmentation plays a significant role in the readability of subtitles. For comparability, it is therefore key that the segmentation in all subtitle tracks is similar. This does not necessarily mean that the subtitle tracks should have optimal segmentation, but that they should have equal amounts of sub-optimally segmented subtitles, optimal segmented subtitles, etc.

The fourth and last component to consider is the subtitles' linguistic complexity. Evidently, the reduction and segmentation of subtitles influence the final complexity of the subtitles (Perego, 2008; Szarkowska et al., 2011), but the syntactical and lexical complexity also play a vital role. Syntactic complexity has an influence on reading time (Clifton et al., 2007), which can be of importance considering the dynamic and fleeting nature of subtitles. With regard to lexical complexity, viewers spend more time on less frequent and more complex words than on frequently occurring words, indicating more effort is required to read lexically complex subtitles (Moran, 2012). With regard to the comparability of subtitle tracks, it is, of course, essential that the lexical and syntactical complexity is relatively similar. For the most part, this complexity will originate from the audiovisual material, which implies that if the material was tested beforehand without subtitles and the complexity was found to be similar, it is more likely that the subtitles will be comparable in this regard as well. However, this may not always be so simple. For edited subtitles, for example, the original is regularly simplified and reduced. Regardless of the source, the degree of simplification may differ across various subtitle tracks disrupting the comparability between them. Another example is the comparative complexity of intralingual and interlingual subtitles. While it might be more straightforward to stick to similar terms and syntactical structures for two languages of the same family, e.g., Dutch and German, it may be more complex for, say, Dutch and Chinese. To our knowledge, there are no clear guidelines on how to produce subtitles of similar syntactical and semantical complexity. Pedersen (2017) proposes the FAR model to assess subtitle quality and Díaz Cintas and Remael (2014) devote attention to the translation process and the transfer of register, style, grammar and lexicon from the original to both interlingual or intralingual subtitles, but both are, of course, more concerned with the quality of the end-product and less with the comparability between separate subtitle tracks. Regardless of the lack of guidelines, linguistic complexity can and should be carefully considered when producing similar subtitle tracks. Practical testing of the subtitle tracks may also shed light on the matter, as illustrated in the following sections.

3 Project Background

The ten-step process we discuss in this article and proposed in a previous article (Van Hoecke et al., 2022a) was developed within the S4AE project. As this project will be used as an example in the discussion of the ten-step process, the project's background and aims will first be briefly summarized below.

The project wishes to examine the effects of subtitles on the cognitive load, i.e., the load imposed on a person to complete a task to a certain level, and comprehension of students in an L2 English lecture. It follows a mixed model design that revolves around a central within-subject component. In this design, Dutch (Flemish) students view three different recorded EMI (English as a Medium of Instruction) lectures on philosophy (named P, R, and T[10]). The lectures are provided in three conditions: (1) with intralingual (English) subtitles; (2) with interlingual (Dutch) subtitles; and (3) without subtitles. The viewing of the lectures takes place in an eye tracking laboratory, which allows us to monitor the students' eye movements, measure cognitive load and assess subtitle reading using the Reading Index for Dynamic Texts (RIDT; Kruger & Steyn, 2014). After each lecture the students are required to fill out a psychometric questionnaire on cognitive load from Leppink and van den Heuvel (2015) and a comprehension test. The use of both psychometric questionnaires and eye tracking allows us to assess cognitive load and triangulate the data of both measures, as recommended by Orero et al. (2018). To measure retention, students are asked to complete the comprehension test again one month after the experiment. The collected data are subsequently correlated with the students' biographical data and language proficiency, which is tested one month prior to the experiment.

4 The Ten Steps

In a previous article (Van Hoecke et al., 2022a) we present a ten-step process to ensure the comparability of materials used in quasi-experimental and experimental subtitling studies that involve multiple conditions, tasks and/or measurements in time. The ten steps are as follows:

1. Careful preparation of materials
2. Content and feature analyses
3. First pilot study
4. Reevaluation
5. Optimization
6. Second pilot study
7. Production of comparable subtitles
8. Subtitle analyses
9. Third pilot study with subtitles
10. Finalization of materials

The first six steps were based on two pilot studies with 75 and 50 participants, respectively, and discussed in Van Hoecke et al. (2022a). Th present article is based on two more studies with 7 and 6 participants, respectively, and discusses the preparation and comparison of the subtitles (steps 7-10), taking into account relevant research and the theoretical framework presented in Section 2. For clarity purposes, we will first briefly summarize the first six steps.

4.1 The First Six Steps

To ensure the validity and strengthen the foundations of a quasi-experimental and experimental subtitling study (especially those involving repeated measures), meticulous preparation is required. The ten-step preparatory process we present illustrates a number of obstacles and key elements that we have encountered in the S4AE project (see Section 3) and may serve as a source of inspiration for similar future research. The process is structured in such a way that it

[10] Henceforth, the lectures are named P, R and T as the topic of the lectures are Thomas Piketty, Jean-Jacques Rousseau and Alexis de Tocqueville, respectively.

gradually introduces and tests the relevant materials for the eventual main study. This is also what is done in the first six steps.

In step 1, the initial materials, which in the case of the S4AE project were the three lectures and the three comprehension tests, are prepared. For a quasi-experimental and experimental study, it is crucial to take into account where a lack of comparability between distinct videos, tools of measurement, etc. could influence the results.

After this initial preparation, step 2 dictates the prepared materials to be analyzed before any field-testing is done. Conducting experiments is time-consuming and correction of any issues in the materials that can be found and eliminated beforehand is warranted. Only after the analyses show no major flaws in the materials and the researcher or research team is convinced the materials (and their comparability) is suited for testing, the next step can be taken.

In step 3 the initial materials are tested in practice. It is important to not add too many experimental components, e.g., subtitles, audiovisual source material or post hoc tests, just yet, because, if the results are skewed, it is easier to identify the cause with a small number of components. Furthermore, the process dictates a gradual increase in components to assure comparability and validity for each separate component.

After this first test, step 4 involves analysis of the data in which the focus lies on finding issues that might originate from the experimental materials. For example, in the case of the S4AE project, if the data show that participants score significantly higher on one of the comprehension tests, it is possible that this one test or the corresponding lecture is easier than the others, and thus not comparable. It is possible that there are no issues with the materials, in which case the materials do not necessarily need to be optimized. However, if needed, there are various ways to optimize the materials without having to start anew, for example by coding not-comparable comprehension tests using the Item Response Theory (Van Hoecke et al., 2022a). This is done in step 5.

Regardless of issues found in step 4 and changes made in step 5, we recommend a second test of the materials (step 6) to ensure no chance-based or sample-related errors. If the data from this second test are promising, the next key component can be added and tested, namely the AVT.

4.2 Step 7: Production of Comparable Subtitles

The production of comparable subtitles should be as much a careful and thorough process as the production or selection of the visual materials and tools of measurement (step 1). In the S4AE project three recorded EMI lectures are used since there are three conditions (no subtitles, English subtitles and Dutch subtitles). This means that for English and Dutch three comparable subtitle tracks need to be produced. This comparability needs to be present between all three English and all three Dutch subtitle tracks separately and between the English and Dutch subtitle track of each lecture.

Based on the theoretical framework discussed in Section 2, we composed a small set of practical rules that could provide an initial anchor for creating similar subtitle tracks (in the same language and between the two languages). Considering the density of the lectures, the expected language proficiency of the intended audience and recent research on subtitle speeds and word skipping for fast subtitles (Kruger et al., 2021; Liao et al., 2021), we set the maximum subtitling speed to 15 CPS. In terms of subtitle length, we allowed a maximum of 40 CPL. With these longer subtitles, we were also able to keep the reduction in the English subtitles to a minimum, which made the English subtitles near-verbatim yielding standard subtitles. Lastly, we preferred two-line subtitles over one-line subtitles, as it has been shown that viewers spend proportionally more time on one-line subtitles than on two-line subtitles (d'Ydewalle & De Bruycker, 2007). Additionally, it reduces the difference in total number of subtitle lines between the lectures. By using predominantly two lines, the same font type and size, near-verbatim/standard subtitles and by positioning the subtitles on the bottom center for all lectures, the subtitle area and appearance were expected to be similar.

After setting up this small guideline, the English subtitles for all three lectures were produced first, since the source texts were written and recorded in English. The comparability of the lectures and lecture transcripts had already been tested and confirmed in the first two steps of the preparatory process. Near-verbatim/standard English subtitles were therefore expected to carry over this comparability. However, subtitles are still distinctly different from a static text, so extra attention was paid to segmenting and reducing the subtitles of all lectures similarly. It is recommended to analyze the subtitles in one language, in this case the English language, before continuing with the subtitles in the other language(s). For structural purposes, however, the analyses of the subtitles are discussed in Section 4.3.

After the English subtitles were produced and found to be comparable in the initial analyses, the Dutch subtitles were made. For the Dutch subtitles, we disregarded the original English transcript of the lecture and used the initial English subtitles as a template. The main aim here was to make the Dutch subtitles match the English in terms of complexity, but also retain the subtitle spotting, duration and segmentation, including sub-optimally segmented parts. Although quality is important, the main goal here was not optimal quality, but comparability of the subtitles in all aspects. In a final effort to make both languages comparable, we reevaluated the English subtitles based on the Dutch subtitles. If certain words or nuances were omitted, pronouns were used or slight segmentation shifts were made during the production of a Dutch subtitle, we corrected the corresponding English subtitle and applied the same changes, keeping in mind idiomatic structures in both languages. Reductions were required, which made the English subtitles slightly less verbatim, but strengthened the similarities between the subtitle tracks of both languages. For cases in which the Dutch segmentation was sub-optimal, the segmentation in English was also altered. This may have led to worse segmentation for the English subtitles, but, again, the goal was comparability of subtitles and, consequently, a more equal number of sub-optimally segmented subtitles.

4.3 Step 8: Subtitle Analyses

To give an initial indication of the comparability of the subtitles, all six subtitle tracks were analyzed. For each comparison within and between languages, we looked at the four components discussed in Section 2.

The first component concerned the reading speed. In this analysis, we included the separate parameters such as the number of one-line and two-line subtitles, CPL, CPS and subtitle duration. As can be seen in Table 1, the mean CPL and mean subtitle duration was relatively similar for all six subtitle tracks, implying similar visual presence of the subtitles on screen. The mean CPL was around 23 to 25 characters. The Dutch subtitles always had a marginally higher mean CPL. Lecture T featured slightly more CPL and longer durations than the other two lectures, but the overall mean CPS was similar for all lectures at approximately 12.5 CPS. More importantly, the variability in CPL and CPS as measured by the standard deviation is comparable both between lectures and between languages, indicating that the CPL and particularly the CPS remains relatively constant throughout the lecture.

Table 1: Subtitle parameters[11]

	Lecture P		Lecture R		Lecture T	
	ENG	**DU**	**ENG**	**DU**	**ENG**	**DU**
Total number of subtitle lines (1-line/2-line)	101 (6/95)	101 (6/96)	103 (3/100)	103 (3/100)	96 (5/91)	96 (5/91)
Mean CPL	23.58	24.24	23.33	23.63	25.11	25.81
Std. dev. CPL	7.10	7.81	7.89	8.23	7.84	8.27
Mean CPS	11.92	12.23	12.49	12.61	12.15	12.49
Std. dev. CPS	1.96	1.88	1.47	1.78	1.78	1.90
Mean subtitle duration	3.88	3.88	3.68	3.68	4.01	4.01
Std. dev. subtitle duration	1.137	1.137	1.004	1.004	0.974	0.974

The second component revolved around the reductions that were made during the production of the subtitles. Using Díaz Cintas and Remael's (2014, p. 151–171) classification of condensations and reformulations, the subtitle tracks were analyzed and compared. We first looked at the English subtitles only. Here we saw a total of 14 reductions for lecture P and 19 each for R and T. Of these reductions, there were 5 total reductions/omissions for P, 7 for R and 6 for T. It is important to note that most omissions were limited to single words, such as adverbs or adjectives. Some examples are 'quite simple' > 'simple' in P or 'is commonly referred to' > 'is referred to' in R. Considering that these omissions are often only single words and that their frequency is similar across the three English subtitle tracks, these were not considered an issue. Regarding the partial reductions, these were generally also limited to the use of pronouns instead of full names, or of shorter synonymous words. Only for lecture T a subtitle could be observed that changed the form of the original soundtrack entirely: 'This is not to say that there are no social-economic classes in a democracy. Of course there are.' > 'However, there are social-economic / classes in a democracy, of course.' While this was the only subtitle track to include such a relatively major change, it was expected not to influence the results, since the comprehension test did not contain a question about this specific piece of information. Additionally, the same syntactical structure was used in the Dutch subtitle, which meant that interlingual comparability was not problematic. We then analyzed the Dutch subtitles. As the English subtitles were used as a template for the production of the Dutch subtitles, all initial reductions made would also be included in the Dutch subtitles. If an additional reduction was necessary for the Dutch subtitle, we tried to further reduce the English subtitles, but this was not always possible without making significant structural changes. Consequently, there were still minor differences in reductions between the Dutch and English subtitles. For lecture P, we highlighted 23 reductions, 2 of which were omissions/additions, R had 22 reductions with 3 omissions/additions and T had 21 reductions, 5 of which were omissions/additions (multiple reductions could occur within one subtitle). The partial reductions that we highlighted concerned small shifts, which most commonly were changes in word class, e.g., 'exceeds' > *is groter* ['is larger'] in P, or 'In studying' > *Tijdens de studie van* ['During the study of'] in T, changing passive to active voice or vice versa, e.g., 'is constituted by' > *bestaat uit* ['consists of'] in R, and a change in subject, e.g., 'It initiates the tragedy' > *Zo begint het drama* ['The tragedy starts with this'] in R or 'It was up to the nations of his day' > *De naties van zijn tijd moesten* ['The nations of his day had to'] in T. After careful consideration, we expected these differences not to significantly influence the results in future experiments.

For the third component, the segmentation and line-breaks of all subtitle tracks were analyzed. We will first discuss the changes in line-breaks between the two languages. For lecture P, we

[11] The lowest *p*-value found in Mann-Whitney U tests comparing the CPL, CPS and subtitle duration of subtitle tracks between languages was $p = 0.112$ for the English and Dutch CPS in Lecture T.

observed a total of 12 shifts in line-breaks comparing the English and the Dutch subtitle track; for R 25 shifts; and for T 12 shifts. To evaluate whether these shifts, and the relatively large number of them in R, would not influence the comparability of the subtitle tracks, we categorized the line-breaks based on what changed (and potentially why it changed). The majority of these shifts seemed to be based on three principles. Firstly, in Dutch the verb is sometimes placed at the end of a sentence while it generally follows the subject in English. To build idiomatic structures, the verb was thus frequently placed in the second subtitle line instead of the first in Dutch. These shifts were never considered problematic, only when they occurred between two subtitles, i.e., the viewer only received the main verb at a later point in time for one of the two subtitle languages. A second reason for these shifts, which caused a number of shifts especially for R, was the change in location of the negation. Whereas the negation generally accompanies the verb at the front of the sentence in English, it is more idiomatic in Dutch to place the negation at the end. A third reason was a change in word order often as a result of a partial reduction. In cases where the word class was changed or the passive voice was used instead of the active voice, the sentence structure changed which often also resulted in a different line-break. These shifts in line-breaks could not always be matched in both languages, so in some cases there were differences in line-breaks that were sub-optimal in one language but not in the other (see Table 2). Because we expect the large majority of these shifts in line-breaks to have no influence, we conclude that the line-breaks are sufficiently similar and of comparable quality and thus fit for the within-subject experiment. As for segmentation between two subtitles, we observed 6 shifts in lecture P, 2 in R and 4 in T. The reasons for these segmentation shifts were generally the same as the ones we mentioned above. In terms of comparability between the subtitle tracks in one language, we also looked at the total number of sub-optimally segmented subtitles (both line-breaks and segmentation) (see Table 2). We observed 7 sub-optimal line-breaks in lecture P, 2 in R and 4 in T. Lecture T also had one case where the segmentation between subtitles was sub-optimal. In most, if not all, cases the direct cause of this sub-optimal segmentation is the syntactic nodes being too long in either of the languages to fit on one subtitle line only. This implies that more than 40 characters are needed for the entire phrase, word group, etc. Because we attempted to match the segmentation between both languages in every case, some of these sub-optimally segmented subtitles could have been prevented in one language, but, maximizing comparability, this was not done.

Table 2: Subtitle segmentation

		Shifts between English and Dutch			Both languages
		Unproblematic	Unfavorable for English	Unfavorable for Dutch	Sub-optimal
Lecture P	**Line-break**	9	2	1	7
	Segmentation	4	0	2	0
	Total	15	2	3	7
Lecture R	**Line-break**	19	4	2	2
	Segmentation	1	0	1	0
	Total	20	4	3	2
Lecture T	**Line-break**	10	0	2	4
	Segmentation	0	2	2	1
	Total	10	2	4	5

Since some reductions that were made directly resulted in shifts in segmentation as well, we also checked how many “exact” subtitle matches, i.e., no reductions or shifts in segmentation, between both languages were present for each lecture. For lecture P, 60 of 101 subtitles (59.41%) are considered exact matches in English and Dutch, for R 62 of 103 (60.19%) are and for T 57 of 96 (59.38%) are. The number of altered subtitles between the subtitle tracks of each lecture is therefore very similar.

The last component concerned the linguistic, i.e., syntactical and lexical, complexity of the subtitles. We expected the linguistic complexity of the English subtitles to mirror the complexity of the original texts as the English subtitles were mostly verbatim, i.e., matching the original soundtrack. The linguistic complexity of these original texts had been shown to be comparable in step 1 of the process (as reported on in Van Hoecke et al. (2022a) and using Perego et al. (2018) as a source of inspiration). However, a more objective assessment is recommendable, especially since the argument put forward above does not apply for the Dutch subtitles. Objectively assessing the lexical or syntactical complexity of subtitle tracks is a difficult endeavor. There are readability measures to evaluate the complexity of a static text (e.g., Flesch Reading Ease Formula), but the use of standard readability formulae and readability indices should be warranted when analyzing linguistic subtitle complexity. Readability indices have been already criticized for their inaccuracy for shorter texts (Kidwell, Lebanon, & Collins-Thompson, 2011), let alone using them for single sentences or separate clauses. One way of avoiding this challenge when analyzing syntactic complexity of subtitles would be to bring the subtitles together in a single text and apply measures like average sentence length or number of clauses. This, however, disregards the segmentation of subtitles. Applying these measures to subtitles, which may consist of one or two sentences, but also just a part of a sentence due to segmentation, will, in our view, therefore not yield any meaningful result. Basic indices to measure lexical complexity such as word length have also received considerable criticism. They have been found to be rather superficial, disregard the entire text structure and overall cohesion and coherence and are not necessarily causally related to linguistic complexity (Kraf & Pander, 2009). Nevertheless, one way to measure lexical complexity in subtitles would be to measure word frequencies. This can be done using the SubtLex corpus for subtitle word frequencies. This corpus exists for multiple languages, e.g., Dutch (Keulers, Brysbaert, & New, 2010), British English (van Heuven et al., 2014) or American English (Brysbaert & New, 2009), and thus also allows the comparison of relative word frequencies across languages. To examine the comparability between the subtitle tracks in our study, both between two languages and within one language, we tokenized and lemmatized the subtitles. Subsequently, we extracted the logarithmic word frequency scores from the Dutch SubtLex corpus and the American English SubtLex corpus (since American English spelling was used for the subtitles). We lemmatized the subtitles as it would give a more accurate indication of the word frequencies of the lemma itself. We also left out every name, year and number as these frequencies might skew the results. Two Kruskal Wallis tests revealed no significant differences between the English subtitle tracks for all three lectures, $H(2) = 1.328$, $p = 0.515$, or between the three Dutch subtitle tracks, $H(2) = 1.203$, $p = 0.548$. This suggests that the lexical complexity based on word frequency is comparable for the subtitle tracks in the same language. We then ran three Mann-Whitney U tests to compare the word frequencies between the English and Dutch subtitle tracks of each video. No significant differences were found comparing the English ($Mdn = 4.437$) and Dutch ($Mdn = 4.319$) tracks for lecture P, $U = 279764$, $z = 0.962$, $p = 0.336$, $r = 0.03$, and none were found when comparing the English ($Mdn = 4.742$) and Dutch ($Mdn = 4.403$) subtitles for lecture R, $U = 301939$, $z = 1.799$, $p = 0.072$, $r = 0.05$. A comparison of the English ($Mdn = 4.488$) and Dutch ($Mdn = 4.453$) tracks for lecture T, $U = 288212.5$, $z = 1.528$, $p = 0.126$, $r = 0.04$, did not yield any statistical difference either. These results suggest that the word frequencies of the Dutch and English tracks for each lecture are also comparable. Comparable syntactic complexity between the English and Dutch tracks is, however, less clear-cut. While an attempt was made to match the structure of the English subtitles, a few minor shifts were still present in all lectures. In turn, comparable segmentation and line-breaks between all subtitle tracks, resulting in relatively similar number of clauses and comparable structures, indicated a baseline syntactical comparability. One important difference in this case, however, is that, as mentioned before, the verb in Dutch is generally placed later on in the sentence. This may lead to a line-break being present between the subject or auxiliary verb and the main verb in Dutch, while they are next to one another in English.

4.4 Step 9: Third Pilot Study with Subtitles

To verify the conclusions drawn from step 8, a pilot study was set up consisting of two small-scale experiments. The first experiment was conducted in October 2020 with 7 students from the 3rd-year of the BA in Applied Linguistics or 1st-year of the MA Interpreting or Translation at the University of Antwerp (only 6 students are considered in the analyses below as 1 participant was excluded based on an eye tracking ratio below 85%). The second experiment took place in March 2021 and included 6 students from the MA Linguistics and Literature or the MA Interpreting at the University of Antwerp. In both experiments, the students viewed all three lectures while being monitored with an SMI RED 250Hz eye tracker, completed the biographical survey, psychometric questionnaires and the comprehension tests and were also interviewed after the experiment. However, in the first experiment (henceforth called ES) all lectures were subtitled in English and in the second (DS) they were subtitled in Dutch. As the groups in both experiments were too small to include between-group variables, such as the student's English proficiency or prior knowledge of the subject (philosophy), these were not included.

For each subtitle track, we collected three types of data. First, we measured the cognitive load using the validated psychometric questionnaire from Leppink and van den Heuvel (2015). This questionnaire consists of eight questions, in which the students had to rate the complexity of the subtitled lecture on a scale from 1 (low complexity) to 10 (high complexity). The first four questions concerned content complexity, providing insight into the perceived intrinsic load. The last four concerned instructional complexity, i.e., perceived extraneous load, and are expected to reveal the effects of subtitles. Second, we had comprehension scores, which could reveal if any of the subtitle tracks influenced comprehension more than the other tracks. Third, eye tracking data was collected to provide insight into differences in subtitle reading behavior, which may also influence the cognitive load ratings and comprehension scores. In these eye tracking data we limit ourselves to fixation counts, average fixation durations and dwell times in the subtitles' areas of interest (AOI). These global measures are indicators of processing (Schotter & Rayner, 2012) and have been shown to be measures of cognitive load (Kruger & Doherty, 2016).

We analyzed the cognitive load ratings, comprehension scores and eye tracking data in each experiment separately to assess the comparability of the subtitle tracks in each of the languages. In view of the small sample size, we consistently used Friedman's tests to assess within-subject differences. For future research, we recommend linear mixed models in which all subtitles can be treated as separate items in the design. This way, intrinsic variability is accounted for and differences can be more accurately measured. While this is something we intend to use in the main study of the project, we limited the comparability analyses to the Friedman's test as our sample size for this analysis is very limited and we still consider this sufficiently robust with the present goal in mind, namely examining comparability.

As shown in Table 3, no significant within-subject effects were found for the cognitive load ratings, the comprehension scores or the eye tracking variables. These findings indicate that, according to these data, there is no significant difference between the three subtitle tracks in English or in Dutch.

Table 3: Within-subject differences

		ES			DS		
		df	*Q*	*p*	*df*	*Q*	*p*
Cognitive Load	**Total**	2	0.609	0.738	2	4.000	0.135
	Intrinsic	2	0.300	0.861	2	5.304	0.070
	Extraneous	2	0.273	0.873	2	0.873	0.293
Comprehension	**Scores**	2	4.000	0.135	2	1.652	0.438
Eye Tracking	**Fixation Count in AOI**	2	0.333	0.311	2	2.333	0.846
	Mean Fixation Duration in AOI	2	5.333	0.069	2	5.333	0.069
	Dwell Time in AOI	2	1.000	0.607	2	1.333	0.513

The comparison of the English and the Dutch subtitle tracks would seem like a logical next step. However, such a comparison - based on the differences in results of the cognitive load questionnaire, comprehension questionnaire and eye tracking measures – should be approached with caution. Such a comparison would rather study the effects instead of the comparability of the subtitle tracks in different languages. Moreover, when a significant difference is found using the same data collection methods as before, it is practically impossible to determine whether this difference was caused by the subtitle complexity or by other confounding factors, such as the matching or contrast between the soundtrack language and the subtitle language or participants' proficiencies, prior knowledge, and subtitle language preference. In terms of cognitive load, a difference in intrinsic load between the two languages would not be expected, as the change in language of the subtitle track is supposed to not have an influence on content complexity, but rather on instructional complexity (i.e., extraneous cognitive load). With regard to comprehension, any significant difference found between languages may very well be due to participants being more proficient or native speakers in one language, thus understanding more because of the subtitle language. This effect of subtitle language on comprehension has been shown in previous research (Lavaur & Bairstow, 2011) and is also one of the research foci in the S4AE project. As for the eye tracking data, any difference found would not necessarily mean that there is a difference between the language tracks either. As previous studies have revealed (Kruger et al., 2014; Hefer, 2013a), the reading behavior of a viewer is not the same for intralingual and interlingual subtitling or for native and foreign language subtitling. For these reasons, statistical comparison of tracks in both languages is not warranted. Significant differences found between the two might not necessarily imply dissimilarities between subtitles, but may simply be caused by different reading patterns in different types (intralingual vs. interlingual subtitles) or different language subtitles.

4.5 Step 10: Finalization of Materials

Step 9 revealed that there were no differences between the subtitle tracks and that the subtitles were perceived to be of adequate quality, i.e., representative for actual subtitles. This means that for this concrete example of the S4AE project, adjustments to the subtitles or the other materials was not considered necessary. However, not all future studies may have this outcome. Therefore, a final optimization of the subtitles (the other materials should already be optimized) could be carried out in step 10 before a potential main study. Ideally, the altered subtitles should then be tested again in practice, but depending on the size of the changes made, this may not be deemed necessary.

5 Conclusion

While an increased number of cognitive AVT studies have been conducted in the past decades (Díaz Cintas, 2020), little attention is devoted to preparing and comparing materials for such studies. Meticulous preparation and practical testing of the research materials is important for any experimental study, but absolutely critical for AVT studies using repeated measures, e.g., multiple clips, multiple language tracks, several measurements in time. This paper builds on our previous proposal (Van Hoecke et al., 2022), in which we lay out a ten-step common-sense preparatory process for quasi-experimental and experimental AVT studies. This paper in particular demonstrates possible key points and obstacles in producing comparable subtitles for such a study. Fundamental is the concept of ‘comparability over quality’. In most cases, a subtitler would be concerned with subtitle quality and thus follow the proposed guidelines of optimal segmentation, reduction, terminology, editing, etc. In the case of a study using repeated measures, quality remains important, but it may be more interesting to reduce more than necessary in certain subtitles or have the line-break at a different, less optimal place if it would make the subtitles of different languages, but also of different clips in the same language, more similar overall. Ensuring the comparability between subtitles in the same language, and especially in different languages, is a complex process. In some cases, e.g., between two languages, it is near impossible to be completely certain there are no significant differences between the two. An effort should be made to examine the comparability in the preparatory phase, but one should keep in mind that these differences can still be accounted for using linear mixed models in the main study.

We hope that these ten steps, inspired by our personal experiences in the S4AE project, may be of use and inspiration for similar future AVT research. Evidently, the ten-step preparatory process may need to be slightly altered to fit specific research goals. Moreover, some of the steps can be refined, e.g., by running linear mixed models to account for the individual subtitles, even when we did not do so in this paper. We also acknowledge that the process of producing comparable subtitles for English and Dutch is most likely more straightforward than doing so for, say, English and Chinese. Nevertheless, this ten-step common-sense preparatory process has shown that, regardless of the complexity of preparing comparable subtitles, it is not impossible.

6 References

Brysbaert, Marc, and Boris New. 2009. "Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English." *Behavior Rsearch Methods 41*(4): 977–990. DOI 10.3758/BRM.41.4.977.

Clifton, Charles, Adrian Staub, and Keith Rayner. 2007. "Eye movements in reading words and sentences." In *Eye movements: A window on mind and brain*, edited by Roger P. G. Van Gompel, Martin H. Fischer, Wayne S. Murray, and Robin L. Hill, 341–371. Amsterdam: Elsevier. DOI 10.1016/B978-008044980-7/50017-3.

Díaz Cintas, Jorge. 2003. *Teoría y práctica de la subtitulación: Inglés-Español*. Barcelona: Ariel.

Díaz Cintas, Jorge. 2020. "Audiovisual translation." In *The Bloomsbury Companion to Language Industry Studies*, edited by Erik Angelone, Maureen Ehrensberger-Dow, and Gary Massey, 209–230. London: Bloomsbury. DOI 10.5040/9781350024960.0014.

Díaz Cintas, Jorge, and Aline Remael. 2014. *Audiovisual Translation: Subtitling*. Manchester: St Jerome. DOI 10.4324/9781315759678.

Doherty, Stephen, and Jan-Louis Kruger. 2018. "The development of eye tracking in empirical research on subtitling and captioning." In *Seeing into screens: Eye tracking and the moving image*, edited by Tessa Dwyer, Claire Perkins, Sean Redmond, and Jodi Sita, 46–56. London: Bloomsbury. DOI 10.5040/9781501329012.0009.

d'Ydewalle, Géry, and Wim De Bruycker. 2007. "Eye movements of children and adults while reading television subtitles." *European Psychologist* 12(3): 196–205. DOI 10.1027/1016-9040.12.3.196.

d'Ydewalle, Géry, Johan Van Rensberger, and Joris Pollet. 1987. "Reading a message when the same message is available auditorily in another language: The case of subtitling." In *Eye Movements from Physiology to Cognition, edited by* J. Kevin O'Regan, and Araine Lévy-Schoen, 313–321. Amsterdam: Elsevier. DOI 10.1016/B978-0-444-70113-8.50047-3.

Fresno, Nazaret, and Katarzyna Sepielak. 2020. "Subtitling speed in Media Accessibility research: Some methodological considerations." *Perspectives*. DOI 10.1080/0907676X.2020.1761841.

Gottlieb, Henrik. 2012. "Subtitles: Readable dialogue?" In *Eye tracking in audiovisual translation,* edited by Elisa Perego, 37–82. Rome: Aracne.

Hefer, Esté. 2013. "Reading first and second language subtitles: Sesotho viewers reading in Sesotho and English." *Southern African Linguistics and Applied Language Studies* 31(3): 359–373. DOI 10.2989/16073614.2013.837610.

Ivarsson, Jan, and Mary Carroll. 1998. *Subtitling*. Simrishamn: TransEdit.

Karamitroglou, Fotios. 1998. "A proposed set of subtitling standards in Europe*." Translation Journal* 2(2).

Keuleers, Emmanuel, Marc Brysbaert, and Boris New. 2010. "Subtlex-NL: A new measure for Dutch word frequency based on film subtitles." *Behavior Research Methods 42*(3): 643–650. DOI 10.3758/BRM.42.3.643.

Kidwell, Paul, Guy Lebanon, and Kevyn Collins-Thompson. 2011. "Statistical estimation of word acquisition with application to readability prediction." *Journal of the American Statistical Association 106*(493): 21–30. DOI 10.1198/jasa.2010.ap09318.

Kraf, Roger, and Henk Pander. 2009. "Leesbaarheidsonderzoek: Oude problemen, nieuwe kansen [Readability research: Old problems, new opportunities]." *Tijdschrift voor Taalbeheersing 31*(2): 97–123. DOI 10.5117/TVT2009.2.LEES356.

Kruger, Jan-Louis, and Stephen Doherty. 2016. "Measuring cognitive load in the presence of educational video: Towards a multimodal methodology." *Australasian Journal of Educational Technology* 32(6): 19–31. DOI 10.14742/ajet.3084.

Kruger, Jan-Louis, Esté Hefer, and Gordon Matthew. 2014. "Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2." *Journal of Eye Movement Research* 7(5): 1–15. DOI 10.16910/jemr.7.5.4.

Kruger, Jan-Louis, María T. Soto-Sanfiel, Stephen Doherty, and Ronny Ibrahim. 2016. "Towards a cognitive audiovisual translatology." In *Reembedding Translation Process Research*, edited by Ricardo Muñoz Martín, 171–193. Amsterdam: John Benjamins. DOI 10.1075/btl.128.09kru.

Kruger, Jan-Louis, and Faans Steyn. 2014. "Subtitles and eye tracking: Reading and performance." *Reading Research Quarterly* 49(1): 105–120. DOI 10.1002/rrq.59.

Kruger, Jan-Louis, Agnieszka Szarkowska, and Izabela Krejtz. 2015. "Subtitles on the moving image: An overview of eye tracking studies." *Refractory: A Journal of Entertainment Media* 25: 1–14.

Kruger, Jan-Louis, Natalia Wisniewksa, and Sixin Liao. 2022. "Why subtitle speed matters: Evidence from word skipping and rereading." *Applied Psycholinguistics 43*(1): 211–236. DOI 10.1017/S0142716421000503.

Lavaur, Jean-Marc, and Dominique Bairstow. 2011. "Languages on the screen: Is film comprehension related to the viewers' fluency level and to the language in the subtitles?" *International Journal of Psychology* 46(6): 455–462. DOI 10.1080/00207594.2011.565343.

Leppink, Jimmie, and Angelique van den Heuvel. 2015. "The evolution of cognitive load theory and its application to medical education." *Perspectives on Medical Education* 4(3): 119–127. DOI 10.1007/s40037-015-0192-x.

Liao, Sixin, Lili Yu, Erik D. Reichle, and Jan-Louis Kruger. 2021. "Using eye movements to study the reading of subtitles in video." *Scientific Studies of Reading 25*(5): 417–435. DOI 10.1080/10888438.2020.1823986.

Moran, Siobhan. 2012. "The effect of linguistic variation on subtitle reception." In *Eye tracking in audiovisual translation*, edited by Elisa Perego, 37–82. Rome: Aracne.

Orero, Pilar, Stephen Doherty, Jan-Louis Kruger, Anna Matamala, Jan Pedersen, Elisa Perego, Pablo Romero-Fresco, Sara Rovira-Esteva, Olga Soler-Vilageliu, and Agnieszka Szarkowska. 2018. "Conducting experimental research in audiovisual translation (AVT): A position paper." *The Journal of Specialised Translation* 30: 105–126.

Pedersen, Jan. 2017. The FAR model: Assessing quality in interlingual subtitling. *The Journal of Specialised Translation* 28: 210–229.

Perego, Elisa. 2008. "Subtitles and line-breaks: Towards improved readability." In *Between Text and Image: Updating research in screen translation*, edited by Delia Chiaro, Christine Heiss, and Chiara Bucaria, 211–223. Amsterdam: John Benjamins. DOI 10.1075/btl.78.21per.

Perego, Elisa, Fabio Del Missier, and Marta Stragà. 2018. Dubbing vs. Subtitling: Complexity matters. *Target 30*(1): 137–157. DOI 10.1075/target.16083.per.

Romero-Fresco, Pablo. 2009. "More haste less speed: Edited versus verbatim respoken subtitles." *Vial-vigo International Journal of Applied Linguistics* 6: 109–133.

Schotter, Elizabeth R., and Keith Rayner. 2012. "Eye movements in reading: Implications for reading subtitles." In *Eye tracking in audiovisual translation*, edited by Elisa Perego, 83–104. Rome: Aracne.

Szarkowska, Agnieszka, and Lidia Bogucka. 2019. "Six-second rule revisited: An eye tracking study on the impact of speech rate and language proficiency on subtitle reading." *Translation, Cognition & Behavior* 2(1): 101–124. DOI 10.1075/tcb.00022.sza.

Szarkowska, Agnieszka, and Gerber-Morón, Oliva. 2018. "Viewers can keep up with fast subtitles: Evidence from eye movements." *PLoS One* 13(6). DOI 10.1371/journal.pone.0199331.

Szarkowska, Agnieszka, Izabela Krejtz, Zuzanna Klyszejko, and Anna Wieczorek. 2011. "Verbatim, standard, or edited? Reading patterns of different captioning styles among deaf, hard of hearing and hearing viewers." *American Annals of the Deaf* 156(4): 363–378. DOI 10.1353/aad.2011.0039.

van Heuven, Walter J. B., Pawel Mandera, Emmanuel Keuleers, and Marc Brysbaert. 2014. "Subtlex-UK: A new and improved word frequency database for British English." *Quarterly Journal of Experimental Psychology 67*(6): 1176–1190. DOI 10.1080/17470218.2013.850521.

Van Hoecke, Senne M., Iris Schrijver, and Isabelle S. Robert. 2022. "Methodological preparation of a within-subject audiovisual cognition, reception and perception study." *Journal of Audiovisual Translation, 5*(1): 94–128. DOI 10.47476/jat.v5i1.2022.163.

Part III: Main Experiments

Chapter 5 – Reception and perception of subtitle presence and language in different EMI lectures

After the foundation provided by the preparatory work in relation to the materials, the experiments to answer the research questions for this thesis could be conducted. The current chapter includes two papers that discuss two components of a large-scale study conducted at the University of Antwerp. Specifically, this study was aimed at examining the effect of the presence and language of subtitles on reception, more specifically the visual attention distribution, and, to a limited extent, reading, comprehension and cognitive load, and perception. For this study, students from the second year of the graduate program Applied Linguistics at the University of Antwerp were recruited. The study consisted of three individual sessions. First, the students were tested for language proficiency, prior knowledge and socio-economic factors. Second, students were shown three EMI lectures in either a talking head or a PowerPoint lecture style. Each time, one lecture had no subtitles, one had English subtitles and one had Dutch subtitles. During the viewing, their eye movements were tracked. The students then had to complete a series of tests for comprehension, perceived cognitive load and verbal working memory span. Lastly, an interview was conducted to assess their perception of subtitles and subtitled lectures. In the third and last stage, students had to complete a delayed comprehension test to assess retention. Together, these two papers discuss the results of the entire study and formulate answers to a number of research questions mentioned in Section 1 of Chapter 3.

Paper 3 – The effects of subtitle presence and language on processing, cognitive load and comprehension in online EMI lectures: An eye-tracking study

This paper was submitted to the Australasian Journal of Educational Technology (2023):

Van Hoecke, S. M., Schrijver, I., Robert, I. S., & Kruger, J.-L. (under review). The effects of subtitle presence and language on processing, cognitive load and comprehension in online EMI lectures: An eye-tracking study. *The Australasian Journal of Educational Technology.*

Formatting has been slightly altered to fit the present thesis.

Abstract

Online lectures have become a common sight in present higher education. Consequently, an opportunity arises to employ subtitles more easily to further increase accessibility and reach larger audiences. Research on how subtitles affect processing, cognitive load and comprehension in asynchronous, online lectures, however, remains limited. This study examines the effects of the presence and language of subtitles in two distinct types of recorded lectures, a talking head lecture and a voiceover PowerPoint presentation. Participants were shown three L2 English lectures in either lecture style in three subtitle conditions: with L1 Dutch subtitles, with L2 English subtitles or without subtitles. Results showed that the inclusion of a PowerPoint presentation is key to improving comprehension and decreasing perceived cognitive load. Subtitles receive a significant amount of attention when present, though learners can effectively process this additional source of information without it having any detrimental effect on comprehension. On the contrary, same-language subtitles are even shown to be beneficial for long-term comprehension. Native-language subtitles, on the other hand, reduce perceived cognitive load as long as no other visual-verbal information in another language is shown on screen.

Key words

Audiovisual translation (AVT), subtitles, cognitive load, comprehension, lecture style, eye tracking

1 Introduction

In the recent pandemic, it became abundantly clear how pivotal the role of multimedia is in higher education. Hybrid forms of lecturing in which teaching happens in person but is also recorded and/or streamed to provide remote access to students, have become increasingly popular and necessary. With the resultant increase in the number of online videos used in education, an opportunity arises for the use of subtitling in the field of education to serve a number of purposes. In today's globalized world, subtitled education provides the opportunity to improve access and significantly increase the number of students who can benefit from education. However, the question of whether and how the effectiveness of subtitles could be improved by using different formats of online educational material, has not been answered comprehensively.

The bulk of research on audiovisual translation (AVT) in education has focused on language learning and accessibility. In recent years, however, more studies have emerged studying the effects of different subtitle languages and subtitle modes (e.g., bilingual subtitles vs. monolingual subtitles) on the hearing audience in different (semi-)educational video formats (e.g., Chan et al., 2019; Hosogoshi, 2016; Liao et al., 2020; Van Hoecke et al., 2022a, 2022b). Despite the growing interest in the topic, the field is yet to reach consensus on the effects of subtitles. Some pertinent questions remain, namely whether subtitles increase or decrease cognitive load, which is the load that is imposed on someone when processing information or completing a task (Sweller et al., 2019), and, relatedly, whether subtitles result in an increase in comprehension and learning.

The dynamic and highly interactive nature of (subtitled) multimedia instruction makes it a highly complex matter. A number of instructional theories, e.g., the cognitive load theory (Sweller et al., 2011; Sweller et al., 2019) and the cognitive theory of multimedia learning (Mayer, 2014a), list a number of principles to consider when optimizing cognitive load for instruction. The majority of these principles are strongly supported by research. However, the findings are at times not consistent. These inconclusive results in the large number of conducted studies on the topic make one thing clear: the effects of (subtitled) instruction are highly dependent on a number of moderating factors. From the literature, we can distinguish moderating factors related to three distinct components of instruction: (1) the audience; (2) the subtitles; and (3) the design of the instructional material.

It is unavoidable that the audience plays a substantial role regarding the load generated by instruction. Two key factors here are the prior knowledge the audience has on a topic (Kalyuga, 2012) and the language proficiency of the viewer in case the instructional material is not in their native language (e.g., Lavaur & Bairstow, 2011). Instructional design has no control over these.

However, when talking about subtitled educational videos, it is important to know that the characteristics of the subtitles themselves have a significant impact on the learning experience as well. In previous research it was already revealed that characteristics, such as the language of the subtitles (e.g., Hefer, 2013a, 2013b), the type or mode of the subtitles (e.g., bilingual vs. monolingual subtitles in Liao et al., 2020; automatically generated vs. corrected in Chan et al., 2019), the presentation of the subtitles (e.g., block vs. scrolling in Rajendran et al., 2013) and the speed of the subtitles (e.g., Liao et al., 2021) change the participants' reading behavior and, consequently, cognitive processing.

More broadly, the design of instructional material has received a considerable amount of attention in previous research. The principles that we referred to earlier are mainly concerned with how instructional material should be designed to optimize cognitive load and learning. Content complexity plays an important role here, but is beyond the scope of the present paper.

Instead, the paper wishes to focus on the visual-textual complexity of the lecture, a component of instruction that has so far received relatively little attention.

In sum, this paper focuses on interactions between and effects of subtitles and visual-textual complexity of the lecture on students. It wishes to answer the question of how the presence and language of subtitles impact processing, cognitive load and comprehension in different styles of online lectures. The paper reports on a study that makes use of eye tracking, self-report psychometric questionnaires and comprehension tests. The paper first sketches a theoretical framework drawing on the cognitive load theory (Sweller et al., 2011; Sweller et al., 2019) and the cognitive theory of multimedia learning (Mayer, 2014a). The theoretical framework is by no means a comprehensive overview of all instructional principles, but instead focuses on a number of instructional principles that played a key role in the present study.

2 Theoretical framework

When creating material for online learners, it is essential to consider the cognitive load imposed on the viewer by the material. Meaningful learning can only be achieved when the instructional load is controlled and within the bounds of the learner's working memory capacity (Paas et al., 2003). However, it rarely happens that instructional design takes the human cognitive architecture into account (Sweller et al., 2011). Yet, when taken into consideration, it has been shown to generate effective instructional methods that at times appear counterintuitive (Sweller et al., 2011). One method to consider is subtitles.

Subtitles in the native language of the audience or in English as a foreign language could serve as an additional, possibly helpful, source of visual-textual information. However, if we consider the previously mentioned human cognitive architecture, subtitles might have an adverse effect on the audience. The cognitive theory of multimedia learning (Mayer, 2014a), built on the cognitive load theory (Sweller, 1988; Sweller et al., 2011; Sweller et al., 1998; Sweller et al., 2019), distinguishes three types of processing: (1) essential processing, i.e., the processing dealing with intrinsic load generated by the inherent complexity of the learning content as well as the prior knowledge of the individual; (2) extraneous processing, which deals with extraneous load generated by the design of the material or details that do not support learning directly; and (3) generative processing, which deals with the germane load and essentially takes care of the distribution of working memory resources from extraneous to intrinsic information. To learn effectively, humans have to process the information actively, select relevant information, organize it cognitively and integrate it with prior knowledge. In other words, the goal in instructional design is to reduce the extraneous load through careful design of material, which will leave more working memory capacity for essential processing. However, humans only have a limited capacity to do this and if the necessary processing effort exceeds the capacity, learning is hampered.

Literature lists a number of principles to consider when minimizing cognitive load generated by the material. These principles form a web of interactions and effects. While they can provide valuable insight into how instructional material should be designed and what should be considered, this paper focuses only on principles relevant for the research at hand, namely the redundancy principle, the modality principle, the transient information principle and the signaling principle. For an overview of all the other principles, we refer to the respective theoretical works (Mayer, 2014a; Sweller et al., 2019).

Offering subtitles during a lecture means there is one more source of information for the student to process. This begs the question whether subtitles increase the load generated by the material. Three cognitive principles play an important role here. If we consider subtitles as a mere copy of already available information, we can consider it redundant. According to the redundancy principle, any material that does not directly add to the essential learning content, such as related but unnecessary imagery or overly elaborate definitions and thus also subtitles, increases extraneous processing load (Kalyuga & Sweller, 2014). It is, however, not that straightforward. Following the modality principle, cognitive load is decreased when information is presented in both channels simultaneously (Low & Sweller, 2014; Sweller et al., 1998, 2019). Subtitles provide information in the visual channel, whereas the soundtrack provides information to the auditory channel. This would suggest subtitles could instead alleviate the load on the learner. A final principle that might provide an answer to the question is the transient information principle. Transient information is assumed not to be as effective in instruction as non-transient information (Leahy & Sweller, 2011, 2015), which implies subtitles would be outperformed by a static full text. This, however, seems to depend heavily on the size of the transient information. In fact, shorter chunks of information that are presented both visually and auditorily might even decrease cognitive load (Leahy & Sweller, 2015), which is another argument in favor of using subtitles.

The potential conflict in these principles is also represented in the literature. If we consider cognitive research into the matter, we see that learners who watch an animated learning video with narration outperform those who watch the same learning video with both narration and on-screen text that either summarizes or duplicates the concurrent narration, supporting the redundancy effect (Craig et al., 2002; Diao et al., 2007; Mayer et al., 2001). In contrast, most research in the field of Audiovisual Translation (AVT) seems to find otherwise. Based on a number of AVT studies conducted, it seems that subtitles, both same-language (intralingual) and different language (interlingual), do not increase cognitive load (Kruger et al., 2013, 2014). Furthermore, while the study of Kruger and Steyn (2014) also found that the presence of subtitles did not have an effect on cognitive load, they did find that when the subtitles were actually read, the test performance of learners improved.

These findings raise questions about the interactions between these principles and reveal the complexity of the matter. As mentioned before, other instructional principles potentially play a role as well. However, there are likely a large number of factors to take into account that go beyond the basics of these principles, which is something Kalyuga (2012) also emphasizes. These are related to three components of subtitled instruction: (1) the audience; (2) the subtitles; and (3) the instructional material. While there has been a considerable amount of research within the field of AVT focusing on the audience and on the characteristics of subtitles and how, for example, language proficiency of an audience, editing of subtitles, subtitle speed, etc. moderate cognitive processing, the interaction between subtitles and instructional material remains relatively underexplored.

Some studies have explored the effects of the language of the subtitles and the language of the narration in education. In two studies by Hefer (2013a, 2013b) conducted in South-Africa on the effects of L1 and L2 subtitles on reading, it was found that it took L2 English speakers longer to read the English subtitles than L1 English speakers. However, the former group still read the L2 English subtitles faster than they read their respective L1 subtitles. It is assumed that this was due to the history and dominance of English in South-Africa. In another South-African study by Kruger et al. (2014), it was found that L2 English subtitles were being read more than L1 Sesotho subtitles. At times, the L1 subtitles were even actively avoided. They speculate that, in the South-African academic context, there might be a preference for English due to the predominant use of English as medium of instruction. It might therefore be interesting to ask viewers about their subtitle preferences before the experiment to see how this might affect reading behavior and, consequently, the effects of subtitles. A more recent study by Liao et al. (2020) on the effect of the language of narration on intra/interlingual subtitle reading showed more subtitle skipping when the narration was in L1 and subtitles in L2 than when they were both in the L2 language of the viewers. They hypothesize that subtitle reading decreases when the need to read the subtitles to be able to understand the source decreases.

While these studies have significant implications regarding the implementation of subtitles in education, it would be interesting to also consider how subtitle processing and cognitive load is affected by the degree of visual-textual complexity of the instructional multimedia material. This has, however, received relatively little attention in research so far. Again, it is nearly impossible to consider this without considering certain instructional principles. The present paper compares voiceover lectures showing PowerPoint slides with lectures that only show the talking head of the lecturer. One other key principle for this paper is thus the signaling principle, which states that cognitive processing is improved when the learner is directed to relevant information. A well-designed PowerPoint presentation is supposed to do exactly that, i.e., serve as a guide containing key words for the viewer. The question, however, remains whether the signaling retains its effectiveness when subtitles are present on screen as well.

One study that explored the interactions between subtitles and visual-textual complexity of a lecture was conducted by van der Zee et al. (2017). It compared subtitled Massive Open Online Courses (MOOCs) of different visual-textual complexity and tested their effects on test performance. While they found no effect of subtitles on test performance or self-reported mental effort ratings, the language proficiency level of the viewer and the visual-textual complexity of the video were significant predictors of test performance. This is in line with the previous AVT studies on cognitive load effects of subtitles, but, as eye tracking was lacking in this study, nothing can be said about how eye movement or actual reading of the subtitles changes the effects or how visual-textual complexity of the material changes reading behavior.

In conclusion, an increasing amount of attention has been devoted to the effects of concurrent on-screen text, e.g., subtitles, on cognitive load and comprehension in an educational context. There does not, however, seem to be consensus on whether it increases cognitive load and improves learning. Research has revealed just how complex the matter of subtitles in educational multimedia is. Numerous instructional principles interact and moderate the effects of the subtitled educational material. On top of that, there are numerous moderating factors at play. This paper wishes to consider how certain instructional principles and moderating factors affect eye movements, self-reported cognitive load and comprehension or test performance. The moderating factors explored relate to the subtitles and the design of the instructional material, more specifically (1) the presence of subtitles; (2) the language of subtitles (intralingual, L2 subtitles vs. interlingual, L1 subtitles); and (3) visual-textual complexity of the source material.

3 Methodology

3.1 Design

The experiment used a 2 x 3 mixed-methods design (2 styles of lectures – between-group: talking head, PowerPoint presentation; x 3 subtitle conditions – within-subject: no subtitles, intralingual - L2 English subtitles, interlingual - L1 Dutch subtitles). The three within-subject conditions were counterbalanced via a Latin-square design.

Ninety-one participants were recruited from the second year of the undergraduate Applied Linguistics program at the University of Antwerp. Their ages ranged between 20 and 27 years old (79% of the participants were younger than 22 years old). Eighty-two percent of the participants were female. Ethics approval for the study was granted by the Ethics Committee for the Social Sciences and Humanities of the University of Antwerp.

The data were collected in three phases. In a first group session, participants were given a biographical survey, a listening comprehension test (the listening part of the 50-minute Education First Standardized English Test [EFSET]) and the Academic, 5000 and 10000 word level from the Vocabulary Levels Test of Schmitt et al. (2001) (a measure of vocabulary size) to assess their linguistic proficiency. For the second session, participants were invited individually to the eye-tracking lab. They were then instructed to watch three lectures, each time completing a psychometric questionnaire (adopted from Leppink and van den Heuvel, 2015) and comprehension test after the lecture. Half of the participants watched all three lectures with a talking head, the other half watched the lecture as a narrated PowerPoint presentation. In a third group session, participants were asked to complete a delayed comprehension test (the same test from the second session). As the second session was done individually for each participant, the delay varied ranging between 7 to 70 days. This is taken into account during the analyses. A total of 83 students participated in all three sessions.

3.2 Stimuli

Three recorded lectures on philosophy were used as stimuli. These three lectures were presented in two lecture styles: a talking head or a PowerPoint presentation. The recordings and subtitles were carefully prepared in a total of four pilot experiments, testing comparability and ensuring the quality of the lectures. This was done following our own recommended ten-step approach of preparing comparable materials for experimental AVT research (Van Hoecke et al., 2022a, 2022b). As shown in Table 1, the recorded lectures and subtitles were comparable in terms of duration, word count and speed (in characters per second or CPS). The subtitles in each lecture were presented at the bottom center of the screen below the video. The PowerPoints were created without taking the subtitles into consideration to maximize the ecological validity.

Table 1: Stimuli characteristics

	Lecture 1	Lecture 2	Lecture 3
Duration	7:21	7:08	7:25
Total words	833	855	833
Mean CPS Eng-Sub	11.92	12.49	12.15
Mean CPS NL-Sub	12.23	12.61	12.49
Total words on slide	189	170	178
Number of slides	7	7	7

3.3 Apparatus

Eye movements were monitored using an SMI 250Hz Remote Eye Tracking Device. The lectures were presented full screen on a Full HD 1920x1080 monitor. Before each lecture viewing, a 5-point calibration and validation were performed with a maximum error of 0.5°. After calibration, participants were asked not to move their heads during the viewing of the lecture to prevent loss of eye tracking accuracy.

3.4 Analyses

All data were analyzed in R (v4.2.0) using Generalized/Linear Mixed Models (lme4 package - v1.1-29). A successive difference/sliding contrast was set up to compare each consecutive level of the factor. The emmeans package (v1.7.4-1) was used to compare all factors after a final model was fit. Throughout the remainder of this paper, only the best model fits are discussed. These rarely included language proficiency scores, indicating language proficiency played a lesser role in the present research.

For the eye movement analyses, a total of 24 participants whose eye tracking ratio did not meet the lower boundary of 80% and/or whose calibration on the eye tracker still showed significant drift, were excluded. This left a total of 59 participants, 30 of whom watched the lectures with PowerPoint slides, 29 with a talking head. Fixations that were shorter than 40ms were not recorded and any fixations that were longer than 627.29ms (the 95% quantile) were discarded.

We created areas of interest for every subtitle, for each line of slide content and for the face of the instructor in the talking head lecture. The initial analyses for the eye movement data were done considering only the data in the area of the subtitles. Before analyzing these indices, we also looked at subtitle skipping in each format and for each subtitle language. A subtitle was skipped when there were no fixations in the subtitle's area of interest (binomial factor of processing). This was also done for slide text, thus slide skipping, later on. In these subtitle analyses, the fixed factors of the best model fit were consistently lecture style and subtitle language. The analysis of the Reading Index for Dynamic Texts (RIDT) scores, which is a composite measure of the processing of the subtitles (for more information, see Kruger and Steyn, 2014), was an exception to this, as the best model fit included listening comprehension as a fixed factor as well. The area of interest of each individual subtitle and individual participants were considered random effects.

To consider slide reading and effects of lecture style and subtitle presence/language, we also compared total fixation duration and fixation counts on the subtitles with the durations and counts on the contents of the slides or the talking head. These were chosen as a rough measure of visual attention distribution and reading. Following a principal component analysis, participants were found to explain a marginal part of the variance. We decided to run multiple linear regressions for these final analyses. In these analyses, the lecture style, area of interest, i.e., subtitles, talking head or slide content, and subtitle language were considered as factors.

In sum, six different eye movement analyses are presented in this paper based on four eye-tracking indices. The first three only considered data from the subtitles' areas of interest, namely (1) subtitle skipping based on fixation counts; (2) rough objective measurement of invested cognitive effort based on mean fixation duration; (3) subtitle reading based on average forward saccade length and unique fixation count as part of the RIDT (Kruger & Steyn, 2014). The fourth analysis considered only data from the slides' areas of interest: (4) slide skipping based on fixation counts. The last two analyses measured visual attention distribution and reading based on data from all areas of interest: (5) total fixation duration; and (6) total fixation counts.

For the subjective cognitive load analyses, lecture style and subtitle condition were used as fixed factors, and participants and the question number of the psychometric questionnaire were treated as random effects. Initially, the lecture was treated as a random effect in these models but following a principal component analysis of the random effects structure, it was revealed to explain very little variance. The same was found for the comprehension models. This was to be expected as the extensive pre-testing ensured comparable lectures and thus simplified the models to some extent. The best fitting model included a by-participant slope based on the subtitle condition.

For the comprehension analyses, the individual answers to each separate question were treated as the dependent variable and thus treated as a binomial factor. Vocabulary size, lecture style, subtitle condition and days between the first test and delayed test were treated as fixed factors, participants and comprehension questions were treated as random effects. As a substructure for the questions, the order of the lectures was also included in the random effects structure.

4 Results

4.1 Eye movement on the subtitle areas of interest

Before comparing specific eye tracking measures, we looked at subtitle skipping. Table 2 shows the percentages of subtitles skipped for each condition. In a binomial GLMM, lecture format was found to have had an effect on subtitle skipping with significantly more subtitles being skipped in the PowerPoint format ($z = 4.905$, $p < 0.01$). No significant effect of subtitle language was found ($z = 0.042$, $p > 0.05$) and no significant interaction between subtitle language and lecture style was found ($z = 0.610$, $p > 0.05$). Predicted effects are presented in Figure 1.

Table 2: Mean subtitle skipping

Subtitles	PowerPoint	Talking head	Average
English	9.17%	4.34%	6.68%
Dutch	12.10%	4.59%	8.19%
Average	10.63%	4.46%	7.43%

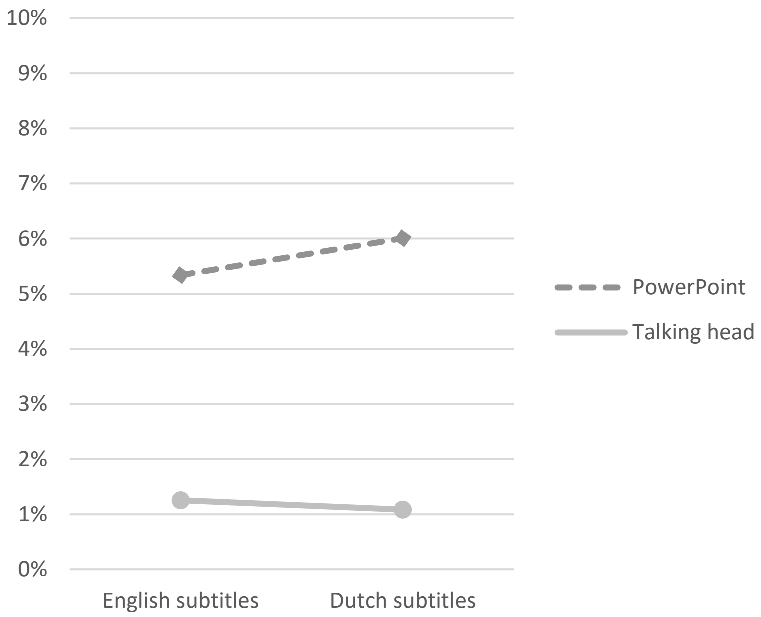


Figure 1: Predicted subtitle skipping

In a next step, mean fixation duration is considered as a rough objective measure of cognitive load. Concretely, the mean fixation duration in the subtitle areas of interest is considered in the different subtitle languages and two lecture formats. The mean fixation duration on the PowerPoint slide or the talking head is not considered in this analysis. The reason for this is that it would not provide valuable insight as reading is generally comprised of more, shorter fixations whereas scene perception tends to lead to fewer and longer fixations (Kruger & Doherty, 2016). While fixations on the subtitles were generally longer when the lecture showed a talking head (as shown in Table 3), a GLMM shows the lecture format to not have a significant effect on mean fixation duration ($t = 0.715$, $p > 0.05$). Additionally, Figure 2 predicts fixations are shorter for Dutch subtitles, the native language of the viewers. This effect is not significant ($t = -1.295$, $p > 0.05$).

Table 3: Mean fixation duration

	M	*SD*
PowerPoint	223.57	60.05
Talking head	231	59.51
English subtitles	230.05	59.97
Dutch subtitles	225.47	59.70

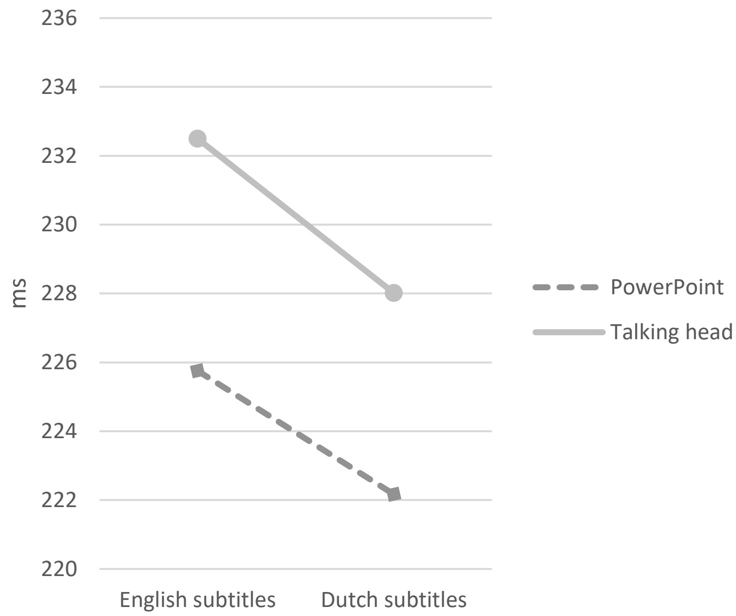


Figure 2: Predicted mean fixation duration

RIDT scores look relatively stable across formats and languages (Table 4). While the effect of subtitle language (t = -1.144, p > 0.05) or lecture style is not significant (t = 0.332, p > 0.05), the interaction between subtitle language and lecture format (represented by the drop in Figure 3) is (t = 2.831, p < 0.01). Additionally, a significant effect is also found for listening comprehension scores. Better listening comprehension scores lead to lower RIDT scores (t = -2.535, p < 0.05), i.e., they read the subtitles less thoroughly.

Table 4: RIDT scores

	M	*SD*	Range
PowerPoint	0.83	0.38	0.08 – 2.34
Talking head	0.87	0.36	0.09 – 2.32
English subtitles	0.87	0.36	0.09 – 2.30
Dutch subtitles	0.84	0.38	0.08 – 2.34

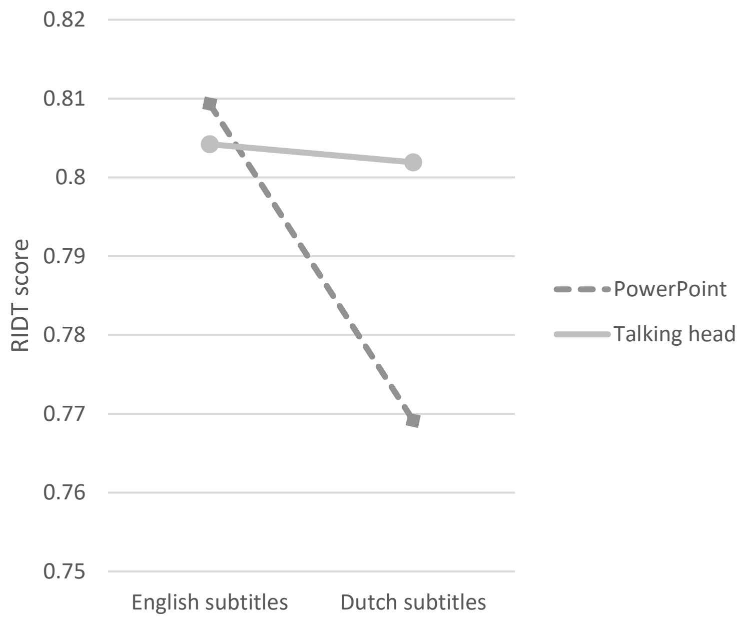


Figure 3: Predicted RIDT scores

4.2 Eye movement on all areas of interest

Earlier we saw that in the lecture showing PowerPoint slides, about 10% of the subtitles were skipped, so the additional visual-textual information on-screen had a clear effect on subtitle reading and it was made clear that participants had to split their attention between the two sources of information. Table 5 shows how barely any content on the slide was skipped when no subtitles were present. The presence of subtitles significantly increased predicted slide skipping (z = 6.070, p < 0.01 - contrast Dutch to no subtitle condition). No significant effect was found for subtitle language (z = 0.615, p > 0.05).

Table 5: Mean slide skipping

	Skipped
English subtitles	7.32%
Dutch subtitles	6.67%
No subtitles	0.67%

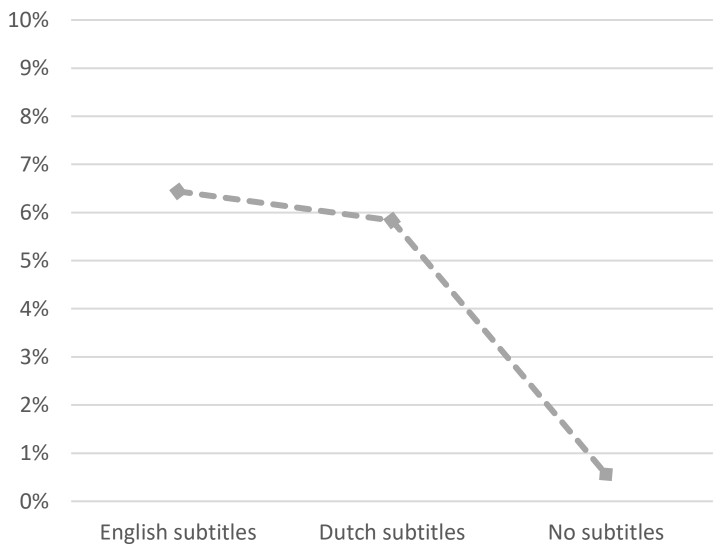


Figure 4: Predicted slide skipping

Figure 5 below shows the predicted effects on total fixation duration comparing the image, i.e., the text on the PowerPoint slides or the talking head, the subtitle areas of interest and other fixations (e.g., on the background behind the talking head or the background of the slides). A linear regression shows that students spent significantly more time on the subtitles than they spent on the talking head of the professor ($t = 10.485$, $p < 0.01$). They also spent more time on the subtitles than on the slides in the lecture with PowerPoint slides ($t = 4.674$, $p < 0.01$). Subtitle language has no effect in the talking head lecture ($t = 0.018$, $p > 0.05$) or the PowerPoint presentation lecture ($t = 0.217$, $p > 0.05$).

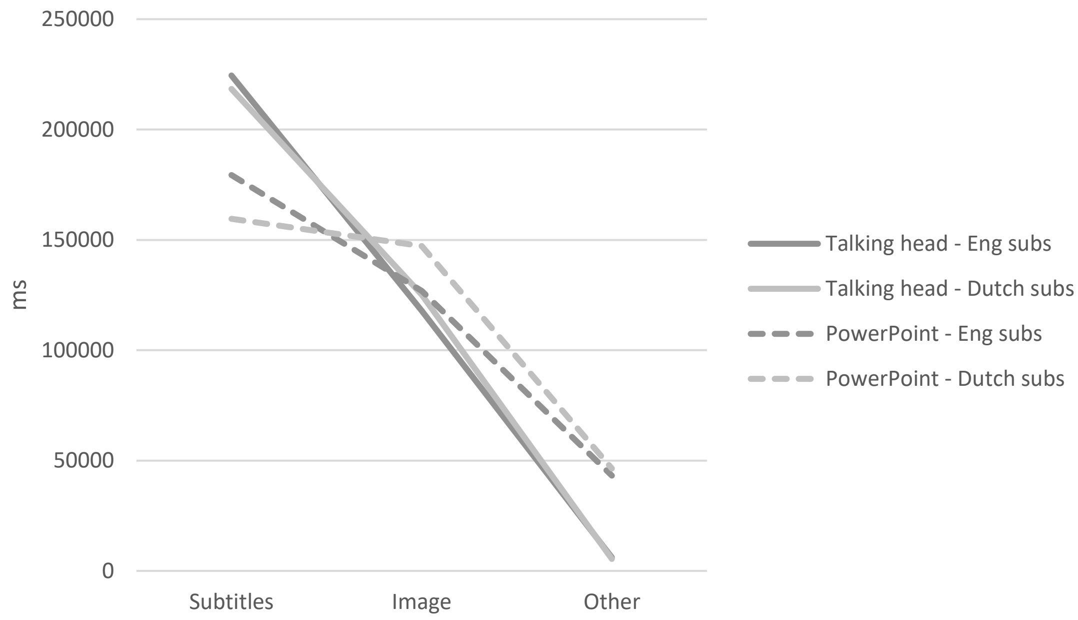


Figure 5: Predicted total fixation duration

Considering fixation counts (Figure 6), a linear regression found a significantly higher number of fixations on the subtitles compared to the talking head ($t = 24.740$, $p < 0.01$) and compared to the PowerPoint slides ($t = 6.595$, $p < 0.01$). There is a notably higher number of fixations on the subtitles in the lecture with a talking head than in the one with PowerPoint slides (Table 7). This signifies the trade-off in additional reading that has to be done in the PowerPoint-styled lecture, whereas the talking head offers little extra information.

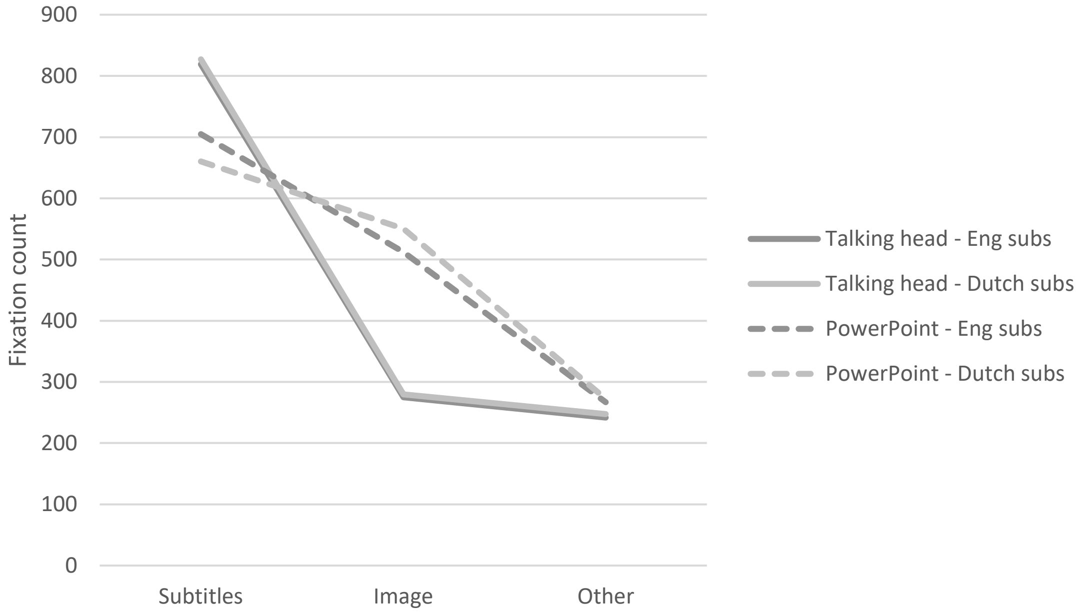


Figure 6: Predicted fixation counts

4.3 Subjective cognitive load

For subjective cognitive load measurement, we used the psychometric questionnaire of Leppink and van den Heuvel (2015). This questionnaire consists of eight questions, the first four entail intrinsic cognitive load/essential processing load, i.e., content complexity, the last four extraneous cognitive load/extraneous processing load, i.e., instructional complexity.

When the first four questions (intrinsic load) are considered (Table 6), the PowerPoint slides seem to cause slightly lower intrinsic loads. One would expect intrinsic load to be relatively constant, as both the addition of subtitles and the change of format does not change the inherent complexity of the content, rather the instructional complexity of the materials, i.e., extraneous load. Regardless, Figure 7 shows a significant effect is predicted going from English to Dutch subtitles ($z = -2.638$, $p < 0.01$) and from Dutch to no subtitles ($z = 2.882$, $p < 0.01$). No significant effect is found changing the style of the lecture ($z = 0.179$, $p > 0.05$). Lastly, the interaction between the change from a talking head lecture to a PowerPoint presentation lecture and English to Dutch subtitles is also significant ($z = -2.287$, $p < 0.05$).

Table 6: Perceived intrinsic load

	M	***SD***
PowerPoint	5	1.96
Talking head	5	2.19
English subtitles	5.5	2.01
Dutch subtitles	5	1.84
No subtitles	6	2.33

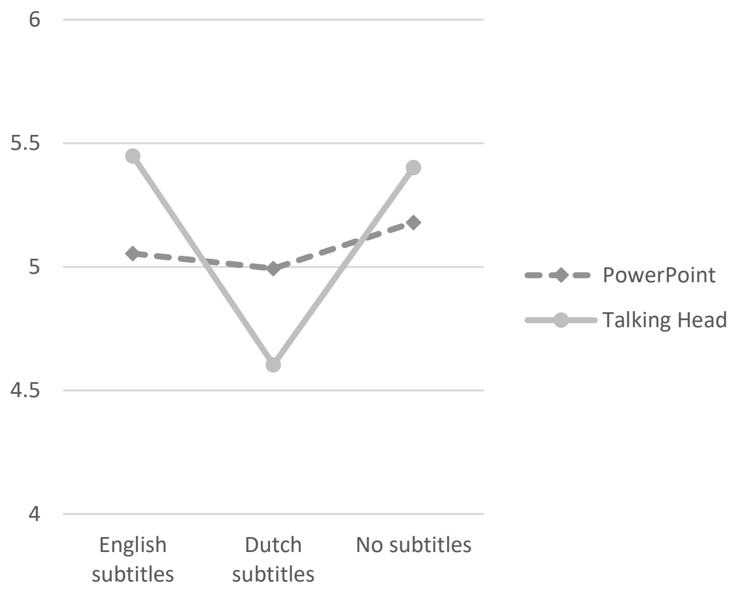


Figure 7: Predicted perceived intrinsic load

For the last four questions, i.e., extraneous load (Table 7), we see a significant effect changing a talking head into a PowerPoint presentation (z = 2.368, $p < 0.05$). While Figure 8 does show Dutch subtitles leading to lower extraneous load ratings in the talking head lecture and higher ratings when PowerPoint slides are shown, the change from English to Dutch (z = -0.016, $p > 0.05$) or from Dutch to no subtitles (z = 0.551, $p > 0.05$) is not significant. For extraneous load, we also do not find a significant interaction between lecture style and subtitle changes.

Table 7: Perceived extraneous load

	M	***SD***
PowerPoint	2	2.03
Talking head	3	2.30
English subtitles	3	2.25
Dutch subtitles	3	1.96
No subtitles	3	2.37

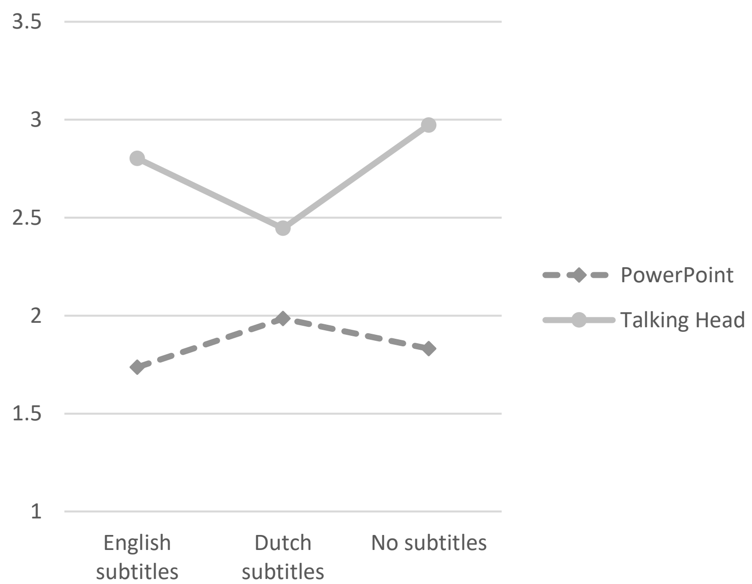


Figure 8: Perceived extraneous load

4.4 Comprehension

As can be seen in Table 8 and Figure 9 below, immediate comprehension was worse for the students who watched the lectures with a talking head (z = -2.791, $p < 0.01$). There were no significant effects of the subtitles, nor were there any significant interactions with lecture style. There was, however, a significant effect of vocabulary size, where a larger vocabulary resulted in better comprehension scores (z = 5.117, $p < 0.01$). When we add the RIDT scores of the participants as a fixed effect, we find no significant effect (z = 0.298, $p > 0.05$). This is a rather surprising find as it opposes what was previously found in a study by Kruger and Steyn (2014), where the presence of subtitles did not improve comprehension but actual reading did. Our results would suggest subtitles in recorded online lectures do not have an effect on immediate comprehension.

Table 8: Immediate comprehension

	M	***SD***	**Range**
PowerPoint	7	1.91	3 – 11
Talking head	6	1.97	1 – 11
English subtitles	7	1.98	2 – 11
Dutch subtitles	7	2.06	2 – 11
No subtitles	6	1.91	1 – 11

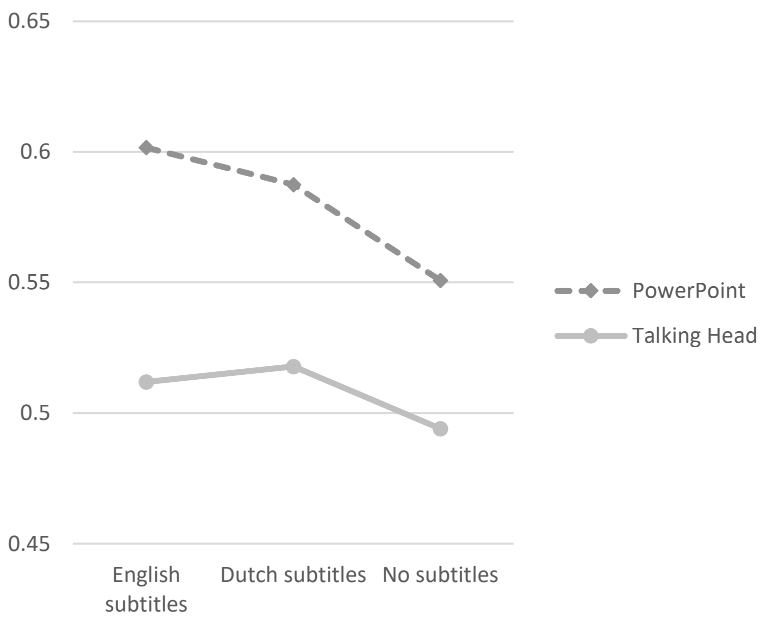


Figure 9: Predicted immediate comprehension

Table 9 and Figure 10 show that the scores are considerably lower for the delayed comprehension test compared to the immediate comprehension test, but the group that watched the lecture with a talking head again scored worse than those who watched a narrated PowerPoint presentation (z = -2.067, $p < 0.05$). Interestingly, we now also see the comprehension scores decrease going from English to Dutch subtitles (z = -2.142, $p < 0.05$). The change from Dutch subtitles to no subtitles is not significant (z = 0.283, $p > 0.05$). We again found a significant positive effect of vocabulary size (z = 4.974, $p < 0.01$) and also found comprehension scores to decrease as time between the immediate and delayed test increased (z = -5.113, $p < 0.01$).

Table 9: Delayed comprehension

	M	*SD*	Range
PowerPoint	5	1.89	0 – 10
Talking head	4	1.64	0 – 8
English subtitles	5	1.78	0 – 8
Dutch subtitles	4	1.76	0 – 9
No subtitles	4	1.81	0 – 10

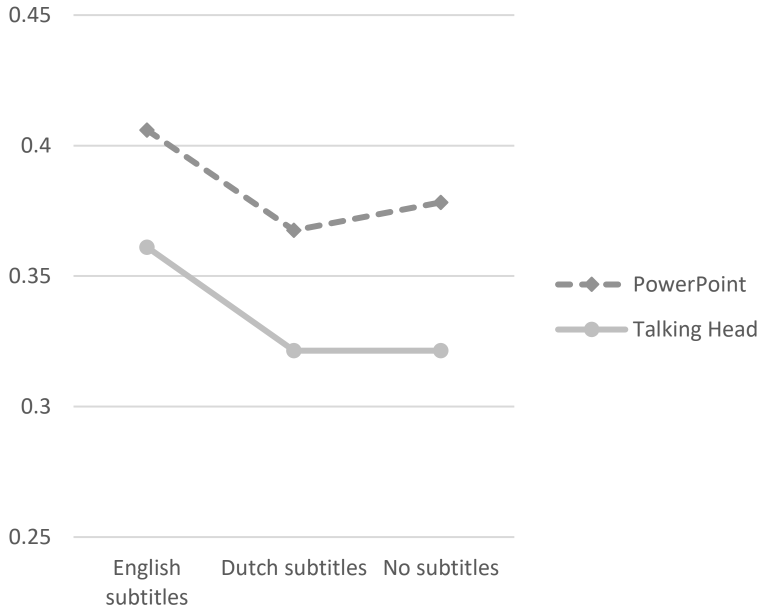


Figure 10: Predicted delayed comprehension

5 Discussion

5.1 Eye movements

The eye movements results in this paper are a clear example of split attention. As more visual-textual sources of information are present on screen, the viewer has to decide where to allocate attention. A trade-off has to occur as paying attention to one source of information comes at the cost of being able to focus on the other sources of information. If subtitles are present on a lecture that includes PowerPoint slides, more subtitles are skipped than in a lecture only including a talking head. In that same PowerPoint-styled lecture, the viewer also skips more content on the slides when subtitles are present than when only a PowerPoint presentation (without subtitles) is shown. This is not affected by the language of the subtitles. While not significant, this is also reflected in the total fixation duration and number of fixations in the subtitle areas of interest. The subtitles are being fixated more in the lecture with a talking head than in the PowerPoint-styled lecture.

Regardless of the differences in subtitle reading between the two lecture styles, subtitles are still being fixated significantly more in both lectures than the contents on the PowerPoint slides or the talking head. This could be expected as subtitles were previously shown to be read automatically (d'Ydewalle & De Bruycker, 2007) and as it is continuously changing information in two arguably static lectures.

If we consider actual reading of the subtitles using the Reading Index for Dynamic Texts (Kruger & Steyn, 2014), the language of the subtitles and the lecture style alone appear to not have any effect on the reading. However, when both are considered, we observe Dutch subtitles to be read significantly less thoroughly in a PowerPoint-styled lecture. This could be another example of split attention, though this effect is not found for English subtitles. Therefore, it is more likely that viewers skimmed the subtitles more when they did not match the language of the verbal information shown on the slides. Additionally, listening comprehension also had a significant effect on reading. As viewers' listening comprehension increased, the reading of the subtitles decreased. With better listening skills, the need for subtitles evidently decreases. We speculate that viewers are aware, either consciously or subconsciously, of their skills and, consequently, rely less on the subtitles as a source of information. This is what Mayer (2011) defines as metacognition. The learner has control of cognitive processing during learning. The influence of listening skills is a clear example of self-managing and learners adjusting learning strategies based on their own needs.

Lastly, we also looked at mean fixation durations on the subtitles as a rough objective indication of cognitive load and active processing. While slightly shorter fixations can be observed for the Dutch subtitles, the native language of the viewers, this difference appears not to be significant. Important to mention is that the average English proficiency of the students in this experiment was rather high. This difference, in line with previous research, is likely to become significant for students with lower proficiencies.

Though not significant, slightly shorter fixations on the subtitles were also observed in the PowerPoint-styled lecture than in the lecture with a talking head. This can be attributed to the fact that viewers had more time to read the subtitles in the talking head lecture as no concurrent visual-textual information was present. It would be interesting to see how reading behavior, including the length of fixations, changes as the amount of concurrent visual-textual information on screen increases, i.e., increasing slide density.

5.2 Subjective cognitive load

Regarding perceived intrinsic load, lecture style is found not to have an effect. With regard to the subtitles, Dutch subtitles are found to significantly decrease the perceived intrinsic load. This effect is marginal for the PowerPoint-style lecture, but significantly stronger when the lecture only contains a talking head. It is clear that, even with a highly proficient group of L2 English-speakers, adding native language subtitles makes the students perceive the lecture to be less complex, even though eye movement data show that they read the Dutch subtitles less thoroughly. The native language subtitles help especially when no other visual-textual information in another language that the audience understands is present on screen. This raises two questions: Firstly, would the effect of native language subtitles on perceived intrinsic load increase when the concurrent visual-textual information is also in the viewer's native language and, secondly, how would this effect change as the viewer's proficiency for the other present language decreases?

For extraneous load, the style of the lecture does play a significant role. A lecture with only a talking head is perceived to be more straining than a PowerPoint-styled lecture, regardless of the subtitle condition, even though the information on the PowerPoint slides can at times be considered redundant. The signaling principle (i.e., directing the learner's attention to what is relevant) is more dominant here. The summarized visual-textual information therefore decreases load instead of increasing it. There is a notable pattern where Dutch subtitles increases perceived extraneous load with concurrent English visual-textual information and decreases load without it. The language of the subtitles, however, has no significant effect on the perceived load, nor does the presence of the subtitles.

5.3 Comprehension

We measured comprehension using a pre-made comprehension test containing both memory questions and questions that require insight. For the immediate test, a significant effect was found for the style of the lecture, where students who watched the lectures including a PowerPoint presentation performed better on the test than those who watched a lecture with a talking head. Even though there was a visible dip in performance when subtitles were not present, the presence and language of subtitles had no significant effect on immediate performance. Lastly, students with higher vocabulary size test scores also performed better on the tests, which underlines the importance of language proficiency in lectures taught in L2 languages.

The effects of lecture style and vocabulary size remained significant even in the delayed test. However, for the delayed test, students performed better when the lectures had English subtitles. Students scored about the same when lectures contained no subtitles or Dutch subtitles. This is a clear example of Dual-Coding, i.e., because the same information is presented both visually and auditory, processing of that information is improved. An important consideration regarding this find is that the language of the test itself was also English. It might be interesting for future research to establish whether test language is the deciding factor in this matter.

Our results also showed subtitle reading to not have a significant effect on comprehension, contrary to what was found in a study by Kruger and Steyn (2014). As for instructional design, based on the results in this paper, the instructor's presence plays a less important role than the additional information present on the slides in a PowerPoint. While we expected the presence of subtitles in addition to text on PowerPoint slides to hamper the comprehension of students due to the screen being clogged with text, students are apparently perfectly capable of focusing on both and filtering relevant information in such a situation. This shows how the current generation of students, arguably students who are very used to subtitles, have strong control over their learning process, i.e., metacognition (Mayer, 2011), when presented with subtitled educational video.

6 Conclusion

This paper sheds light on the interactions and consequent effects of subtitles and visual-textual complexity of educational video on cognitive processing of students. Specifically, it attempts to answer the question of how the presence and language of subtitles impact processing, cognitive load and comprehension in different styles of online lectures. Especially the latter part, namely how the style of the lecture moderates the effects of the subtitles, has not yet been thoroughly examined.

We found that when subtitles were present on the lecture, significantly more time was spent on them than on the talking head or the content on the slides. The language of the subtitles only seems to matter when there is other concurrent on-screen text in another language. In that case, more attention is devoted to the text that is in the same language as the narration, which was the text shown on the PowerPoint slides in the present study.

This, however, did not have any effect on the perceived cognitive load. Perceived cognitive load for the lecture containing a PowerPoint presentation remained relatively constant regardless of whether English, Dutch or no subtitles were shown. While one could initially expect the combination of subtitles and on-screen text to increase cognitive load and potentially lead to cognitive overload because of the sheer amount of redundant information the learner is presented with, it does not in this case. Even more so, the PowerPoint presentation lecture induces lower cognitive loads than the lecture with a talking head. Only when Dutch subtitles are included does the lecture with a talking head come close to the perceived cognitive load in the PowerPoint presentation lecture. The strength of the PowerPoint presentation is probably due to the dominance of the signaling effect. Though the learner is presented with redundant information, they are directed to what is relevant for the lecture at hand and can use it as a guide. A potential reason why even subtitles do not affect the strength of a PowerPoint might lie in the audience. The current generation is constantly exposed to subtitles. Consequently, they may have become experts at dealing with subtitled content and have developed strategies to cope with both subtitles and large amounts of information on screen (cf. metacognition Mayer, 2011).

In line with the consistently lower perceived cognitive loads for a lecture that includes a PowerPoint presentation, student performance also increases. The presence or language of subtitles has no effect on immediate performance, but on a long-term basis English subtitles do seem to help students to tie in the information with previous knowledge and store it in their long-term memory.

Tying these comprehension and cognitive load results in with the eye tracking data, brings us to three provisional conclusions: Firstly, native language, interlingual subtitles can decrease cognitive load as long as the lecture includes no other relevant (verbal) information. This is not the case for L2, intralingual subtitles, regardless of whether they are both read to the same extent. While this did not have any effect on comprehension scores in this study, we believe one possible explanation for this was that the test itself was entirely in English. Secondly, an effective PowerPoint presentation decreases the perceived cognitive load of students in a lecture, indicating the signaling effect overrides the redundancy effect in a way. Adding subtitles alters the reading behavior of students and makes them divide their attention between both sources of information, but they are able to do so effectively and are not overloaded by the information. We believe this emphasizes the role of metacognition in multimedia learning (Mayer, 2011, 2014a). Learners know how to manage their cognitive processing and how they do this evidently has an effect on the cognitive load they experience. We do not expect this to be a consistent find. Rather, depending on slide density and the linguistic proficiency of the students, we expect there to be a tipping point where subtitles do in fact cause cognitive overload with a potentially detrimental effect on comprehension in the case of lower proficiency students. In any case, our finding calls for a significant overhaul of the redundancy principle . Lastly, English subtitles in combination with English slide content improves long-term memory storage (in English).

There are, however, a number of limitations that should be taken into account when interpreting the results of the current study. Firstly, the average linguistic proficiency of the present group is high. Considering we found listening comprehension and vocabulary size to be significant predictors of eye movements, perceived cognitive load and comprehension test scores, it would be interesting to consider students with lower proficiency levels or, ideally, conduct a study on a group of students with a large variation of linguistic proficiency levels. Secondly, while the present paper includes the Reading Index for Dynamic Text (Kruger & Steyn, 2014) as a measure of reading, the eye tracking system used includes no word-level data. This prevents us from making more fundamental statements about reading behavior. More research is needed to study how reading behavior, including word skipping, reading speed, word frequency and word length effects, etc., changes based on subtitle presence/language in differently styled lectures.

7 References

Chan, W. S., Kruger, J.-L., & Doherty, S. (2019). Comparing the impact of automatically generated and corrected subtitles on cognitive load and learning in a first- and second-language educational context. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*(18), 237–272. doi:10.52034/lanstts.v18i0.506

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94*(2), 428–434. doi:10.1037//0022-0663.94.2.428

d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist, 12*(3), 196–205. doi:10.1027/1016-9040.12.3.196

Diao, Y., Chandler, P., & Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology, 120*(2), 237–261. doi:10.2307/20445397

Hefer, E. (2013a). Reading first and second language subtitles: Sesotho viewers reading in Sesotho and English. *Southern African Linguistics and Applied Language Studies, 31*(3), 359–373. doi:10.2989/16073614.2013.837610

Hefer, E. (2013b). Reading second language subtitles: A case study of Afrikaans viewers reading in Afrikaans and English. *Perspectives, 21*(1), 22–41. doi:10.1080/0907676x.2012.722652

Hosogoshi, K. (2016). Effects of captions and subtitles on the listening process: Insights from EFL learners' listening strategies. *JALT CALL Journal, 12*(3), 153–178. doi:10.29140/jaltcall.v12n3.206

Kalyuga, S. (2012). Instructional benefits of spoken words: A review of cognitive load factors. *Educational Research Review, 7*(2), 145–159. doi:10.1016/j.edurev.2011.12.002

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Vol. 2, pp. 247–262). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9781139547369.013

Kruger, J.-L., & Doherty, S. (2016). Measuring cognitive load in the presence of educational video: Towards a multimodal methodology. *Australasian Journal of Educational Technology, 32*(6), 19–31. doi:10.14742/ajet.3084

Kruger, J.-L., Hefer, E., & Matthew, G. (2013). *Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts.* Paper presented at the 2013 Conference on Eye Tracking South Africa, Cape Town, South Africa. doi:10.1145/2509315.2509331

Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research, 7*(5), 1–15. doi:10.16910/jemr.7.5.4

Kruger, J.-L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly, 49*(1), 105–120. doi:10.1002/rrq.59

Lavaur, J.-M., & Bairstow, D. (2011). Languages on the screen: Is film comprehension related to the viewers' fluency level and to the language in the subtitles? *International Journal of Psychology, 46*(6), 455–462. doi:10.1080/00207594.2011.565343

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25*, 943–951. doi:10.1002/acp.1787

Leahy, W., & Sweller, J. (2015). Cognitive load theory and the effects of transient information on the modality effect. *Instructional Science, 44*, 107–123. doi:10.1007/s11251-015-9362-9

Leppink, J., & van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education, 4*(3), 119–127. doi:10.1007/s40037-015-0192-x

Liao, S., Kruger, J.-L., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of Specialised Translation*(33), 70–98.

Liao, S., Yu, L., Reichle, E. D., & Kruger, J.-L. (2021). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading, 25*(5), 417–435. doi:10.1080/10888438.2020.1823986

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Vol. 2, pp. 227–246). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9781139547369.012

Mayer, R. E. (2011). *Applying the Science of Learning*. London, England: Pearson. doi:10.1016/B978-0-12-387691-1.00003-X

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43–71). New York, NY: Cambridge University Press. doi:10.1017/CBO9781139547369.005

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187–198. doi:10.1037//0022-0663.93.1.18

Paas, F. G. W. C., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63–71. doi:10.1207/s15326985ep3801_8

Rajendran, D. J., Duchowski, A. T., Orero, P., Martínez, J., & Romero-Fresco, P. (2013). Effects of text chunking on subtitling: A quantitative and qualitative examination. *Perspectives, 21*(1), 5–21. doi:10.1080/0907676X.2012.722651

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55–88. doi:10.1177/026553220101800103

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285. doi:10.1207/s15516709cog1202_4

Sweller, J., Kalyuga, S., & Ayres, P. (2011). *Cognitive load theory* (1 ed.). New York, NY: Springer. doi:10.1007/978-1-4419-8126-4

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. doi:1040-726X/98/0900-0251

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292. doi:10.1007/s10648-019-09465-5

van der Zee, T., Admiraal, W., Paas, F., Saab, N., & Giesbers, B. (2017). Effects of subtitles, complexity, and language proficiency on learning from education videos. *Journal of Media Psychology, 29*(1), 18–30. doi:10.1027/1864-1105/a000208

Van Hoecke, S. M., Schrijver, I., & Robert, I. S. (2022a). Methodological preparation of a within-subject audiovisual cognition, reception and perception study. *Journal of Audiovisual Translation, 5*(1), 94–128.

Van Hoecke, S. M., Schrijver, I., & Robert, I. S. (2022b). Preparing and comparing subtitles for quasi-experimental and experimental research in audiovisual translation studies. *Translation Spaces, 11*(1), 113–133

Paper 4 – Perception of subtitles in online education: What does the learner want?

This paper was submitted to Across Languages and Cultures (2023):

Van Hoecke, S. M. (under review). Perception of subtitles in online education: What does the learner want? *Across Languages and Cultures.*

Formatting has been slightly altered to fit the present thesis.

Abstract

Audiovisual Translation (AVT) is still a relatively young and growing research discipline. For some time, the focus has been on accessibility and language learning. However, in the last decade, more studies have emerged that examine the reception of AVT. Reception research investigates how a specific audience consumes AVT and how AVT affects viewing and reading patterns, cognition and comprehension, among others. It is, however, also vital to assess the audience's perception as any proven benefit of AVT revealed in reception studies may go to waste if the audience appears to simply not want to use AVT. The present paper wishes to contribute to the field by interviewing students that were shown three English-taught lectures with either no subtitles, L2 English subtitles or L1 Dutch subtitles. The students were asked how they experienced the subtitled lectures, what effects they thought the subtitles had on them and whether they would like to have subtitles for lectures in the future. The results of this paper are a clear endorsement of subtitles in education. Subtitles are generally appreciated by students as they are perceived to aid with meaning construction and make viewing more comfortable.

Key words

Perception, subtitles, educational design, audience, qualitative research

1 Introduction

In the past decades, the importance of audiovisual translation (AVT) has grown following the increasing emphasis on overall accessibility and inclusivity and the commercial pressure to reach larger, multilingual audiences. Even more so, in the aftermath of the pandemic, during which 'remote' became key and multimedia played an even bigger role in everyday life, the need for AVT has sky-rocketed. Consequently, research into AVT has never been more important. While the research field of AVT has been steadily developing since the turn of the millennium, the crucial focal point of the audience, studied in reception and perception studies, has only begun to receive attention in the last decade (Di Giovanni & Gambier, 2018).

Because the line between reception and perception of AVT is thin and the terms are easily confused, it is important to clarify how this paper defines perception and distinguishes it from reception. Reception of AVT concerns the ways in which AVT is processed, consumed, understood and retained (Gambier, 2018). Studies into the reception of AVT examine, for example, the reading of AVT, how viewers watch video with AVT, whether AVT affects comprehension or cognitive load or how the presence and the characteristics of AVT play a role in the viewing of audiovisual material (for an overview see Di Giovanni & Gambier, 2018; Kruger, 2016; Liao & Kruger, 2023; among others). Perception, on the other hand, envelops how viewers think they watch film, what effects viewers believe AVT has on them and what the audience prefers with regard to AVT. As defined by Gambier (2018, p. 56): "Perception is made of opinions and impressions and varies over time." It is important to mention that this is also distinctly different from perception in other fields of research, such as psychology, where perception is defined as the cognitive processing and understanding of sensory information (Grondin, 2016).

An increasing number of reception studies have been conducted in AVT over the past decade (e.g., Gerber-Morón et al., 2018; Hefer, 2013a, 2013b; Kruger et al., 2013, 2014; Kruger et al., 2022; Liao et al., 2022; Liao et al., 2021; Montero Perez et al., 2014; Perego et al., 2010; Szarkowska & Bogucka, 2019; Szarkowska et al., 2021; Van Hoecke & Kruger, forthcoming; Van Hoecke et al., forthcoming). These studies, however, rarely focus on the viewer's habits, preferences and attitudes towards AVT. The audience is, however, a vital component for AVT. Research may prove numerous benefits gained from AVT, but if the audience does not perceive these as such and is consequently not willing to use AVT, these potential benefits are lost. We thus argue that perception is equally important. In the past decade, there have also been quite some studies focusing on perception (e.g., Burnham et al., 2010; Kruger & Mazur, 2012; Orrego-Carmona, 2018; Perego et al., 2016; Rajendran et al., 2013; Romero-Fresco, 2012; Taylor, 2008; Szarkowska & Boczkowska, 2022; Szarkowska et al., 2021), but few have focused on perception of subtitles in education (e.g., Montero Perez et al., 2013).

This paper wishes to contribute to the field by reporting on a study that examines the effects of the presence and language of subtitles in lectures that use English as a medium of instruction (EMI). We showed 83 undergraduate students at the University of Antwerp three EMI lectures with English/intralingual, Dutch/interlingual or no subtitles in either a talking head style or a voice-over PowerPoint style. We then interviewed some of these students, asking them about their use of subtitles, how they thought the subtitles impacted them, their preferences and their watching habits at home. This paper aims to provide valuable insight into the opinion of the hearing student on subtitles in education, answering the key question: What does the learner want?

2 AVT reception and perception research

AVT reception research has so far mainly focused on benefits for language learning audiences and audiences with disabilities (Díaz Cintas, 2020). In the past years, however, there has also been a growing interest in the other benefits of AVT, more specifically those of subtitles, for hearing audiences in different contexts, such as education. These studies tend to focus on the processing of subtitles (e.g., Kruger et al., 2022; Liao et al., 2022; Liao et al., 2021; Van Hoecke & Kruger, forthcoming) or the effects of subtitles on comprehension or cognitive load (e.g., Hosogoshi, 2016; Kruger & Doherty, 2016; Kruger et al., 2014; Montero Perez et al., 2014; Vulchanova et al., 2015; Van Hoecke et al., forthcoming). Perception of subtitles has also been studied extensively (e.g., Kruger & Mazur, 2012; Linder, 2016; Orrego-Carmona, 2018; Perego et al., 2016; Szarkowska & Boczkowska, 2022; Szarkowska et al., 2021). Perception of AVT is often not the sole focus of a study, likely due to the fine line between reception and perception and because researchers attempt to maximize data collected during experiments. Regardless, we support any research that does consider perception as well, as we argue the audience's perception of AVT should be considered at least as important as any effects measured in reception studies.

When we look at studies that consider perception of AVT, one aspect that has received considerable attention is the viewer's attitude towards subtitled vs. dubbed multimedia (e.g., Deckert & Bogucki, 2019; Di Giovanni, 2016; Perego, Laskowska, et al., 2016; Szarkowska & Laskowska, 2015). For this paper, however, we are particularly interested in subtitles.

First and foremost, it is important to know how often subtitles (for the deaf and hard of hearing) or captions are used in general. In 2000, the Television Caption Users Survey was conducted in Australia (Burnham et al., 2010). The survey received 1,311 complete responses. It showed that between 66% and 70% of the viewers used captions at home when available. The number increased as the viewers' experience with captions (having used captions before) increased and as the hearing level of users decreased. The age of the viewer was also an important factor. Eighty-one percent of the viewers under the age of 26 used captions when available while this was only 67% of the users above the age of 60. More recently, Stagetext, a deaf-led charity aimed at boosting accessible media in the UK, and Sapio Research collaborated and surveyed 2,003 people in the UK about their use of subtitles (Youngs, 2021). The results were similar to those in the Australian survey. A large proportion of viewers, and especially the younger generation, used subtitles regularly. Between the ages of 18 and 24, 80% reported they used subtitles on a regular basis. Between the ages of 46 and 55, only 37% used subtitles often. These results were yet again confirmed in a survey conducted by Preply in the United States (Zajechowski, 2022). Of 1,260 respondents, 50% replied they watched content with subtitles most of the time. In the 15-25 age group, this number went up to 70%. Additionally, this survey asked viewers to explain why they used subtitles. The most common reasons were to understand muddled audio or hard to understand accents, to be able to watch a show quietly at home, to stay focused on the screen or to learn a new language.

Each of these surveys make clear that subtitles are ubiquitous and especially popular among younger generations. They are generally positively received by the audience and the large majority of viewers tends to make use of them when watching television at home. However, the context of television is distinctly different from a purely educational context. Furthermore, while the Preply survey (Zajechowski, 2022) does provide some information as to why users turn on subtitles and thus what effects users believe subtitles have on them, the insight gained from them on this topic of perception remains limited.

As mentioned before, perception of subtitles is rarely the main goal of a study. Of the studies included in this review, this paper only discusses the results that concern perception for the sake of brevity. One study that examined perception was conducted by Orrego-Carmona (2014). He investigated the audience's reception and perception of non-professional and professional subtitles. While the audience mentioned that non-professional subtitles available online tend to be of low quality, they were unable to identify differences between non-professional and professional subtitles during the experiment. To gauge the quality of subtitles, the audience only mentioned speed and appearance as key factors. Furthermore, in all subtitle conditions participants indicated they had good comprehension of the material, even when their comprehension test scores showed different levels of comprehension. This shows that participants' perceptions of subtitles do not necessarily correspond to reality or, in this case, that comprehension defined by participants may not match the level of comprehension required to successfully complete the test.

Timing, speed and appearance are factors that are frequently mentioned when the quality of prerecorded and live subtitles is assessed by viewers. Di Giovanni (2016) examined the audience's reception and perception of subtitles at three different film festivals. The audience indicated that two of the most important factors for the perceived quality of subtitles were a good translation and synchronicity. In a survey on *live* subtitles for the deaf and hard of hearing in the UK (cf. Romero-Fresco, 2012), 80% of the participants replied they had experienced problems with subtitles before. The main issues are, surprisingly, not the unavailability of subtitles (7%), but instead the delay with the audio (25%) and the inaccuracy of the subtitles (17%). To shed more light on the matter, Romero-Fresco (2012) also examined the reception and perception of live subtitles in the UK. They collected 434 responses from UK viewers, 423 of whom had hearing impairments. The majority of the participants were above the age of 60 (58.7%) or at least above the age of 35 (91.7% in total). When asked what they used subtitles for, the two main answers were to gain access to the dialogue or to help them understand. In line with the survey from the Royal National Institute for the Deaf, a large portion (55%) of the participants indicated that live subtitles could be better. The main issue appeared to be the delay, which, according to the participants, can and should be reduced as much as possible. The group of participants was divided when it came to how subtitles should be presented, as blocks or word by word.

The presentation mode of pre-recorded subtitles was also investigated by Rajendran et al. (2013). They investigated the effects of text chunking on processing speed and comprehension of subtitled video. They showed participants four videos, one without segmentation, one with word-by-word segmentation, one with phrase by phrase segmentation and one with sentence-by-sentence segmentation. Confirming the findings of Romero-Fresco (2012), they found the audience did not have a clear preference with regard to segmentation.

Another aspect of presenting subtitles is the placement. They can be at the bottom of the screen or can also be integrated with the content. Black (2022) studied children's reception of content with integrated and non-integrated interlingual subtitles. They found that integrated subtitles were approved by children and were found easy to read and understand.

The studies mentioned above highlight how characteristics of subtitles, such as their presentation speed, the delay in the case of live subtitles, or the presentation mode and segmentation can influence the audience's perception. These all have important implications for the use of subtitles in education, especially if they were to be live subtitles. Recorded lectures, however, can be subtitled in advance. It would therefore also be interesting to examine how conventional subtitles are perceived.

Kruger et al. (2017) studied the effects of subtitles on transportation, immersion, character identification, presence and perceived realism. Using subtitled and non-subtitled conditions with participants of a variety of native languages, they found subtitles to increase viewer's perceived transportation, character identification and realism. This would imply that subtitles are not perceived as a distraction by the audience, but instead facilitate the involvement of the audience in the story. The native language of the audience had no impact.

In education specifically, it would, again, be interesting to examine the current use of subtitles. Linder (2016) conducted a survey with 2,124 student participants above the age of 18. First and foremost, this survey revealed almost all participants had at least one course that included some kind of audiovisual content. Half of these students stated they were aware of closed captions being available for the videos in their courses. Approximately 35% of the students indicated they always or regularly used closed captioning during online courses when available. The main reasons for this were to help them focus, retain information or overcome poor audio quality. This survey shows the applicability of AVT in education and also that a large portion of the students, if aware of the option, would turn on subtitles whenever possible.

Kruger et al. (2017) showed higher involvement in the presence of subtitles. This can also be useful when watching educational video as it could be reflected in higher engagement. Dommett et al. (2022) studied the perceived value of subtitles and transcripts in recorded lectures during a full semester. Similar to the results in Kruger et al. (2017), subtitles increased engagement with the lectures. However, they only did in the second half of the semester, which highlights the possibility that for subtitles to be perceived as useful, students first have to get accustomed to them. Subtitles were also found to be generally appreciated by students. Students reported that the subtitles made it easier to find certain parts in the recordings and improved their understanding of the lecture and their learning experience.

Subtitles were also positively received in the study by Taylor (2008). They examined processing strategies of students learning Spanish when watching subtitled Spanish video. In their experiment they also evaluated students' attitudes towards subtitles. They found that the majority of students both with higher and lower proficiencies of Spanish had a positive attitude towards subtitles and found subtitles helpful. Some even reported that they would not have understood anything if subtitles were omitted. One issue, however, was that students were revealed to rely less on the audio and more on the subtitles instead. This result was replicated by Hayati and Mohmedi (2011). While investigating listening comprehension of students in subtitled and non-subtitled English media, the group of students who had interlingual L1 Persian subtitles reported that the subtitles distracted them from the audio and that they started relying more on the subtitles and less on the soundtrack. The students who had intralingual L2 English subtitles, however, reacted very positively to the subtitles. The students believed the subtitles aided their understanding and test performance, in addition to their improved ability to comprehend, spell and recall new English words.

Finally, Montero Perez et al. (2013) studied the effectiveness of different types of subtitles (full captions vs. keyword captions) for listening comprehension. In terms of perceived usefulness, students seemed to be confident that they could understand the material without subtitles, yet most would activate them if they were available. Subtitles were found to be useful and not considered distracting. Some of the most common reasons were: (1) to improve understanding; (2) to deal with accents, unclear pronunciation and background noise; (3) to have a more relaxed viewing experience; (4) because bimodal input makes the content easier to understand; and (5) because reading is easier than listening.

The review above draws a picture that is clearly in favor of subtitles. Viewers of television or students exposed to subtitled instructional videos tend to approve of subtitles. The audience perceives them as ways to improve comprehension, deal with difficult speech, support focus and offer another mode to access the content. However, it is clear that the quality of the subtitles, the presentation mode (to a lesser extent) and the language of the source and subtitles play a vital role in how subtitles are perceived. This paper wishes to contribute to the state-of-the-art by interviewing students after they had to watch a number of subtitled lectures. In doing so, the paper aims to provide further insight into how students perceive subtitles, examining the effects they think subtitles have on them, if and why they would prefer to have subtitles and what language they would like to have subtitles in.

3 Method

3.1 Participants

Sixty students from the second year of Applied Linguistics at the University of Antwerp participated in this study. Their ages ranged between 19 and 26. Thirteen participants were male. Their native languages were Dutch, with two exceptions who had native languages other than English but near-native proficiency of Dutch. The study was approved by the Ethics Committee for the Social Sciences and Humanities of the University of Antwerp.

3.2 Design

The study employed a 2 x 3 mixed-methods design (2 styles of lectures x 3 subtitle conditions). Participants were first divided into two groups. One group would watch all lectures as a voice-over PowerPoint lecture (N = 32); the other group would watch all lectures as a talking head lecture (N = 28). Participants were then invited to attend an individual session in the eye tracking laboratory at the University of Antwerp. They were shown a total of three recorded lectures about philosophy of approximately seven minutes long. The lectures were taught in English. For each participant, one lecture was subtitled in English (intralingual), one lecture was subtitled in Dutch (interlingual) and one lecture did not include subtitles. The comparability of the lectures and the subtitle tracks was thoroughly tested following the ten-step approach (Van Hoecke et al., 2022a, 2022b). The order of the lectures was randomized following a Latin Square design.

After the participants had watched all three lectures and completed a number of tests (reported in Van Hoecke et al., forthcoming), they were interviewed for approximately 20 minutes about their experiences with subtitles in the experiment and at home, their preferences with regard to the presence and language of subtitles in education and their opinion regarding subtitles in the specific style of lectures. Using the data gathered in these interviews, this paper aims to shed light on the students' preferences and perception of subtitles in online education.

3.3 Analyses

Interviews were first transcribed and translated to English. Thematic analysis was then used to identify themes in the transcripts and examine the opinion of students on these themes. Certain questions and their corresponding answers could be quantified. The paper uses this quantified thematic data and supports this with excerpts from the qualitative data. In doing so, it wishes to convey the general opinions and elucidate relevant nuances pertaining to the topic at hand.

4 Results

We set out to examine the students' preference and perception of subtitles when watching recorded lectures taught in L2 English. Firstly, we asked the students whether they believed that they had read the subtitles often. Table 1 shows the perceived frequency of reading. Only a small minority of students reported to have read the subtitles only rarely. This draws a clear picture that when subtitles are present, students think they make regular use of them. In both lecture styles, students believed that they had read the English subtitles more often than the

Dutch subtitles. When asked about their reading frequency of the subtitles in both languages, 25% of the students first reported to not have noticed the different language of the subtitles between the two subtitled lectures. This could be a sign of being accustomed to subtitles in both languages and the reading thereof and/or having sufficient proficiency in English to no longer experience the language as a barrier or obstacle. For these students, we did ask them whether they believed that they had read the subtitles in one of the lectures more than in the other to gauge potential conscious or unconscious bias. If this was not the case, the reported answer was considered the same for both languages. Interestingly, the students that watched the talking head lectures reported they read the Dutch subtitles more than those that watched the PowerPoint lecture.

Table 1: Perceived subtitle reading behavior

		Always	Very often	Often	Rarely	Very rarely
PPT	ENG	21.88%	53.13%	21.88%	0%	3.13%
	NL	15.63%	28.13%	37.50%	9.38%	9.38%
TH	ENG	35.71%	25%	28.57%	10.71%	0%
	NL	25%	35.71%	32.14%	0%	7.14%

To complete this picture, we asked them why they read the subtitles. We identified a number of themes and subthemes in the data and performed frequency counts per theme and subtheme. Participants had the option to mention multiple reasons, hence the sum of the percentages in Table 2 surpasses 100%. Table 2 shows the percentages of participants in each lecture style that mentioned the adjacent perceived benefit gained by reading the subtitles.

Table 2: Perceived benefits of reading subtitles

		PPT	TH
Meaning construction	Total	75.00%	89.66%
	To better understand the content and the speech	37.50%	67.86%
	Because bimodal input (reading and listening) is easier	53.13%	25.00%
	To better retain content	25.00%	46.43%
Viewer's comfort	Total	50.00%	55.17%
	To make following the stream of information easier	25.00%	50.00%
	To improve focus and minimize distractions	34.38%	17.86%
	To look at something other than the main content on screen	12.50%	17.86%
Automaticity	Because it is a habit and happens automatically	34.38%	25.00%
Vocabulary acquisition	To foster vocabulary acquisition or learn spelling	18.75%	21.43%
Experimental design	Because of the use of an eye tracker or the order of the conditions	3.13%	3.57%

The first theme we identified is meaning construction (similar to Montero Perez et al., 2013). The large majority of students, 75% in the PowerPoint condition and a staggering 90% in the talking head lecture, mentioned meaning construction as one of the reasons to read the subtitles. It is subdivided into three subcategories. The first category is: to understand the content and the speech better. This appears to be a major reason to read the subtitles in the talking head lecture. Students mentioned this benefit for the Dutch subtitles:

> "Especially with the Dutch subtitles, I could really understand everything. The more difficult words he said I could understand as well and could follow better because I had the Dutch translation." (P17 – TH)

And for the English subtitles:

> "When there were words that I did not really understand I could link it to what it looks like so that I could still understand." (P76 – TH)

A second commonly perceived benefit was the fact that subtitles present the already available auditory information visually. It presents the same information in two modes and allows this to be processed in the visual and auditory channel simultaneously. While this is the subject of a very lively debate in research (dual-coding vs. redundancy) that goes beyond the scope of this paper, it is interesting to note that students themselves report this bimodal input as a benefit:

> "I just thought it [reading the subtitles] was easier. You hear it and you see it as well, both in English and in Dutch" (P64 – PPT)

> "I guess I always think it's more comfortable to also have a text when someone is talking. I can pay attention better and process it better." (P74 – PPT)

While the majority of students perceived the bimodal input as a benefit, mainly in the PowerPoint lecture, not all students prefer to have subtitles. Some, who reported they did not use subtitles at home either, considered subtitles to be a distraction and the bimodal input to have a detrimental impact on their processing:

> "I noticed that for me it's quite confusing to listen to something and read the same again at the same time. I should either read the subtitles or focus solely on what is being said." (P89 – PPT)

A final subcategory of meaning construction is retention. We see similar patterns as in the category of improved understanding. A large number of students perceive subtitles to aid with retention and, again, this is mainly the case for the talking head lecture.

> "I just take it in better because I can really see the words. That way I can memorize it better." (P66 – TH)

Combined with the first subcategory, it seems that subtitles serve as a replacement for the PowerPoint in terms of anchor points. They are perceived to be less important when these anchor points are already present. This was also mentioned by the students:

> "It would have been better to have a PowerPoint. I think the subtitles are kind of like a replacement for a PowerPoint in this case." (P83 – TH)

The second theme we identified concerns viewer's comfort. It was mentioned by about half of the students as a reason to read the subtitles in both lecture conditions. It is subdivided into three categories: (1) ease of following; (2) improved focus; and (3) alternative focal point. Firstly, students report the subtitles serve as something to hold on to or fall back on. Subtitles make it easier to follow the stream of information and not lose track. The second and third category of viewer's comfort lie close to one another. Since subtitles are dynamic and constantly changing on screen, students perceive them as a feature that increases their focus and makes the entire lecture more interesting. If the lecture does not offer enough information, e.g., a PowerPoint slide that remains on screen for a significant period of time or a talking head of the instructor that does not offer extra information, the subtitles add an additional element to focus on.

> "I enjoyed it more when subtitles where present because then I am actively doing something. It works better than just staring at him." (P57 – TH)

"I get a little bit awkward because of the eye contact." (P19 – TH)

About one fourth of the students reported they also read the subtitles because this happens automatically. The automatic reading of subtitles has been shown in previous research (d'Ydewalle & De Bruycker, 2007). For some students, this automaticity dominates over any other potential benefit to read the subtitles.

"To be honest, I think just because it's a habit. I can hear perfectly what he is saying and understand everything and still I read the subtitles." (P57 – TH)

"The text attracted my attention. It keeps changing so you automatically go to the translation. You can hear the voice, you can hear the words and so you automatically read the subtitles like you would for a movie." (P72 – PPT)

Lastly, another frequently proven benefit of subtitles is language learning and vocabulary acquisition (Bird & Williams, 2002; Danan, 2004; Montero Perez, 2020). While not all students report it as a reason to read the subtitles, some do perceive it as a benefit and are interested in learning new words, terms and spelling.

"Sometimes I look at the subtitles to learn a new word or when a nice word is being used that I wouldn't use myself." (P41 – PPT)

"[I read the subtitles] to see if I understood it properly and to check specific terms so I know how to write them." (P6 – TH)

It warrants mentioning that the participants in this experiment are language students and therefore automatically considered to be more interested in language than the average student. When asked specifically whether students believed they had learned new vocabulary because of the subtitles, 41.94% of the students in the PowerPoint condition and 68.97% of the students in the talking head condition answered yes.

Important to note is that two students reported to have altered their reading behavior because of the experimental setting. One student was very aware of the eye tracker present in the room while the other frequently read the subtitles in the first lecture, felt disappointed after the test and consequently avoided the subtitles in the second lecture. It highlights the need for researchers to be aware of the impact of an experimental setting and the possible intrusive use of eye trackers.

While the perceived benefits of subtitles mentioned above weigh heavily in favor of the use of subtitles in education, not all students perceived the subtitles to be helpful. Table 3 shows that about 10-20% of the students see no gain in having subtitles present. This seems to be especially the case for Dutch subtitles. Furthermore, Table 3 shows that the overall helpfulness of the English subtitles in the talking head lecture is considerably higher than the perceived benefit with regard to understanding the content of the lecture better. This, again, highlights the subtitles potentially being used as a replacement PowerPoint, which is perceived to assist in following the lecture and increasing retention. The opposite pattern can be observed in the PowerPoint lecture, where the Dutch subtitles are overall less helpful whereas they are perceived to aid in understanding the content better. Thus, they may not be perceived to help with following the lecture because a PowerPoint is already present or with retention for the English test, but do allow students to understand everything in the lecture as they offer a translation to their native language.

Table 3: Perceived helpfulness of subtitles

		General			Understanding		
		Yes	Neutral	No	Yes	Neutral	No
PPT	ENG	67.74%	19.35%	12.90%	71.88%	3.13%	18.75%
	NL	48.39%	25.81%	25.81%	65.63%	3.13%	25.00%
TH	ENG	89.29%	7.14%	3.57%	68.97%	6.90%	20.69%
	NL	67.86%	14.29%	17.86%	65.52%	3.45%	27.59%

Most students also report issues with the Dutch subtitles because of the presence of another language they can understand, i.e., the English audio in both lecture styles and, additionally, the on-screen English text in the PowerPoint lecture. The test language is also mentioned as a reason why the Dutch subtitles do not necessarily help. About 40% of students mentioned the use of Dutch subtitles with these features to not always be beneficial or even confusing at times.

> "I sometimes looked at the Dutch subtitles to remember what was being said but then when I got the test in English, I had the Dutch terms in my head. So I had to think back to what they said exactly." (P8 – TH)

> "The first lecture with Dutch subtitles was confusing because what was being said was in English and the PowerPoint slides were in English, so with the Dutch subtitles it was just confusing." (P77 – PPT)

To conclude the interview, we asked students whether they would like to have subtitles for recorded lectures in the future and what language those subtitles should be in. Approximately half of the students (52.46%) reported they would like to have subtitles in future recorded or online lectures. A smaller portion (14.75%) said they would like to have subtitles, but would not want to use them all the time. A quarter (27.87%) said they would not like to have subtitles in lectures and the remaining students said they did not really have a preference.

When asked what language they would like to have the subtitles in, almost all students said they wanted to have the subtitles in the language of instruction (LOI), provided this was a language they could understand adequately, e.g., English (Table 4). Had it been in a language in which they were less proficient, such as German or French, they would have been more inclined to have wanted the subtitles in their native language, Dutch. Some would still like to use the language of instruction, mainly to allow easier note-taking and learn the language via content-integrated language learning. A few mentioned the examination language as the determining factor. Had the lecture been taught in a language the students were not familiar with, they would have preferred Dutch. One in five students mentioned English as their preferred language because they claimed to be more used to English subtitles and foreign media.

Table 4: Preferred language of subtitles in education

	Proficient	Less-proficient	Unfamiliar
LOI	96.77%	28.13%	0%
NL	3.23%	59.38%	64.52%
ENG	0%	0%	19.35%
NL or ENG	0%	6.25%	16.13%
EXAM	0%	6.25%	0%

These results are a clear endorsement of subtitles in education. However, it should be noted that students generally do not watch recorded lectures the way they did in this experiment. We asked students how they watched lectures at home and discovered a wide range of different study approaches. Some reported to put their phone away, sit at their desk and take a significant number of notes while watching a recorded lecture, while others reported they watched them in bed while being on their phone or even just used the recordings similar to podcasts in the background:

> "I sit at my desk and take notes of everything the professor is saying. I also go back to make sure I understood everything correctly. I do not have my phone with me so I do not use it." (P6 – TH)

> "Sometimes I just put it on in the background and then I clean my room or fold my clothes while I am just listening to it." (P85 – TH)

When asked whether they would still look at the subtitles at home, 48.33% of the students said they still would. One in three students said they would look at the subtitles less, though most of these students reported they would still look at the subtitles frequently. The large majority of the students did mention that, as exemplified in the previous paragraph, they would look away from the screen more because they are taking notes, averting their gaze to improve listening or are just occupied otherwise, e.g., using their phone, walking around, cooking or cleaning their room.

Furthermore, 25.93% of the students reported they always sped up recorded lectures and 22.22% said they sped it up occasionally depending on the speech rate of the professor, difficulty of the lecture or proximity of the examination period. The selected speed was often between 1.5 to 2 times the normal speed, but went up to 3, 4 or even 5 times the speed for a handful of students. This has significant implications on the processing of subtitles, as Liao et al. (2021) have shown that the presentation rate of subtitles has a considerable impact on the reading and processing of subtitles.

5 Discussion

In line with previous research (Dommett et al., 2022; Hayati & Mohmedi, 2011; Montero Perez et al., 2013; Taylor, 2008) and surveys (Burnham et al., 2010; Linder, 2016; Youngs, 2021; Zajechowski, 2022), this paper shows most students would appreciate subtitles in online education. The reasons why they would like to have subtitles and what benefits they perceive depend on the language of the subtitles and the style of the lecture they are used in. Most commonly, subtitles serve as an alternative to a slide presentation in lectures lacking slides. In such case, subtitles are perceived to help with understanding the lecture, retaining the content and following the stream of information. When a PowerPoint presentation is present, the need for subtitles decreases, though students still believe that they make frequent use of them and enjoy the bimodal input. These reasons lie close to the main reason to have subtitles reported by participants in Montero Perez et al. (2013), namely meaning construction.

In terms of language, the general preference seems to go to the language of instruction provided the students are sufficiently proficient in that language. Similar to what was found by Hayati and Mohmedi (2011), interlingual, native language subtitles are less appreciated and sometimes found distracting. This was especially the case when another language is shown on screen, e.g., a PowerPoint presentation in the instructional language.

While the reception and perception of subtitles is very positive in this experiment, it should be noted that the experimental setting is different from the home environment in which these lectures would normally be watched in. In the context of home, students walk around more, pay less attention and also take notes. Consequently, they spend considerably less time viewing the screen and possibly reading the subtitles. Furthermore, if students speed up subtitled lectures, like they often do for recorded lectures, the benefits of subtitles may go to waste as high subtitle presentation rates have been shown to negatively impact subtitle processing (Liao et al., 2021).

This discrepancy between an experimental lab setting and real-life is also highlighted in educational research focusing on a classroom environment (cf. Jarodzka et al., 2021). This should therefore always be considered during the analysis of results from future research.

6 Conclusion

This paper examined the perception of students regarding the use of subtitles in online education. Specifically, Dutch-speaking students watched three recorded EMI lectures with intralingual English subtitles, interlingual Dutch subtitles or no subtitles in either a voice-over PowerPoint style or a talking head style. The results revealed that the majority of students (1) would like to have subtitles in online education in the future; (2) often make use of subtitles when they are present for a variety of reasons; (3) prefer the language of instruction for subtitles in education, provided they can sufficiently understand this language. We therefore advocate the introduction of the option for subtitles in all asynchronous, online education, similar to Linder (2016). Not only would this increase overall accessibility and inclusivity, this paper shows the average hearing student also appreciates and perceives benefits from subtitles in education.

The results, however, should be interpreted carefully. As most students in this paper stated, the process of viewing a lecture at home may be different from the viewing during the experiment. Students tend to look at the lecture less because they are taking notes, walking around, on their phone or even cleaning their room or cooking. At home, the subtitles might therefore be used much less than they have been during the experiment. Consequently, the perceived benefits of the presence of subtitles might also be considered less prevalent than they were in this experiment. Another possibility would be that the reasons are simply different, e.g., using the subtitles to watch lectures quietly or as a way to overcome surrounding noise, similar to what was found in the Preply survey (Zajechowski, 2022). Not to mention the fact that a small majority of the students report that they occasionally speed up the lectures and would thus also speed up the presentation rate of subtitles, which may have a detrimental impact on the processing thereof (Liao et al., 2021). Lastly, the students in this experiment were all language students. As such, they are likely to be more interested and well-versed in language and subtitles than the average student.

We identify two limitations to the present study. Firstly, the recordings used in this experiment were approximately seven minutes long. While similar lengths can be observed in most AVT research (Ghia, 2012; Hefer, 2013b; Liao et al., 2021), a recorded lecture of extended length may yield different results. Secondly, as was shown in the results of this paper and similar to other educational research (cf. Jarodzka et al., 2021), the experimental setting likely impacts the use and thus also the perception of subtitles.

Consequently, this paper highlights the need for further research. Firstly, it would be interesting to explore the behavior of students watching subtitled recorded lectures in a home environment and reassess their preferences and perceptions of the benefits of subtitles. Secondly, in the present study, the participants were mostly native speakers of Dutch. They reported the use of Dutch subtitles in an EMI lecture to reduce the benefits of subtitles, at times even having detrimental effects on their processing. Future research should be conducted to assess the perceptions of native speakers on native language subtitles in an L1 lecture. Beyond these two avenues for future research, we encourage more research into the perception of subtitles and the preferences of the audience. As shown in pervious research and this paper, valuable insights can be gained by examining the matter.

7 References

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*(4), 509–533. doi:10.1017/S0142716402004022

Black, S. (2022). Could integrated subtitles benefit young viewers? Children's reception of standard and integrated subtitles: A mixed methods approach using eye tracking. *Perspectives, 30*(3), 503–519. doi: 10.1080/0907676X.2020.1849324

Burnham, D., Jones, C., Leigh, G., Noble, W., Brown, H., Varley, A., . . . Reid, A. (2010). Who uses television captions, when, and why? Analyses based on the Australian Television Caption Users Survey. *Asia Pacific Journal of Speech, Language and Hearing, 13*(2), 135–144. doi:10.1179/136132810805335083

d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist, 12*(3), 196–205. doi:10.1027/1016-9040.12.3.196

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta, 49*(1), 67–77. doi:10.7202/009021ar

Deckert, M., & Bogucki, L. (2019). Polish audience preferences regarding audiovisual translation: A reception study. In M. Deckert (Ed.), *Audiovisual translation: Research and use - Second Expanded Edition* (pp. 67–80). Berlin, Germany: Peter Lang.

Di Giovanni, E. (2016). Reception studies in audiovisual translation research: The case of subtitling at film festivals. *Trans-kom, 9*(1), 58–78.

Di Giovanni, E., & Gambier, Y. (2018). *Reception studies and audiovisual translation*. Amsterdam, Netherlands: John Benjamins.

Díaz Cintas, J. (2020). Audiovisual translation. In E. Angelone, M. Ehrensberger-Dow, & G. Massey (Eds.), *The Bloomsbury Companion to Language Industry Studies* (pp. 209–230). London, England: Bloomsbury Publishing Plc.

Dommett, E. J., Dinu, L. M., Van Tilburg, W., Keightley, S., & Gardner, B. (2022). Effects of captions, transcripts and reminders on learning and perceptions of lecture capture. *International Journal of Educational Technology in Higher Education, 19*.

Gambier, Y. (2018). Translation studies, audiovisual translation and reception. In E. Di Giovanni & Y. Gambier (Eds.), *Reception studies and audiovisual translation* (pp. 43–66). Amsterdam, Netherlands: John Benjamins.

Gerber-Morón, O., Szarkowska, A., & Woll, B. (2018). The impact of text segmentation on subtitle reading. *Journal of Eye Movement Research, 11*(4). doi:10.16910/jemr.11.4.2

Ghia, E. (2012). The impact of translation strategies on subtitle reading. In E. Perego (Ed.), *Eye tracking in audiovisual translation* (pp. 157–182). Rome, Italy: Aracne.

Grondin, S. (2016). *Psychology of perception*. Bern, Switzerland: Springer.

Hayati, A., & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology, 42*(1), 181–192. doi:10.1111/j.1467-8535.2009.01004.x

Hefer, E. (2013a). Reading first and second language subtitles: Sesotho viewers reading in Sesotho and English. *Southern African Linguistics and Applied Language Studies, 31*(3), 359–373. doi:10.2989/16073614.2013.837610

Hefer, E. (2013b). Reading second language subtitles: A case study of Afrikaans viewers reading in Afrikaans and English. *Perspectives, 21*(1), 22–41. doi:10.1080/0907676x.2012.722652

Hosogoshi, K. (2016). Effects of captions and subtitles on the listening process: Insights from EFL learners' listening strategies. *JALT CALL Journal, 12*(3), 153–178.

Jarodzka, H., Skuballa, I., & Gruber, H. (2021). Eye tracking in educational practice: Investigating visual perception underlying teaching and learning in the classroom. *Educational Psychology Review, 33*, 1–10. doi:10.1007/s10648-020-09565-7

Kruger, J.-L. (2016). Psycholinguists and audiovisual translation. *Target, 28*(2), 276–287. doi:10.1075/target.28.2.08kru

Kruger, J.-L., & Doherty, S. (2016). Measuring cognitive load in the presence of educational video: Towards a multimodal methodology. *Australasian Journal of Educational Technology, 32*(6), 19–31. doi:10.14742/ajet.3084

Kruger, J.-L., Doherty, S., & Soto-Sanfiel, M. T. (2017). Original language subtitles: Their effects on the native and foreign viewer. *Comunicar, 25*(50), 23–32. doi:10.3916/c50-2017-02

Kruger, J.-L., Hefer, E., & Matthew, G. (2013). *Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts.* Paper presented at the 2013 Conference on Eye Tracking South Africa, Cape Town, South Africa.

Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research, 7*(5), 1–15. doi:10.16910/jemr.7.5.4

Kruger, J.-L., & Mazur, I. (2012). Pear stories and audio description: Language, perception and cognition across cultures. *Perspectives: Studies in Translatology*, *20*(1), 1–3. doi:10.1080/0907676X.2012.633769

Kruger, J.-L., Wisniewska, N., & Liao, S. (2022). Why subtitle speed matters: Evidence from word skipping and rereading. *Applied Psycholinguistics, 43,* 211–236.

Liao, S., & Kruger, J.-L. (2023). Cognitive processing of subtitles: Charting the future by mapping the past. In A. Ferreira & J. W. Schwieter (Eds.), *The Routledge handbook of translation, interpreting and bilingualism* (pp. 161–176). London, England: Routledge.

Liao, S., Yu, L., Kruger, J.-L., & Reichle, E. D. (2022). The impact of audio on the reading of intralingual versus interlingual subtitles: Evidence from eye movements. *Applied Psycholinguistics*, *43*(1), 237–269. doi:10.1017/S0142716421000527

Liao, S., Yu, L., Reichle, E. D., & Kruger, J.-L. (2021). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading, 25*(5), 417–435.

Linder, K. (2016). Student uses and perceptions of closed captions and transcripts. Retrieved from https://content-calpoly -edu.s3.amazonaws.com/ctlt/1/images/3PM%20Student%20Survey%20 report_final%2010-25-16_Final_Remediated.pdf

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, 1–25. doi:10.1017/s0272263119000706

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology, 18*(1), 118–141.

Montero Perez, M., Peters, E., & Desmet, P. (2013). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL, 26*(1), 21–43. doi:10.1017/s0958344013000256

Orrego-Carmona, D. (2014). Where is the audience? Testing the audience reception of non-professional subtitling. In E. Torres-Simon & D. Orrego-Carmona (Eds.), *Translation Research Projects 5* (pp. 77–92). Tarragona, Spain: Intercultural Studies Group.

Orrego-Carmona, D. (2018). Audiovisual translation and audience reception. In L. Pérez-González (Ed.), *The Routledge Handbook of Audiovisual Translation* (pp. 367–382). London, England: Routledge.

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, *13*(3), 243–272. doi:10.1080/15213269.2010.502873

Perego, E., Laskowska, M., Matamala, A., Remael, A., Robert, I. S., Szarkowska, A., . . . Bottiroli, S. (2016). Is subtitling equally effective everywhere? A first cross-national study on the reception of interlingually subtitled messages. *Across Languages and Cultures, 17*(2), 205–229. doi:10.1556/084.2016.17.2.4

Rajendran, D. J., Duchowski, A. T., Orero, P., Martínez, J., & Romero-Fresco, P. (2013). Effects of text chunking on subtitling: A quantitative and qualitative examination. *Perspectives, 21*(1), 5–21. doi:10.1080/0907676X.2012.722651

Romero-Fresco, P. (2012). Quality in live subtitling: The reception of respoken subtitles in the UK. In A. Remael, P. Orero, & M. Carroll (Eds.), *Audiovisual translation and media accessibility at the crossroads* (pp. 109–131).

Szarkowska, A., & Bogucka, L. (2019). Six-second rule revisited: An eye-tracking study on the impact of speech rate and language proficiency on subtitle reading. *Translation, Cognition & Behavior*, *2*(1), 101–124. doi:10.1075/tcb.00022.sza

Szarkowska, A., & Boczkowska, J. (2022). Colour coding subtitles in multilingual films–a reception study. Perspectives, 30(3), 520–536. doi: 10.1080/0907676X.2020.1853186

Szarkowska, A., Díaz Cintas, J., & Gerber-Morón, O. (2021). Quality is in the eye of the stakeholders: What do professional subtitlers and viewers think about subtitling? *Universal Access in the Information Society, 20*(4), 661–675. doi:10.1007/s10209-020-00739-2

Szarkowska, A., & Laskowska, M. (2015). Poland - a voice-over country no more? A report on an online survey on subtitling preferences among Polish hearing and hearing-impaired viewers. In L. Bogucki & M. Deckert (Eds.), *Accessing audiovisual translation* (pp. 179–198). Bern, Switzerland: Peter Lang.

Szarkowska, A., Silva, B., & Orrego-Carmona, D. (2021). Effects of subtitle speed on proportional reading time: Re-analysing subtitle reading data with mixed effects models. *Translation, Cognition & Behavior, 4*(2), 305–330.

Taylor, G. (2008). Perceived processing strategies of students watching captioned video. *Foreign Language Annals, 38*(3), 422–427. doi:10.1111/j.1944-9720.2005.tb02228.x

Van Hoecke, S. M., & Kruger, J.-L. (forthcoming). *Reading in asynchronous, online lectures: A study into eye movements, comprehension and cognitive load*.

Van Hoecke, S. M., Schrijver, I., & Robert, I. S. (2022a). Methodological preparation of a within-subject audiovisual cognition, reception and perception study. *Journal of Audiovisual Translation, 5*(1), 94–128.

Van Hoecke, S. M., Schrijver, I., & Robert, I. S. (2022b). Preparing and comparing subtitles for quasi-experimental and experimental research in audiovisual translation studies. *Translation Spaces, 11*(1), 113–133

Van Hoecke, S. M., Schrijver, I., Robert, I. S., & Kruger, J.-L. (forthcoming). *The effects of subtitle presence and language on processing, cognitive load and comprehension in online EMI lectures: An eye-tracking study.*

Vulchanova, M., Aurstad, L. M. G., Kvitnes, I. E. N., & Eshuis, H. (2015). As naturalistic as it gets: Subtitles in the English classroom in Norway. *Frontiers in Psychology, 5*(1510). doi:10.3389/fpsyg.2014.01510

Youngs, I. (2021). Young viewers prefer TV subtitles, research suggests. *BBC*. Retrieved from https://www.bbc.com/news/entertainment-arts-59259964

Zajechowski, M. (2022). Why America is obsessed with subtitles. Retrieved from https://preply.com/en/blog/americas-subtitles-use/

Chapter 6 – Reading and processing of different subtitled lecture styles

To further our understanding of how lecture styles and subtitles interact and impact students' comprehension, perceived cognitive load and visual attention distribution, another study was conducted at Macquarie University using the same lectures, each time with English subtitles, in three different styles, namely a talking head, a voiceover PowerPoint and a composite, picture-in-picture lecture. Using an SR EyeLink 1000+ eye tracker, the study could examine word-level data and provide insight into reading and processing of subtitles and other on-screen content during the lecture. Additionally, the audience used in this study would be L1 English speakers and could therefore potentially respond entirely differently to the subtitles and the lectures. This chapter includes a single paper that discusses the results of this study. In doing so, this chapter adds to the answers formulated in Chapter 5 and deepens our understanding of the processing of subtitled lectures, answering another key research question of this thesis formulated in Section 1 of Chapter 3. As subtitle conditions were the same in all three lectures, this chapter predominantly focuses on lecture styles and their impact on the student.

Paper 5 – Reading in asynchronous, online lectures: A study into eye movements, comprehension and cognitive load

This paper was submitted to the Journal of Educational Psychology (2023):

Van Hoecke, S. M., & Kruger, J.-L. (under review). Reading in asynchronous, online lectures: A study into eye movements, comprehension and cognitive load. *The Journal of Educational Psychology*.

Formatting has been slightly altered to fit the present thesis.

Abstract

In the past decades, the availability of online education has grown exponentially. Despite this growth, there are still no guidelines that discuss how to design effective online lectures. While research lists numerous instructional principles that relate to the cognitive processing involved in learning and that can be considered while producing online education, these instructional principles can be ambiguous and contradictory at times. Additionally, with the increased importance of accessibility, subtitles may soon become a common aspect in lectures as well. It is therefore important to test various types of lectures with subtitles to examine how these instructional principles interact and how each contribute towards the experience of the lecture by the learner. Eye tracking can be a crucial tool to measure the effect of each aspect of lectures on cognitive processing and learning. The present paper uses eye tracking to examine the reception of three different styles of subtitled lectures, namely a talking head lecture, a voice-over PowerPoint lecture and a composite, picture-in-picture lecture. Specifically, 60 undergraduate Linguistic students were recruited to examine: (1) the viewing and reading patterns and how these are affected by each lecture style; and (2) the impact of lecture styles on comprehension/recall and perceived cognitive load. The results show that eye movements only differed in the picture-in-picture lecture. Furthermore, even though this is a clear example of split attention, no significant impact was found on perceived cognitive load or comprehension. The paper therefore raises questions about the redundancy principle, calling for an overhaul of its specifics.

Key words

Audiovisual translation (AVT), higher education, eye movement, reading, learning, online learning, cognitive load

1 Introduction

Understanding the cognitive processes involved in learning is key to improving educational design and learning outcomes. With the rise of the internet and multimedia a new branch of educational research has emerged, namely the study of e-learning. Many advances have been made regarding the concept and complexities of multimedia learning and how to design effective e-learning courses (e.g., Anmarkrud et al., 2019; Horton, 2011; Lopes & Soares, 2022; Mayer, 2014a). To our knowledge, there are, however, still no guidelines on how to present, or what to consider when producing, video lectures (Ilioudi et al., 2013) and research on advantages and disadvantages of different styles of video lectures also seems to be lacking (Chen & Wu, 2015).

There are a large number of multimedia lecture styles. Some of the more common ones are talking head lectures, voice-over PowerPoint lectures, picture-in-picture lectures (a combination of the previous two) or Khan-style lectures (showing a screencast of a digital whiteboard controlled by the lecturer or showing a lecturer drawing). Although there have been a number of studies on how the human cognitive architecture deals with multimodal educational material, very few deal with dynamic contexts such as subtitled video. The current paper wishes to address this gap and examines instructional effects and eye movements in different styles of subtitled lectures.

1.1 Instructional principles for multimedia instruction

The cognitive load theory (Sweller, 1988; Sweller et al., 2011; Sweller et al., 1998, 2019) and the cognitive theory of multimedia learning (Mayer, 2014a) provide an account of how the human mind is assumed to work during instruction and list a number of instructional principles that lecturers can consider while designing their courses. While these instructional principles are generally thoroughly researched, some ambiguities seem to remain in the theory. Firstly, one of the most common types of *online* lectures are talking head lectures. The image principle (Mayer, 2014b) states that learners do not necessarily benefit from a speaker's image being present on screen. The embodiment principle, on the other hand, assumes that people learn more deeply when the on-screen agent makes eye contact, has facial expressions and employs humanlike gestures and movement (Fiorella et al., 2019; Kizilcec et al., 2015; Stull et al., 2021; F. Wang et al., 2018). The apparent contradiction between these two principles makes it difficult to determine whether to include a talking head or not. It stands to reason that this will also be impacted by the nature and amount of information that accompanies the talking head (e.g., slides, subtitles).

The tension between the image and the embodiment principles is evident in the literature. Regarding processing load (cf. Mayer, 2014a) or cognitive load (cf. Sweller et al., 2019), one study found the image of the instructor to lead to increased cognitive load (Homer et al., 2008), while another found a decrease in cognitive load when the video of the instructor was present (Chen & Wu, 2015). Performance measures also reveal this tension. Two studies (Kizilcec et al., 2014; Lyons et al., 2012) found a positive reaction and heightened feeling of learning among students but no actual effect on test performance when the instructor was present in online lectures. Korving et al. (2016) confirmed such a positive reaction but also found an increase in recall in the presence of a talking head format under certain conditions.

Another ambiguity is revealed when a second common style of lectures, namely a voice-over PowerPoint lecture, is examined closely. The slides present information that is likely also being delivered in speech by the lecturer during the course. If this is the case, this information can be considered redundant. According to the redundancy principle, redundant material unnecessarily increases cognitive load and can thus negatively impact learning (Jamet & Le Bohec, 2007; Kalyuga & Sweller, 2014; Mayer et al., 2001). Furthermore, this redundancy in PowerPoint lectures is not perfectly synchronized with the lecturer's speech since slides appear and then remain on screen for a period of time. This means that learners are exposed preemptively to information that has not been addressed yet or remain exposed to information that has already been addressed earlier in the lecture. According to the temporal contiguity principle, this lack of synchronization is detrimental to learning (Mayer & Fiorella, 2014).

In contrast, there are two principles that assume the opposite would happen. Firstly, because the slides present the information visually whereas the narration is auditory, the information is presented in two modes and this is assumed to improve learning following the modality principle or dual-coding theory (Harskamp et al., 2007; Leahy & Sweller, 2011; Low & Sweller, 2014; Paivio, 1986). Secondly, slides present fragments of information that are deemed relevant for the lecture. The signaling principle claims learners can benefit from instruction when relevant words, sentences, concepts, etc. are highlighted (cf. Holsanova, 2014; van Gog, 2014).

The complexity of integrating different components in multimodal online learning is therefore emphasized in research on a number of principles of multimedia learning. In seemingly simple lectures, such as a talking head or voice-over PowerPoint lecture, multiple instructional principles appear to conflict with one another. This complexity grows even further when more dynamic lecture styles, such as a picture-in-picture lecture that combines a talking head with other information such as a PowerPoint presentation, or lectures containing animations or video, are considered. For these lectures, the sheer number of sources of information on screen can also impact the processing of the lecture. This leads to split attention (Ayres & Sweller, 2014; Mayer et al., 2001), which has a detrimental effect on cognitive processing. However, sometimes multiple sources of information do not cause split attention as, for example, the study by Jarodzka et al. (2015) has shown that students may not always be eager to process all sources of information, which means selective attention overrides split attention in such cases.

These (and other) instructional principles do not offer a uniform answer as to how a lecture should be designed. Consequently, it is important that research is conducted that concretely compares multiple lecture designs and illustrates the interactions between different principles. In particular, there is a need for more studies on the design of video lectures combined with different formats of media. The present paper will therefore focus on the processing of three common video lecture styles, namely talking head, voice-over PowerPoint, and a combination of these in picture-in-picture. Eye movements of participants will be analyzed to get a better understanding of reading patterns and attention distribution, while questionnaires will be used to measure self-reported cognitive load. Finally, comprehension tests are employed to measure the impact of lecture format on learning, which remains the most important aspect in instructional design.

1.2 Eye movements in online education

Eye tracking is a key tool in educational research (Jarodzka et al., 2017; Holmqvist et al., 2011). Eye movements provide a reliable way to determine where viewers attend (and, therefore, how they process different formats). This is because, in most cases, the mind attends where the eye is fixated. This well-known concept is called the eye-mind hypothesis (Just & Carpenter, 1980). While this is not completely true as parafoveal information can also be processed in the mind and the mind can continue attending to information after a fixation on that source of information has been terminated, attend to the location of the next saccade target or, indeed, wander (Anderson et al., 2004; Reichle et al., 2009), the gaze is still an important indicator of cognitive processing. However, cognitive processing capacity is limited. Learners have to constantly select what to pay attention to (by fixating the eyes) and what to process. More information makes the decision on what to attend to more onerous, which could deplete cognitive resources.

A talking head can be expected to be a clear focal point for the eyes. In normal human interaction, the speaker's mouth and eyes attract the gaze of the conversational partner for a variety of reasons (Clark, 1996). As shown in Louwerse et al. (2009), the on-screen agent elicits eye movement behavior that is comparable to a real social environment. Viewers look at characters on screen in the same way they would look at human conversational partners. This also means that they receive a considerable amount of attention, e.g., 26% of dwell time in a picture-in-picture animated lecture (Wang & Antonenko, 2017), 35-37% of dwell time in an easy or difficult PowerPoint lecture with an integrated video of the instructor (Wang et al., 2020) or 41% dwell time in a picture-in-picture PowerPoint lecture (Kizilcec et al., 2014). The amount of time spent viewing the instructor depends on the presentation of the instructor, the difficulty of the lecture and the presence of other concurrent information (both on screen and off screen). Although the amount of attention received by the instructor therefore varies, the presence of the instructor will have an impact on cognitive processing and viewing patterns in online lectures.

In addition to the talking head of the lecturer, many educational videos provide another source of supporting visual information such as PowerPoint slides or a video of a whiteboard. As mentioned before, relatively few studies have explored the integration of information on PowerPoint slides combined with voice-over or an inset of the lecturer in online education and even fewer have investigated the reading of PowerPoint slides in these contexts.

A possible third source of information that is relatively uncommon in present day education but may soon be introduced due to commercial, societal and, in Europe, even legal measures (i.e., the EU Accessibility Act – Directive 2019/882 – and the Audiovisual Media Service Directive – Directive 2010/13/EU) is subtitles. Subtitles offer hard-of-hearing and deaf audiences and viewers who are less proficient in the language of instruction, improved access to the source material. Benefits of subtitles for foreign language learning have also been revealed (e.g., Bairstow & Lavaur, 2017; Birulés-Muntané & Soto-Faraco, 2016; Gernsbacher, 2015; Montero Perez, 2020). Lecture design and subtitles, however, are rarely both considered in a single study.

Subtitles in online lectures compete with other information on screen for the attentional resources of the viewer. Subtitles are highly dynamic, automatically drawing the attention of the viewer to their on-screen location (d'Ydewalle & De Bruycker, 2007). Eventually, this leads to a considerable amount of viewing time being spent on the subtitles. How much time is spent on the subtitles depends on a number of factors, such as language of the subtitles and the source text, the speed of the subtitles or content on-screen. Chan (2020), for example, finds viewers to spend about 24% of the viewing time on L1, interlingual subtitles in a Khan-style lecture, whereas those same viewers spend around 45% of the total viewing time on L2, intralingual subtitles in a similar lecture.

With the limited processing capacity of humans in mind, and the additional resources required to read subtitles, the idea of subtitles in educational video for accessibility and better understanding appears to be self-contradictory. The effect subtitles have on comprehension and cognitive load are still under discussion. Some studies are in support of the redundancy principle, i.e., similar information presented twice (subtitles and audio) unnecessarily increases cognitive processing load and/or decreases comprehension (e.g., Kalyuga et al., 2000, 2004; J. Lee et al., 2012). Others find no effect of concurrent on-screen text on cognitive load (e.g., Craig et al., 2002; Kruger et al., 2013, 2014) and at times even improved comprehension (Van Hoecke et al., forthcoming), supporting the modality principle instead.

Kruger and Steyn (2014) highlight the importance of quantifying subtitle reading if we want to make any claims about the impact of subtitles. They quantify reading using the Reading Index for Dynamic Texts, a composite measure of unique fixations, saccades and average word length when no word-level eye movement data is available. With more accurate eye trackers, however, the possibility emerges to investigate the process of subtitle reading in a multimodal environment at word level. So far, the process of reading subtitles in a multimodal environment has received relatively little attention. One study (Liao et al., 2021) that goes into detail examines reading of subtitles of different speed and in the presence or absence of video. They found global changes in eye movement patterns when task demand changed as a result of higher subtitle speed. As subtitle speed increased, viewers started spending more time on the subtitles to try and read the subtitles before they disappeared. This was evidenced by a decrease in crossover saccades and dwell time on the video content. They also conclude that depending on the demand, eye-movement control shifts from local control to more global control. The idea is that in a non-demanding condition, reading happens under local control. The reader decides when and where to fixate next based on the word currently fixated. However, they found that word-length effects, word-frequency effects and wrap-up effects were modulated by the presence of video and the speed of subtitles. This indicated that as the task became more demanding, viewers started to pace their reading, e.g., skipping shorter words and skipping more words at the ends of sentences, with less regard for potential reading problems. Readers essentially adapt and start skimming rather than reading thoroughly.

To visualize the reading and processing of subtitles, Liao and colleagues (Liao et al., 2022; Liao et al., 2021) developed the Multimodal Integrated Language Framework. It illustrates how reading of subtitles can occur at a limited expense of viewing video content. After an on-screen feature and its location has been identified, viewers are capable of holding this information in their working memory, allowing them to track this feature in their peripheral vision. While the object is being tracked, viewers have time to read the subtitles. The identification of the object can also aid in the processing of the subtitles as certain words become more predictable, e.g., if the screen shows a polar bear, the viewer can expect a reference to a polar bear in the subtitles (for a more detailed description of the framework, we refer to Liao et al., 2022 and Liao et al., 2021). In online education, however, the concurrent video-content frequently consists of texts, static images or the image of an instructor and not a dynamic video of the topic. The question that arises then is how this content specifically affects viewing, reading and processing.

2 Current study

As mentioned before, the current study examines different styles of subtitles lectures and how each aspect of the lecture impacts the perceived cognitive load, comprehension, viewing/reading and cognitive processing of the lecture by the student. Ethics approval was granted by the Macquarie University Human Research Ethics Committee.

The goal of this paper is twofold: The first goal is to examine the impact of educational design on the reading of subtitles and viewing/reading of other material on screen. Specifically, the study compares three frequently used lecture styles (Figure 1), namely a talking head lecture, a voice-over PowerPoint lecture and a composite lecture showing the PowerPoint and including the talking head video in the bottom-right corner of the screen (also called a picture-in-picture PowerPoint lecture). By looking at eye movements, such as saccades and fixations, the study investigates global reading behavior. Additionally, this paper explores local reading behavior as evidenced by word-length and word-frequency effects. In doing so, the paper wishes to shed further light on the effects of concurrent video-content on eye-movement control in reading. Global measures are also used to provide insight into the reading of PowerPoint slides.

A second goal of the paper is to examine the effects of these lecture styles on comprehension and perceived cognitive load. As shown in the literature reviewed above, a number of instructional effects can play a role in the matter. Specifically, this study wishes to examine the image principle, the redundancy principle, the modality principle, the split attention principle and the signaling principle. By measuring comprehension, perceived cognitive load and eye movements in each lecture, the study provides insights into how these principles interact or possibly override each other.

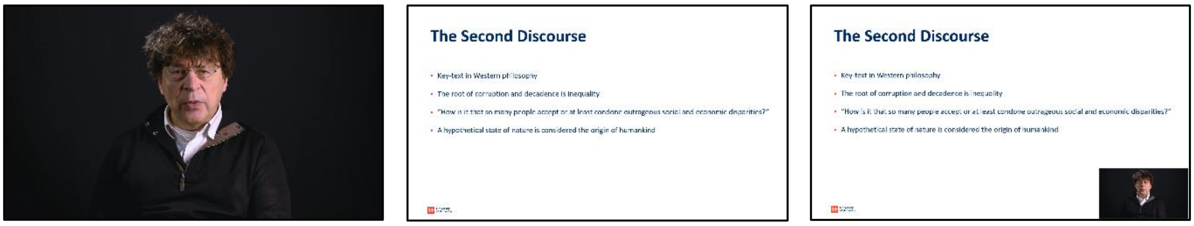


Figure 1: The three lecture styles: (1) talking head; (2) voice-over PowerPoint; (3) composite

We formulate the following two research questions:

1) *What is the impact of the presence of a PowerPoint presentation on students?*
2) *What is the effect of having both a talking head and a PowerPoint presentation as a picture-in-picture lecture on students?*

With regard to these two questions, we formulate two related hypotheses:

Firstly, we expect concurrent on-screen text (in the form of a PowerPoint presentation) to have a significant impact on subtitle reading following the split attention principle. When no text is present, viewers have more time to focus on the subtitles and can process them more extensively and effectively. We therefore expect talking head lectures to exhibit more thorough subtitle reading, i.e., longer fixations, shorter saccades, fewer words or subtitles skipped. In lectures that contain other text, subtitle reading is assumed to be sped up. This can be because the other text demands more attention than a talking head but also because subtitles may become more predictable. Given the idea that an identified object on screen can speed up the processing of subtitles due to words becoming more predictable (Liao et al., 2021), it would be logical to expect that if a certain word on a slide was already read by the viewer moments before said word appears in the subtitles, the processing of that word in the subtitles speeds up significantly. Despite this sped-up reading and split attention, previous research (Van Hoecke et al., forthcoming) leads us to expect that the subtitled PowerPoint lecture will still lead to lower perceived cognitive load and improved comprehension compared to the subtitled talking head lecture. We expect subtitles, in line with previous research (Kruger et al., 2014; Kruger & Steyn, 2014; Vulchanova et al., 2015), to have no effect or a slightly positive effect on learning and cognitive load, following the modality principle and opposing the redundancy principle. When a PowerPoint slide is added, we assume the benefit of signaling gained from the presence of the slide overrides the additional redundancy generated by having visual-verbal information on the slides and in the subtitles (the same mode) that is at times identical. Previous research also highlighted the importance of the signaling principle and confirmed that slide presentations could be beneficial as long as they do not include extensive extraneous material (Bartsch & Cobern, 2003; Kizilcec et al., 2014; Nouri & Shahid, 2005). Van der Zee et al. (2017) also showed that visual-textual complexity of a subtitled lecture is important for knowledge transfer. The level of visual-textual complexity and extent of extraneous material in our PowerPoint presentations, however, is beyond the scope of the present research. This should be kept in mind when interpreting the results. In sum, our first hypothesis states that lectures that include a

PowerPoint presentation will lead to sped-up subtitle reading, but also to lower perceived cognitive loads and improved comprehension when compared to an admittedly idle subtitled talking head lecture.

Secondly, we expect that a composite, picture-in-picture lecture offers too many sources of information, which will lead to split attention (Ayres & Sweller, 2014). This in turn will have a detrimental impact on learning and lead to faster and less thorough reading of both subtitles and content on slides. In terms of subtitle reading, we therefore expect fewer and shorter fixations, longer saccades and less time spent on the subtitles than in a talking head lecture or a voice-over PowerPoint lecture. Because the talking head in a picture-in-picture lecture is usually reduced to a portion in the corner of the screen, we also expect that any benefit that may be gained from the image of the instructor will be lost. These expectations follow the image principle. This is because, as stated before, a talking head presumably only improves learning when appropriate gestures or facial expressions, among others, are included as per the embodiment principle (Fiorella et al., 2019). Because of the smaller size of the talking head in this lecture style, these features are likely too small to track in the peripheral vision while reading the subtitles. The benefit would therefore be lost, unless viewers actively and frequently attend to the talking head at the expense of other information. In conclusion, our second hypothesis states that the composite lecture will lead to even faster reading and more split attention and will not gain any benefit from the presence of the instructor following the image principle.

Beyond these hypotheses, one key consideration should be made. Previous research has shown that language proficiency and native language play a key role regarding the effects of subtitles and education. Higher proficiency in the source language improves knowledge transfer (van der Zee et al., 2017; Van Hoecke et al., forthcoming). In such case, (native-language) subtitles may lead to higher perceived cognitive loads. It is therefore possible that as proficiency increases, subtitles gradually start becoming redundant, reducing their benefits.

3 Method

3.1 Participants and Design

The study used a three-factorial within-subjects design, comparing a talking head lecture, a voice-over PowerPoint presentation lecture and a composite lecture, i.e., picture-in-picture, showing both a talking head and a PowerPoint presentation. The participants were 60 native English-speaking undergraduate students from Linguistics at Macquarie University. Eye movement data from 14 participants were excluded because their data showed significant drift. Data from one more participant were excluded as the participant closed their eyes during the viewing, leading to a limited amount of data collected for those trials. Seven participants were male. The mean age of the participants was 22 years (SD = 6.49). Participants received either course credit or a $20 gift voucher for their participation.

Participants were first asked to complete a short biographical survey, after which they were shown a recorded lecture in one of the aforementioned styles. They were not allowed to take notes. After the lecture, they were given an 8-item psychometric questionnaire on cognitive load (adopted from Leppink and van den Heuvel, 2015) and a 12-item comprehension test, to measure perceived cognitive load and comprehension, respectively. This process was repeated two more times with new lectures in another style. Each time, the lectures and styles were counterbalanced via a Latin-square design to minimize order, fatigue or carry-over effects.

3.2 Stimuli

Three recorded lectures on philosophy were used as stimuli. All lectures were taught in English and included English subtitles. The subtitles were presented on a black background below the lecture so as not to contaminate the image of the lecture. The lectures were presented in three styles: as a talking head (showing only the face of the instructor on a neutral background), as voice-over PowerPoint presentation (showing only a PowerPoint), or in a composite/picture-in-picture style (showing the PowerPoint presentation full-screen and the talking head of the instructor in a small box in the bottom-right corner of the screen). Comparability between recordings and between subtitle tracks were achieved following the ten-step approach as recommended by Van Hoecke et al. (2022a, 2022b).

3.3 Apparatus

Eye movements were recorded using an SR EyeLink 1000+ eye tracker set to a sampling rate of 2000Hz. The lectures were presented on a Full HD 1920x1080 monitor. They were presented at the center and top of the screen with a resolution of 1280x720 and subtitles were presented below the video in a 1280x360 black box. Participants sat approximately 90cm from the screen and the chair was adjusted to match the screen height and an appropriate visual angle for the eye tracker. A chin-and-forehead rest was also used to minimize head movements. Only one eye was tracked per participant.

3.4 Analyses

Data were analyzed using Generalized/Linear Mixed Models with the lme4 package (v1.1-29) in R (v4.2.0). The G/LMMs were selected based on the process recommended by Zuur et al. (2009), starting with a full, overfitted model, trimming the random effects structure based on the restricted maximum likelihood Akaike information criterion (using ANOVAs) and, lastly, sorting out the fixed effects structure. A treatment contrast was used to compare different levels. The composite lecture style was used as the intercept. The emmeans package (v1.7.4-1) was used to compare all factors after a final model was fit. Only significant results are reported for brevity.

4 Results

4.1 Reading of subtitles

Table 1 and Figure 2a show that the average fixation duration on the subtitles is shorter in the composite style. Indeed, the results of a linear mixed effect model show significantly longer average fixations for the PowerPoint mode ($t = 14.76$, $p < 0.001$) and for the talking head mode ($t = 15.04$, $p < 0.001$) compared to the composite mode. The shorter average fixation duration in the composite lecture likely relates to the amount of information available in that lecture. To attend to all information, less time is available to read the subtitles, meaning the subtitles are being read faster, resulting in shorter fixations. There is no significant difference in average fixation duration between the PowerPoint and talking head lecture.

A similar pattern can be observed when looking at the average fixation count per subtitle shown in Table 1 and Figure 2b. The composite lecture received the lowest number of fixations per subtitle, followed by significantly higher number of fixations per subtitle in the PowerPoint lecture ($t = 7.7$, $p < 0.001$) and in the talking head format ($t = 11.96$, $p < 0.001$). Though the number of fixations per subtitle increases going to the talking head lecture, the difference between the PowerPoint and talking head lecture is not significant.

As both the number of fixations per subtitle and the average duration of these fixations are lower in the composite lecture, it can be expected that the average dwell time spent on each subtitle (as shown in Table 1 and Figure 2c) is lower in the composite lecture compared to the other two lectures. This is the case when the composite lecture is compared to the PowerPoint lecture (t = 15.48, p < 0.001) or to the talking head lecture (t = 18.66, p < 0.001).

Table 1 and Figure 2d also show that the average forward saccade length in the area of the subtitles is significantly shorter in the PowerPoint lecture (t = -5.73, p < 0.001) and in the talking head lecture (t = - 14.09, p < 0.001) when compared to the composite lecture. An emmeans analysis revealed that the saccades in the PowerPoint lecture were significantly longer than in the talking head lecture (z = 8.31, p < 0.001) These longer saccade lengths likely relate to slightly faster, and possibly less thorough, subtitle reading behavior in the composite lectures and possibly even in the PowerPoint lectures.

As shown in Table 1 and Figure 2e, there were also significantly more crossovers between the subtitles and the other content on screen in the composite lecture when compared to the PowerPoint lecture (t = -5.406, p < 0.01) or the talking head lecture (t = 3.732, p < 0.01). Although they also reveal slightly fewer crossovers in the PowerPoint lecture compared to the talking head, this difference is not significant.

Table 1: Means and standard deviations for global subtitle reading measures

	Mean Fixation duration (ms)	**Sub fixation count**	**Sub dwell time (ms)**
Composite	241.9(171)	6.8(4.57)	1872(1110)
PowerPoint	262.1(185)	7.4(4.80)	2191(1213)
Talking head	265.5(188)	7.8(4.59)	2247(1209)

	Mean forward saccade length (°)	**Total sub crossovers**
Composite	3.24(1.20)	154.6(42.66)
PowerPoint	3.13(1.18)	123.0(38.79)
Talking head	2.95(1.10)	132.8(56.36)

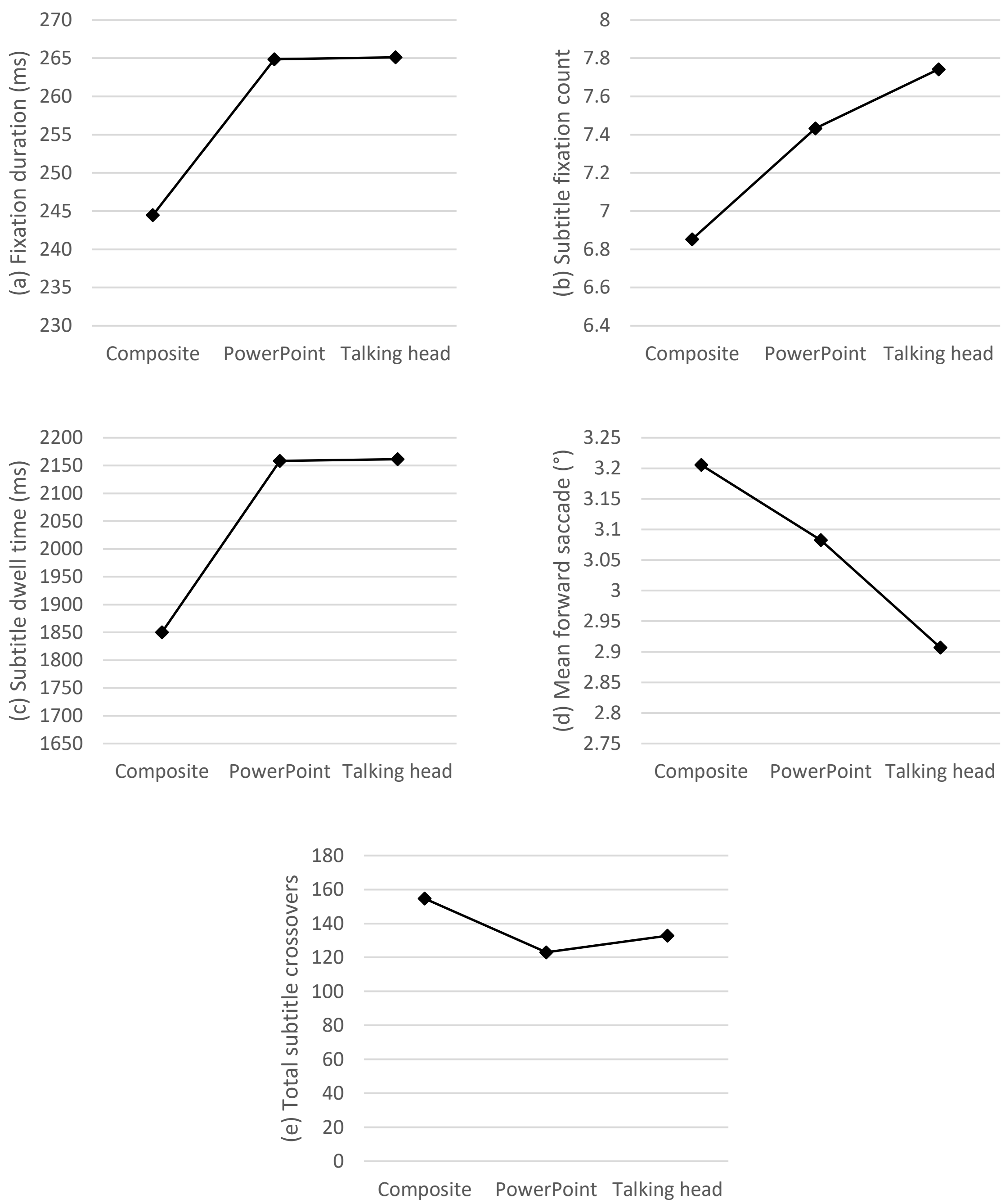


Figure 2: LMMs predicted lecture effects for global subtitle reading measures

Results so far point towards the subtitles being read faster and possibly less thoroughly in the composite lecture compared to the other two lectures. This leads to two more questions: 1) Do they skip more words in the subtitles?; and 2) Are the subtitles still read to completion?

Table 2 and Figure 3a show that significantly fewer words in the subtitles are skipped in the PowerPoint lecture (t = -6.54, $p < 0.001$) and in the talking head lecture (t = -4.32, $p < 0.001$) compared to the composite lecture. The considerable standard deviation for the talking head lecture does indicate a certain amount of personal variation when it comes to reading the subtitles in a talking head lecture. Fewer words were skipped in the PowerPoint lecture when compared to the talking head lecture, but more subtitles were skipped entirely in the PowerPoint lecture.

Table 2 and Figure 3b also show that in each format subtitles are still being read to (near) completion. The subtitle reading percentage shows the (word) location of the fixation closest to the end of the subtitle as a percentage of all words in the subtitles. For example, if the last word was fixated, it would be 100%. If the seventh word in a subtitle with ten words was focused and the last three words were skipped, it would be 70%. Subtitles with a single fixation that was not on the first word were excluded. The LMM predicts no significant differences between the lectures, meaning in all lectures, subtitles were generally read to completion.

In conclusion, subtitles in the composite lecture are being read faster, at times more akin to skimming rather than thorough reading, than in the other two lectures. Despite the difference in content shown on screen, subtitle reading behavior in the PowerPoint and the talking head lecture is mostly similar. Two significant differences found were shorter saccades in the talking head lecture and fewer words skipped in the PowerPoint lecture.

Table 2: Mean and standard deviation for complete subtitle reading measures

	Subtitle words skipped percentage	Subtitle reading percentage
Composite	50.00(8..73)	88.80(16.96)
PowerPoint	47.66(9.87)	89.26(16.88)
Talking head	48.44(16.29)	87.61(18.89)

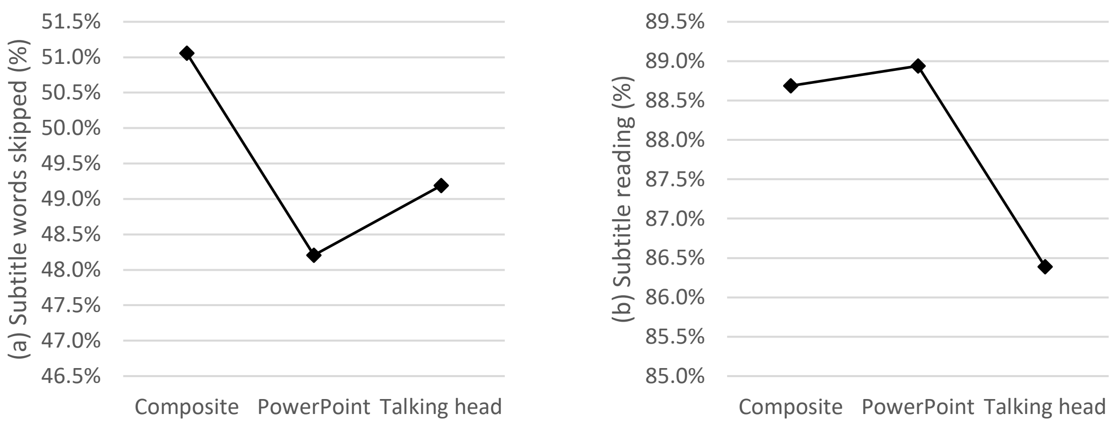


Figure 3: LMMs predicted lecture effects for complete subtitle reading measures

We were also interested in more local measures of reading to shed some light on the active processing load while reading/watching these lectures. We were mainly interested in word length and word frequency effects that could potentially be attenuated by processing load. A third measure that we also included was the time to first fixation to examine possible delays in viewing the subtitles because of the additional lexical processing in the PowerPoint and composite lecture that included text on the slides in addition to the subtitles.

The word frequency effect states that lower frequency words elicit longer fixations. Liao et al. (2021) showed that this frequency effect is attenuated when less time is available to read and process the content, but that it is still present even at high subtitle speeds. Our previous analyses showed faster subtitle reading (fewer, shorter fixations and longer saccades) in the composite lecture, so we would expect a less prevalent word frequency effect in this lecture compared to the other lectures. Figure 4b, however, reveals that this is not the case. While the word frequency effect is still present (t = -26.095, $p < 0.001$), the decline in dwell time per word (log-transformed) is similar for all three lectures. A significant interaction between word frequency and the talking head lecture reveals the word frequency effect is more prevalent in the talking head lecture compared to the other two (t = -3.273, $p < 0.01$).

A second measure of more local eye movement control is the word length effect. Assuming that reading pace is decided based on active lexical processing and information gathered in the parafovea, word length (similar to word frequency) impacts whether a word will be fixated and how long, (e.g., shorter words are more likely to be skipped, cf. Reichle et al., 2003). Liao et al. (2021) also found an attenuated effect of word length on dwell time per word in more demanding conditions. Similarly to the word frequency analysis, our results (Figure 4c) find a significant effect of word length on dwell time ($t = 27.786$, $p < 0.001$). Additionally, in support of Liao et al. (2021), a significant interaction is found between word length and lecture type, with a stronger word length effect in the talking head lecture ($t = 4.388$, $p < 0.001$).

Looking at Table 3 and Figure 4a, similar times to first fixation can be seen for the composite lecture and the PowerPoint lecture. However, when they are compared to the talking head lecture, a significant drop in time can be observed (PowerPoint: $z = -3.176$, $p < 0.01$; Composite: $z = -4.024$, $p < 0.001$). Moving from the image of the instructor to the subtitles is done faster than moving from on-screen text. This likely relates to the additional ongoing lexical processing delaying the saccade, whereas the lexical processing in the talking head is limited to audio only. In the PowerPoint lecture, students are also found to move their eyes back away from the subtitles later than in the composite lecture ($z = -5.830$, $p < 0.001$) or the talking head lecture ($z = -2.761$, $p < 0.05$). This explains why even when viewers move their eyes to the subtitles later in the PowerPoint lecture than in the talking head lecture, the overall reading pattern of the subtitles in these two lecture styles was still found to be similar.

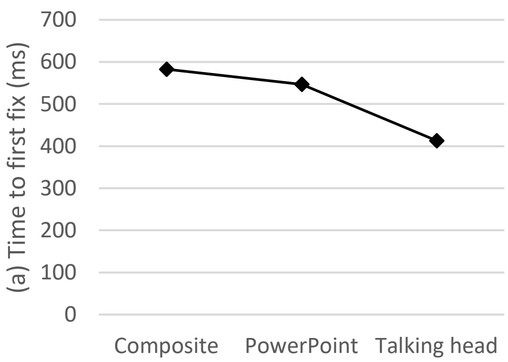


Table 3: Means and standard deviations for time to first fixation

	Time to first fixation
Composite	569.7(762)
PowerPoint	529.7(748)
Talking Head	366.3(519)

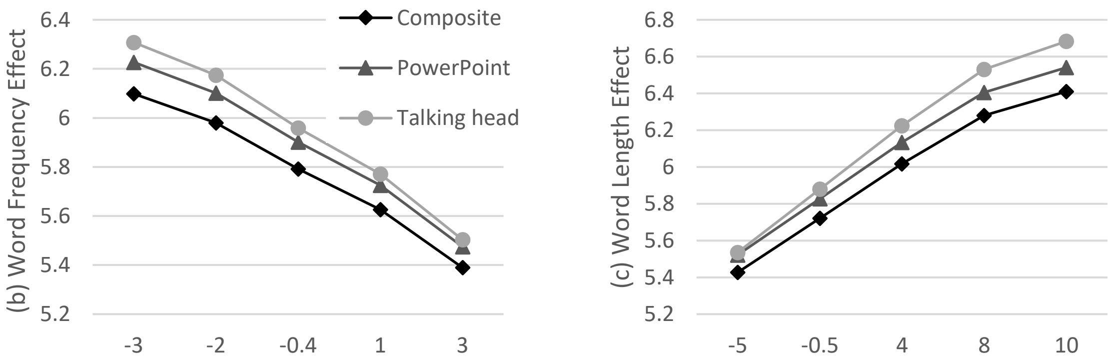


Figure 4: Predicted lecture effects for local subtitle reading measures

4.2 Reading of slide content

The following four analyses focus on the eye movement data on the slide content. Images were excluded, meaning only data from text on the slides were considered. This also means that the following analyses only contain data from the composite and the PowerPoint lecture.

Table 4 and Figure 5a show the mean fixation duration on slide content. A marginally shorter mean fixation duration can be observed in the composite lecture. This difference is significant ($t = 3.686$, $p < 0.001$). This would imply faster reading in the composite lecture alongside the already established sped-up reading of the subtitles in this lecture, due to the increase in the number of visual elements to process.

A similar pattern can be observed when looking at the average number of fixations per slide in Table 4 and Figure 5b. A significant difference is found with fewer fixations per slide in the composite lecture compared to the PowerPoint lecture ($t = 2.333$, $p < 0.05$).

As expected the average dwell time per slide further corroborates this finding. Table 4 and Figure 5c show a lower dwell time per slide for the composite lecture compared to the PowerPoint lecture. Again, this difference is significant ($t = 3.06$, $p < 0.01$).

The results so far indicate faster reading of slide content in the composite lecture. This is a somewhat logical finding as participants who desire to attend to all three sources of information present in the composite lecture, will have to split their attention and have less time to read the slide content. However, one could also expect that once the slide content has been read, participants would skip the slide content and only focus on new information, i.e., the talking head and the subtitles. Figure 5d shows the total number of crossovers between the subtitles, slide content and talking head per quarter of slide duration. Interestingly, the total number of crossovers between the slide and other content does not differ between the two lecture styles. This means that although the composite lecture offers one extra source of information, the number of crossovers between the sources of information is similar to that in a PowerPoint lecture. Figure 5d also shows how the number of crossovers decreases significantly as the time a slide is shown on screen progresses. Comparing Q1 to Q4 reveals a significant drop ($t = 16.991$, $p < 0.001$).

Table 4: Means and standard deviations for global slide reading measures

	Mean fixation duration (ms)	**Slide fixation count**	**Slide dwell time (ms)**
Composite	230(119.17)	73.7(38.47)	17393.63(9678)
PowerPoint	240(124.75)	79.4(41.90)	19775.77(11262)

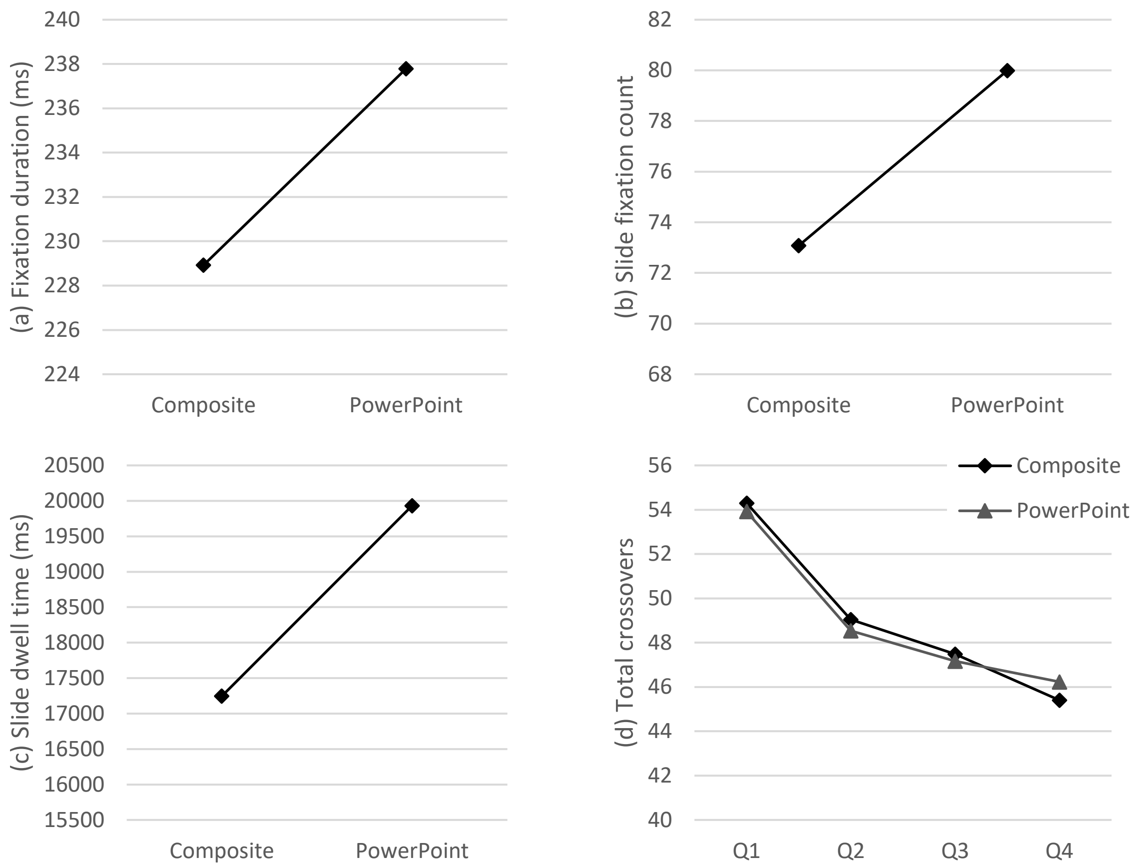


Figure 5: LMMs predicted lecture effects for global slide reading measures

4.3 Distribution of visual attention

To get a sense of the overall visual attention distribution, dwell time was considered once more. However, the following two analyses consider dwell time as a percentage of the visible time of the interest area, i.e., percentage of time spent on a subtitle based on subtitle duration or percentage of time spent on a slide based on slide duration. Finally, a short overview of dwell time distribution for the individual components of each lecture is also included.

Table 5 and Figure 6 show that participants spent about 55% of the time a subtitle is visible watching that subtitle. The time spent is about the same in the PowerPoint lecture and the talking head lecture. However, comparing the composite lecture to the PowerPoint lecture reveals a significant increase in time spent on the subtitle ($t = 6.708$, $p < 0.01$).

A similar observation can be made when looking at dwell time on slides (Table 6 and Figure 7) with an increase of time spent on the slides in the PowerPoint lecture when compared to the composite lecture ($t = 2.887$, $p < 0.01$). Evidently, this follows from all the data and analyses that have already been mentioned earlier in this paper and yet again reveals a picture of split attention in lectures with more simultaneous sources of information.

To get an overview of how viewers distribute their attentional resources based on eye movement data, Figure 8 shows the respective dwell times on each of the components in each of the lectures. Here we see considerably less time on the talking head in the composite lecture when compared to the talking head lecture, slightly less time spent on the slides and slightly less time spent on subtitles in the composite lecture. Time spent on the subtitles and image (slide or talking head) in the PowerPoint and talking head lecture is somewhat similar. This would imply reading slide content or viewing the image of an instructor leads to a comparable distribution of attentional resources.

Table 5: Means and deviations for subtitle dwell time (%)

	Dwell time percentage
Composite	48.97 (26.39)
PowerPoint	57.81(28.83)
Talking Head	59.15 (27.73)

Table 6: Means and deviations for slide dwell time (%)

	Dwell time percentage
Composite	25.16 (11.42)
PowerPoint	27.49 (12.95)

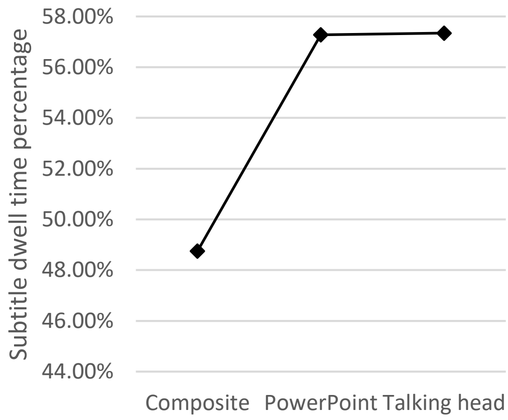


Figure 6: Predicted lecture effects for subtitle dwell time (%)

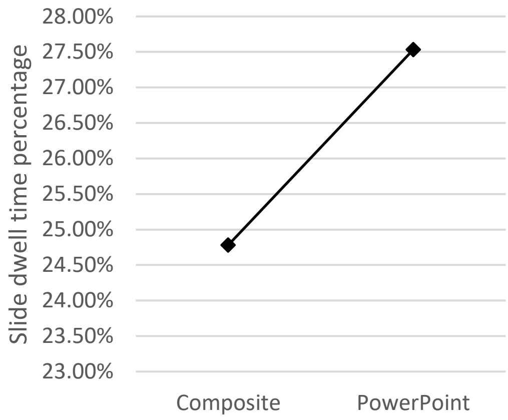


Figure 7: Predicted lecture effects for slide dwell time (%)

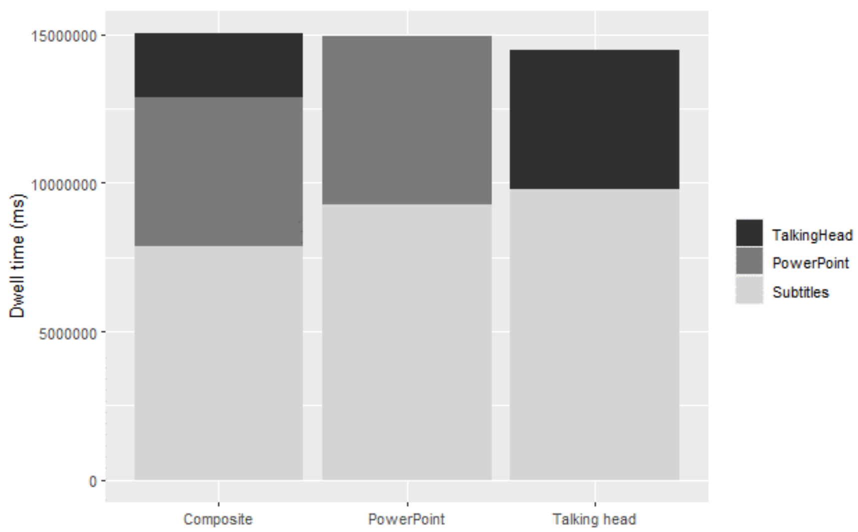


Figure 8: Visual attention distribution

4.4 Perceived cognitive load

The psychometric questionnaire used (Leppink & van den Heuvel, 2015) consists of eight questions. The first four questions measure, intrinsic cognitive load and the last four extraneous cognitive load.

When we consider only the intrinsic load questions (Table 7), we find no significant differences between the three lecture styles. The data do show a gradual increase in perceived intrinsic load ratings going from the composite style to the PowerPoint style with the highest ratings for the talking head style, though this increase is marginal and not significant.

Extraneous load is rated considerably lower than intrinsic load. When we compare extraneous load across the different lecture styles, no significant difference is found either. The data reveal a similar increase in perceived extraneous load for the talking head style. This, again, is not significant. The PowerPoint and composite style are very similar.

Table 7: Mean and deviations for perceived cognitive load

	Intrinsic	Extraneous
Composite	4.88(2.25)	2.19(2.20)
PowerPoint	5.00(2.20)	2.23(2.15)
Talking Head	5.18(2.16)	2.71(2.45)
	M(SD)	**M(SD)**

4.5 Comprehension

The 12-item comprehension test that participants were required to complete after watching each lecture yielded no significantly different scores between the different lecture styles (Table 8). The tests that followed the composite style were done marginally better than other tests, though the difference remains insignificant.

Table 8: Mean and deviation for comprehension

	Score on 12
Composite	6.67(1.80)
PowerPoint	6.43(1.93)
Talking Head	6.30(2.19)
	M(SD)

5 Discussion

The goal of this paper was twofold: (1) examining reading of subtitles and other on-screen textual content and overall viewing patterns in different styles of lectures and (2) investigating the effects of subtitled lecture styles on comprehension and perceived cognitive load.

5.1 The impact of the presence of a PowerPoint presentation

The first research question posed in this paper asked about the impact of the presence of a PowerPoint presentation on (1) reading and eye movements and (2) perceived cognitive load and comprehension. Our hypothesis stated that a lecture with a PowerPoint presentation would lead to sped-up subtitle reading, but also lower perceived cognitive loads and improved comprehension when compared to a talking head lecture. Our hypothesis was not supported:

(1) The presence of a PowerPoint presentation did not necessarily lead to sped-up subtitle reading. Subtitle reading patterns were largely similar between the voice-over PowerPoint lecture and the talking head lecture. The only significant differences were that in the PowerPoint lecture, fewer words were skipped in the subtitles and saccades were slightly longer. Another interesting difference we did find between subtitle reading in a talking head lecture and a PowerPoint lecture is the time to first fixation. The time to first fixation from when the subtitle appears was significantly later in the PowerPoint lecture. To compensate for this delay, viewers were also shown to exit the subtitles later than they would in a talking head lecture. This highlights the increased time required for lexical processing as opposed to visual processing before a new saccade can be executed. This similar reading pattern contradicts our expectations.

We can think of three possible reasons for this. Firstly, it is possible that for the audience in this study, the PowerPoint lecture did not include sufficient information to require faster reading or lead to split attention. A second possibility would be that the predictability of words gained from having these similar or identical content on the slide did not impact subtitle reading, contrary to our expectations that were based on the Multimodal Integrated Language Framework (Liao et al., 2022; Liao et al., 2021). This raises some questions about the interactions between the visual system and the declarative memory in the framework. Further research would be necessary to make definitive statements on the relation between on-screen information density, the effect of concurrent visual-verbal information and eye movements in multimodal contexts.

(2) We expected improved comprehension scores and lower perceived cognitive load ratings in the presence of a PowerPoint. We expected this as the signaling principle (van Gog, 2014) was previously found to override additional redundancy generated by the PowerPoint presentation in a subtitled lecture (Van Hoecke et al., forthcoming). In this study, this was not the case. Despite the signaling not being beneficial for knowledge transfer, the additional information on the PowerPoint presentation also did not increase cognitive load or decrease comprehension. This raises questions about the redundancy principle (Jamet & Le Bohec, 2007; Kalyuga & Sweller, 2014; Mayer et al., 2001) and possibly calls for a serious overhaul of the principle. This finding likely relates to the fact that the audience in the present paper are native speakers of English and thus have a decreased need for additional material, such as a PowerPoint, to understand the lecture. If this were the case, it is still interesting to see that they spend about 50% of the time the subtitles are visible reading the subtitles, despite having no real use for them, which could be further proof that subtitles are being read automatically and subconsciously (d'Ydewalle & De Bruycker, 2007). This idea can be strengthened as the present results could also be an indication that the current younger generation, a population that is arguably strongly accustomed to reading subtitles, can do so effectively even in highly demanding contexts. They adapt their reading and viewing patterns with no expense to comprehension or perceived cognitive load of the lecture. Following the study of van der Zee et al. (2017), we assume that with different degrees of slide density in PowerPoint lectures results might be different. Further research would be required to examine when a PowerPoint has too much content for subtitles to start having a detrimental effect on learning or cognitive load. Additionally, this finding also shows the talking head lecture does not outperform the other lectures. These results are similar to what was found in Kizilcec et al. (2014), but different from Homer et al. (2008), Y.-H. Lee et al. (2014) and J. Wang and Antonenko (2017), among others, who find either an increase or a decrease in comprehension, recall and/or cognitive load when a talking head was present. Admittedly, the talking head used in this study was relatively idle. A neutral background to simulate a lecturer sitting behind a computer at home and a lack of social cues, e.g., hand gestures, facial expressions or interaction, likely led to the talking head not influencing the results, in line with Stull et al. (2021), Fiorella et al. (2019) and Kizilcec et al. (2015). This also supports the image principle, stating learners do not necessarily benefit from the speaker's image, provided of course that the embodiment principle was not considered during the production of the lecture (Mayer, 2014b). In any case, this lack of differences between the talking head lecture and the PowerPoint lecture does raise some questions about the aforementioned principles and may possibly call for an overhaul or further specification of certain principles to some extent.

5.2 The effect of a composite lecture style

The second research question asked about the effect of having both a talking head and a PowerPoint presentation as a picture-in-picture lecture on (1) reading and eye movements and (2) perceived cognitive load and comprehension of students. We argued that due to the number of sources of information, the composite lecture would lead to split attention (Ayres & Sweller, 2014; Mayer et al., 2001), which would in turn lead to even faster reading of content (Liao et al., 2021). The addition of a talking head was expected not to yield any benefits. Our second hypothesis was supported.

(1) Our results clearly indicate that the composite lecture is the outlier regarding reading and viewing patterns. In terms of subtitle reading, the composite lecture showed longer forward saccades, more words were skipped, fixations were shorter and less frequent and dwell time on the subtitles decreased. Our results also showed that subtitles were still largely read to completion in the composite lecture, meaning that even with four sources of information, the students still had time to read the subtitles, albeit hurriedly. In line with Liao et al. (2021), a transition from global eye movement control to local eye movement control can also be observed as the word frequency and word length effects were shown to be less strong when compared to a talking head, though not when compared to a PowerPoint lecture. The lack of difference between the composite lecture and the PowerPoint lecture seems to imply concurrent text is a key factor in the matter.

(2) Despite the sped-up reading and viewing to be able to attend to all sources of information, perceived cognitive load or performance did not increase or decrease. It is interesting to see that even though the eye movements are a clear example of split attention, the lecture style had no detrimental effect on learning. Again, we argue that the lecture might not have been sufficiently difficult for the native-speaking audience and thus the style of the lecture had no strong impact on learning. Nevertheless, it is interesting to see that, once again, the introduction of the PowerPoint or inclusion of a talking head in a combined lecture had no impact. It maintains support for the image principle, though, once again, raises questions about the redundancy principle and the signaling principle.

6 Conclusion

This paper shows how students can effectively adapt their reading and viewing behavior in different styles of lectures. Depending on the other content available on screen, students consciously or unconsciously decide to speed up their reading to be able to attend to all information on screen. In this study, students were always seen to attend to all sources of information, never deciding to ignore one component to be able to focus better on another. This has important implications for educational design as it shows that every single detail added to a lecture will have an impact on eye movements.

The study also shows how even with adapted viewing and reading behavior, comprehension and perceived cognitive load remain constant among students. It highlights the complex web of instructional principles and their interactions. Specifically, this study wished to increase the understanding of the split attention, image, signaling, redundancy and modality principles.

The results tell us that, as mentioned before, each component in a lecture will affect eye movements and thus potentially lead to split attention. Split attention is assumed to have a detrimental effect on learning. However, in this study, this is not the case. One explanation would be that the lecture was not challenging enough for split attention to lead to decreased knowledge transfer. Another would be that other components (and related principles) contributed to the lecture negating the split attention effect.

Both the composite lecture, including the talking head, and the talking head lecture did not affect comprehension or perceived cognitive load. Consequently, this study supports the image principle stating the addition of the image of the lecturer does not necessarily improve learning. We acknowledge that the talking head used in this study was relatively idle and thus do not wish to make any statements about whether or when the embodiment principle could override the image principle.

Despite our previous support for the signaling principle, the present study did not reveal any benefits from the introduction of a PowerPoint presentation. We believe this relates to the aforementioned difficulty of the lecture itself. The current student population consisted of native English-speaking students, whereas in a previous study (Van Hoecke et al., forthcoming) the participants were L2 English speakers. For L1 English speakers, it is possible that the lecture was not sufficiently difficult, hence signaling was not required to improve learning or comprehension.

The last and arguably biggest question raised by this study concerns the modality and redundancy principles. Though this study can make few statements about the modality principle and its relation to subtitles in education, it showed that even when subtitles were present, the addition of more, possibly identical, verbal information on a PowerPoint slide did not affect learning. This calls for a significant overhaul of the redundancy principle and its specifics. It is clear that redundant information, even when presented in the same mode, does not necessarily increase processing/cognitive loads or decrease comprehension. Further research would be required to see when and which information can in fact be considered redundant and detrimental for learning. We also believe one possible explanation would be that the current younger generation is accustomed to high-information situations and the use of subtitles. Consequently, they have adopted strong strategies to cope with these situations at a limited expense to their processing.

While the study has made a contribution and advances our understanding of eye-movements and instructional principles in multimodal, subtitled educational environments, there are a number of limitations that should be addressed in future research. Firstly, this paper merely explores PowerPoint slide reading at a superficial level. Future research could further explore PowerPoint viewing and reading behavior with experiments and materials designed around that specific scientific goal. Secondly, the present study employs thoroughly prepared materials. Research, similar to the studies on instructional principles referred to in the literature review of this paper, on specific details of each individual lecture style and how it influences the global picture is necessary to provide a more complete view of effects of lecture styles on viewing patterns, comprehension and cognitive load. We see this research as one step towards understanding the reception of online lecture styles and eventual optimization of the design of these lectures.

7 References

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. *Psychological Science, 15*(4), 225–231. doi:10.1111/j.0956-7976.2004.00656.x

Anmarkrud, Ø., Andresen, A., & Bråten, I. (2019). Cognitive load and working memory in multimedia learning: Conceptual and measurement issues. *Educational Psychologist, 54*(2), 61–83. doi:10.1080/00461520.2018.1554484

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Vol. 2, pp. 206–226). Cambridge, England: Cambridge University Press.

Bairstow, D., & Lavaur, J.-M. (2017). Sous-titrage, compréhension de films et acquisition de vocabulaire. *Psychologie Française, 62*(3), 249–261. doi:10.1016/j.psfr.2015.05.001

Bartsch, R. A., & Cobern, K. M. (2003). Effectiveness of PowerPoint presentations in lectures. *Computers & Education, 41*, 77–86. doi:10.1016/S0360-1315(03)00027-7

Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PLoS One, 11*(6). doi:10.1371/journal.pone.0158409

Chan, W. S. (2020). *An investigation of subtitles as learning support in university education* [Unpublished doctoral dissertation]. Macquarie University, Sydney, Australia.

Chen, C.-M., & Wu, C.-H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education, 80*, 108–121.

Clark, H. H. (1996). *Using language*. Cambridge, England: Cambridge University Press

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94*(2), 428–434. doi:10.1037//0022-0663.94.2.428

d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist, 12*(3), 196–205. doi:10.1027/1016-9040.12.3.196

Fiorella, L., Stull, A. T., Kuhlmann, S., & Mayer, R. E. (2019). Instructor presence in video lectures: The role of dynamic drawings, eye contact, and instructor visibility. *Journal of Educational Psychology, 111*(7), 1162–1171. doi:10.1037/edu0000325

Gernsbacher, M. A. (2015). Video captions benefit everyone. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 195–202. doi:10.1177/2372732215602130

Harskamp, E. G., Mayer, R. E., & Suhre, C. (2007). Does the modality principle for multimedia learning apply to science classrooms? *Learning and Instruction, 17*, 465–477. doi:10.1016/j.learninstruc.2007.09.010

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, England: Oxford University Press.

Holsanova, J. (2014). Reception of multimodality: Applying eye tracking methodology in multimodal research. In C. Jewitt (Ed.), *Routledge Handbook of Multimodal Analysis* (Second edition ed., pp. 285–296). London, England: Routledge.

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior, 24*(3), 786–797.

Horton, W. (2011). *E-learning by design*. Hoboken, NJ: Wiley.

Ilioudi, C., Giannakos, M. N., & Chorianopoulos, K. (2013, April 8 – April 9). *Investigating differences among the commonly used video lecture styles.* Paper presented at the Workshop on Analytics on Video-based learning, Leuven, Belgium.

Jamet, E., & Le Bohec, O. (2007). The effect of redundant text in multimedia instruction. *Contemporary Educational Psychology, 32*, 588–598.

Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research, 10*(1), 1–18. doi:10.16910/jemr.10.1.3

Jarodzka, H., Janssen, N., Kirschner, P. A., & Erkens, G. (2015). Avoiding split attention in computer-based testing: Is neglecting additional information facilitative? *British Journal of Educational Technology, 46*(4), 803–817. doi:10.1111/bjet.12174

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*(1), 126–136. doi:10.1037//0022-0663.92.1.126

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors, 46*(3), 567–581. doi:10.1518/hfes.46.3.567.50405

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Vol. 2, pp. 247–262). Cambridge, England: Cambridge University Press.

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology, 107*(3), 724–739. doi:10.1037/edu0000013

Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). *Showing face in video instruction: Effects on information retention, visual attention, and affect.* Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Toronto, Canada.

Korving, H., Hernández, M., & de Groot, E. (2016). Look at me and pay attention! A study on the relation between visibility and attention in weblectures. *Computers & Education, 94,* 151–161.

Kruger, J.-L., Hefer, E., & Matthew, G. (2013). *Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts.* Paper presented at the 2013 Conference on Eye Tracking South Africa, Cape Town, South Africa.

Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research, 7*(5), 1–15. doi:10.16910/jemr.7.5.4

Kruger, J.-L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly, 49*(1), 105–120. doi:10.1002/rrq.59

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25*, 943–951. doi:10.1002/acp.1787

Lee, J., Lin, L., & Robertson, T. (2012). The impact of media multitasking on learning. *Learning, Media and Technology, 37*(1), 94–104. doi:10.1080/17439884.2010.537664

Lee, Y.-H., Hsiao, C., & Ho, C.-H. (2014). The effects of various multimedia instructional materials on students' learning responses and outcomes: A comparative experimental study. *Computers in Human Behavior, 40*, 119–132.

Leppink, J., & van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education, 4*(3), 119–127. doi:10.1007/s40037-015-0192-x

Liao, S., Yu, L., Kruger, J.-L., & Reichle, E. D. (2022). The impact of audio on the reading of intralingual versus interlingual subtitles: Evidence from eye movements. *Applied Psycholinguistics, 43*(1), 237–269. doi:10.1017/S0142716421000527

Liao, S., Yu, L., Reichle, E. D., & Kruger, J.-L. (2021). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading, 25*(5), 417–435.

Lopes, A. P., & Soares, F. (2022). *Online distance learning, course design and multimedia in E-learning*. Hershey, PA: IGI Global.

Louwerse, M. M., Graesser, A. C., McNamara, D. S., & Lu, S. (2009). Embodied conversational agents as conversational partners. *Applied Cognitive Psychology, 23,* 1244–1255.

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Vol. 2, pp. 227–246). Cambridge, England: Cambridge University Press.

Lyons, A., Reysen, S., & Pierce, L. (2012). Video lecture format, student technological efficacy, and social presence in online courses. *Computers in Human Behavior, 28,* 181–186.

Mayer, R. E. (2014a). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43–71). New York, NY: Cambridge University Press.

Mayer, R. E. (2014b). Principles based on social cues in multimedia learning: Personalization, Voice, Image, and Embodiment principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 345–368). New York, NY: Cambridge University Press.

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 279–315). New York, NY: Cambridge University Press.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187–198. doi:10.1037//0022-0663.93.1.18

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, 1–25. doi:10.1017/s0272263119000706

Nouri, H., & Shahid, A. (2005). The effect of PowerPoint presentations on student learning and attitudes. *Global Perspectives on Accounting Education, 2*, 53–73.

Paivio, A. (1986). *Mental Representations: A dual coding approach*. Oxford, England: Oxford University Press.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*, 445–526.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16*(1), 1–21.

Stull, A. T., Fiorella, L., & Mayer, R. E. (2021). The case of embodied instruction: The instructor as a source of attentional and social cues in video lectures. *Journal of Educational Psychology, 113*(7), 1441–1453. doi:10.1037/edu0000650

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.

Sweller, J., Kalyuga, S., & Ayres, P. (2011). *Cognitive load theory* (1 ed.). New York, NY: Springer.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. doi:1040-726X/98/0900-0251

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292. doi:10.1007/s10648-019-09465-5

van der Zee, T., Admiraal, W., Paas, F., Saab, N., & Giesbers, B. (2017). Effects of subtitles, complexity, and language proficiency on learning from education videos. *Journal of Media Psychology, 29*(1), 18–30.

van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 263–278). New York, NY: Cambridge University Press.

Van Hoecke, S. M., Schrijver, I., & Robert, I. S. (2022a). Methodological preparation of a within-subject audiovisual cognition, reception and perception study. *Journal of Audiovisual Translation, 5*(1), 94–128.

Van Hoecke, S. M., Schrijver, I., & Robert, I. S. (2022b). Preparing and comparing subtitles for quasi-experimental and experimental research in audiovisual translation studies. *Translation Spaces, 11*(1), 113–133

Van Hoecke, S. M., Schrijver, I., Robert, I. S., & Kruger, J.-L. (forthcoming). *The effects of subtitle presence and language on processing, cognitive load and comprehension in online EMI lectures: An eye-tracking study.*

Vulchanova, M., Aurstad, L. M. G., Kvitnes, I. E. N., & Eshuis, H. (2015). As naturalistic as it gets: Subtitles in the English classroom in Norway. *Frontiers in Psychology, 5*(1510). doi:10.3389/fpsyg.2014.01510

Wang, F., Li, W., Mayer, R. E., & Liu, H. (2018). Animated pedagogical agents as aids in multimedia learning: Effects on eye-fixations during learning and learning outcomes. *Journal of Educational Psychology, 110*(2), 250–268. doi:10.1037/edu0000221

Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior, 71*, 79–89.

Wang, J., Antonenko, P. D., & Dawson, K. (2020). Does visual attention to the instructor in online video affect learning and learner perceptions? An eye-tracking analysis. *Computers & Education, 146*. doi:10.1016/j.compedu.2019.103779

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York, NY: Springer.

Part IV: Discussion and conclusion

Chapter 7 – Conclusion

This concluding chapter of the manuscript aims to provide a concise overview and discussion of its results and the implications it has for practice. First, it answers the research questions asked in Chapter 3, tying the answers back to the findings and previous research. Subsequently, it briefly highlights some limitations present in the research conducted as part of this project. The chapter ends with a brief conclusion and a short discussion on important future research tracks that were highlighted during this research.

1 Discussion of findings

1.1 The effect of lecture styles and subtitles on perceived cognitive load

The first research question ("What is the effect of lecture styles and subtitles on perceived cognitive load?") was addressed in Chapter 5 and Chapter 6. In both studies, the same recorded English-taught lectures were used but with different experimental parameters. Chapter 5 had three subtitle conditions (English, Dutch or no subtitles) as a within-subject factor and two lecture styles (talking head and voiceover PowerPoint) as a between-group factor. Participants were students from Belgium, i.e., native speakers of Dutch/L2 English speakers. Chapter 6 had only one subtitle condition (English subtitles) and three lecture styles (talking head, voiceover PowerPoint and a composite style) as within-subject factor. Participants in Chapter 6 were L1 English students from Australia.

Both studies yielded a similar answer to this first research question, namely the presence of subtitles has no effect on perceived cognitive load as measured by the psychometric questionnaire from Leppink and van den Heuvel (2015). This means that these findings align with previous research from the field of AVT (Chan, 2020; Kruger et al., 2013, 2014; Kruger & Steyn, 2014; Liao et al., 2020; van der Zee et al., 2017). It contradicts research from the field of education (Craig et al., 2002; Diao et al., 2007; Mayer et al., 2001) and raises questions with regard to the redundancy principle (Kalyuga & Sweller, 2014; Mayer & Fiorella, 2014). The redundancy principle states that learning is impaired when the same or similar information is presented concurrently in multiple forms (Kalyuga & Sweller, 2014).. Instead of supporting the redundancy principle, our findings seem to support the modality principle (Low & Sweller, 2014), i.e., knowledge construction is more effective when information is presented both visually and auditory, and the dual-coding theory (Paivio, 1986). However, it should be noted that subtitles did not decrease cognitive load either.

Despite the subtitles not decreasing perceived cognitive load as measured by the questionnaire, they are very positively received and considered helpful, similar to what was found in Taylor (2008), Montero Perez et al. (2013) and Dommett et al. (2022). A significant portion of the students in Chapter 5 report the bimodal input (subtitles and audio) is helpful and the subtitles make it easier to follow the stream of information and focus on the lecture.

These results should, however, be interpreted cautiously. While there is no increase in cognitive load when subtitles are present, the proficiency of students in addition to the language of the subtitles, source and other material do play a vital role in the matter. In Chapter 6, it can be argued that the L1 English students did not need the English subtitles to have access to or understand the presented content to the same degree as the second language speakers reported in Chapter 5. Hence, they do not decrease cognitive load. In Chapter 5 on the other hand, subtitles could have served as an aid to understand the lecture. English subtitles can already help, but interlingual Dutch subtitles would be expected to help even more. This was also the case, as Dutch subtitles did decrease the perceived cognitive load but only for students that watched a talking head lecture. Returning to the modality principle, we argue that native language subtitles can be beneficial in foreign language lectures, as long as these lectures do not show visual-verbal information in the foreign language on screen alongside the subtitles. It is beneficial for learners to present information visually and auditorily, but if two languages are presented in one of these modes simultaneously, the benefit from using the two modes is largely lost. Two languages in a single mode possibly creates a additional cognitive load for that processing channel. We would therefore extend the modality principle, defining it as follows: Knowledge construction is improved when information is presented visually and auditory, provided the information presented in each mode is limited to a single language that the audience can understand. Our assumption is further supported by the interviews conducted in Chapter 5, during which multiple students mentioned the use of Dutch subtitles and English PowerPoint slides led to confusion.

As for the lecture styles, the impact they have on perceived cognitive load seems to depend on the language proficiency of the student audience. For the L1 English students in Chapter 6, lecture styles did not have an impact on perceived intrinsic or extraneous cognitive load. Lecture styles did, however, affect the perceived extraneous load of L2 English speakers in Chapter 5. A PowerPoint presentation was found to decrease the extraneous load experienced by students when compared to a talking head lecture. This is in line with the students' perceptions and preferences with regard to lecture styles as the large majority prefer to have PowerPoint slides present during a lecture, ideally in combination with the image of the instructor. We believe this finding shows the importance of the signaling principle, which states cognitive load is decreased when learners are directed to relevant information (van Gog, 2014). Additionally, we assume the talking head that was used in this study was very idle. As was shown in Fiorella et al. (2019), Kizilcec et al. (2015) and Stull et al. (2021), hand gestures or facial expressions, for example, play a significant role in the processing of talking head lectures. As for the L1 English students, we unexpectedly found no significant differences in perceived cognitive load between the lecture styles. While a tendency for higher extraneous load ratings for the talking head lecture could be observed when compared to the other two lecture styles, the lack of significance is striking. We argue that the experienced load for this audience was already low due to the language of the lecture being their first language, hence the difference between the lectures was too minute to notice any significant discrepancies.

1.2 The effect of lecture styles and subtitles on comprehension

The second research question ("What is the impact of different styles of subtitled lectures on comprehension?") was also addressed in the studies of both Chapter 5 and Chapter 6. As expected, the findings lie close to what was found with regard to cognitive load. Chapter 5 showed no effect of the presence or language of subtitles on immediate comprehension for L2 English speakers. As all conditions in Chapter 6 had English subtitles, nothing can be said about the impact of the presence of subtitles on comprehension for L1 English speakers. In contrast to Kruger and Steyn (2014), Chapter 5 finds no interaction between actual subtitle reading, as measured by the Reading Index for Dynamic Texts (Kruger & Steyn, 2014), and test performance. This would imply that regardless of the student's approach, subtitles do not aid immediate comprehension. The only measure that did positively correlate with test performance was English vocabulary size. This is in line with the studies by Vulchanova et al. (2015) and van der Zee et al. (2017), in which higher language proficiency was shown to lead to improved comprehension. For our study, we argue that all students in Chapter 5 had sufficient proficiency in English to comprehend the lecture. We expect there would be a positive effect of subtitles on comprehension for students with lower proficiencies, similar to Lavaur and Bairstow (2011) and Hayati and Mohmedi (2011). This remains to be explored further.

It is important to note that regardless of the results above, numerous students did think the subtitles were helpful. About 70% of the L2 English students for either lecture style indicated the English subtitles helped them understand the content of the lecture better. This went down to about 65% for the Dutch subtitles. Subtitles were considered especially helpful in the talking head lecture to better understand the content and speech and retain the content.

As part of the study in Chapter 5, we also had students complete a delayed comprehension test. In the delayed test, English subtitles were shown to significantly improve scores. We link this to the dual-coding assumption (Paivio, 1986) and the modality principle (Low & Sweller, 2014): A benefit stands to be gained by presenting the same information in both the visual and auditory channel. Similar to the results regarding cognitive load, the benefit is only present when the subtitles match the source language. However, for retention, another potential cause for this result can be the fact that the test was also in English. Moreover, delayed test performance positively correlates to vocabulary size scores, similarly to immediate test performance.

As for lecture styles, L2 English speakers performed better on the immediate and delayed test when they were shown a voiceover PowerPoint lecture. As was mentioned in the previous section, the majority of these students preferred to have a PowerPoint present over only seeing the instructor’s image. Despite L1 English speakers also preferring the composite lecture or at the very least a voiceover PowerPoint lecture, the presence of the slides did not impact their test scores. In addition to the importance of the design of a PowerPoint presentation for learning gains (Bartsch & Cobern, 2003; Kizilcec et al., 2014; Nouri & Shahid, 2005), our results underline the role of the relation between source language and audience language in the matter. Regardless of the result for native language speakers, we do highly recommend the use of PowerPoints presentations in online lectures as opposed to only a talking head, because of the benefit it has for L2 speakers and the preference and large support it receives in both student groups.

1.3 Attention allocation and reading in subtitled lectures

This section addresses research question three (“How is attention allocation affected by lecture styles and subtitles?”) and four (“How is reading affected by lecture styles and subtitles?”). Question three was addressed in both Chapter 5 and 6, whereas we only delved deeper into the process of reading in Chapter 6. Because they are closely related, we address both in this section.

Overall, we observed more than half of the total viewing time was spent watching the subtitles in all conditions and for all audiences. The language of the subtitles did not play a role. In all cases, a large portion of the viewing time (proportional reading time is not reported) is spent on the subtitles. This is in line with some of the proportional reading times per subtitles measured in previous research, e.g., about 35% in d'Ydewalle and De Bruycker (2007), 40% in Szarkowska and Gerber-Morón (2018) or 36% in Szarkowska and Bogucka (2019). Students were also aware of their viewing behavior as many indicated they read the subtitles often or even all the time. The reasons for reading the subtitles varied, for example, to be able to understand and/or retain the content better, focus on a dynamic part of the screen or because the bimodal input was appreciated. Additionally, in line with the assumption that subtitles are being read automatically (d'Ydewalle & De Bruycker, 2007), one in four students in Chapter 5 mentioned this automaticity as one of the reasons they spent so much time on the subtitles.

Personal preferences and language proficiency evidently play a role in the time spent viewing the subtitles. Listening comprehension scores were shown to lead to reduced reading of the subtitles in Chapter 5. As stated in Liao et al. (2020) and Szarkowska and Bogucka (2019), actual reading of the subtitles decreases as the need to read the subtitles decreases. Another important factor is the presence of other information on screen. If other content is present, the viewer can and likely will desire to attend to that information as well. In such case, split attention occurs, i.e., the viewer has to allocate limited cognitive resources and viewing time to each source of information and thus divide their attention (Ayres & Sweller, 2014). Liao et al. (2021) showed that in most media, viewers are able to identify an item and its corresponding location on screen and then attend to it in their peripheral vision. This allows for subtitles to be processed and read at a limited expense of being able to process the other content on screen.

In our case, however, the other content is additional instructional material that can contain important information in the form of gestures, facial expressions and gaze (talking head), or text and images (PowerPoint). We expected the talking head to have a limited impact on the reading of subtitles, whereas the addition of text on slides would likely have a significant impact as subtitles could no longer be processed while the visual-verbal information on the slides was being processed. Our findings are in line with these expectations. We see that in Chapter 5, the presence of presentation slides leads to more subtitles being skipped entirely, less time being spent on the subtitles and more time being spent on the concurrent on-screen content. When measured with the RIDT, however, we see that the reading of the subtitles that were not skipped stays the same across all subtitle conditions and lecture conditions. Only the Dutch subtitles are being read significantly less thoroughly in a PowerPoint lecture when compared to a talking head lecture. This is likely a very conscious choice of students to avoid mixing two languages in their visual processing channel. In Chapter 6, the results are rather striking. In line with our expectations, the composite lecture, i.e., the lecture containing the largest number of information sources, leads to less time being spent on the subtitles. Students attempt to attend to all sources of information and thus limit the attention devoted to the subtitles. We expected this to be similar in the PowerPoint lecture, albeit to a lesser extent than in the composite lecture. This was, however, not the case as viewing patterns were similar between the PowerPoint and talking head lectures. We believe this once again can be brought back to the language proficiency of these students. The subtitle presentation rate is slow enough for these students to attend to two sources of information, even when the second source is verbal information on a PowerPoint. It exceeds their cognitive capacities when the third source of information is introduced. Only then do they decide to divide their attention and limit the focus on each source of information.

The results regarding attention allocation highlight one concept in particular, namely metacognition (Mayer, 2011). Metacognition refers to learners being capable of consciously or subconsciously adjusting their learning approach to maximize gains and optimize processing load. When subtitles are present, the current generation of students, which are likely accustomed to subtitles, can distribute their attentional resources so that this additional source of information does not negatively impact learning. Depending on the lecture, personal preferences and language proficiency, they further adjust their viewing pattern and do so very effectively.

This adjusted viewing behavior becomes even more apparent when we analyze the viewing and reading in the subtitled lectures more closely in Chapter 6. The clear outlier is the composite lecture style. In the composite lecture style, subtitles are read noticeably faster and less thorough than in the other two lecture styles. The average fixation duration is shorter, there are fewer fixations, the average forward saccade length is longer and there are more crossovers between the subtitles and the other content on screen. Interestingly, the subtitles are still read to completion in this lecture style. They are just read faster, more words are skipped and overall less dwell time is spent on the subtitles. What we did not expect was that the subtitle reading patterns in the PowerPoint lecture were very similar to those in the talking head lecture. There were only a few differences.

Firstly, the average forward saccade length was slightly longer in the PowerPoint lecture than it was in the talking head lecture. We do not yet have an explanation for this difference. Further investigation would be necessary. Secondly, fewer words are skipped in the PowerPoint lecture, but there were more subtitles that were skipped entirely. This results in similar dwell times for both lectures, even with a difference in words skipped. This probably relates to the moments when a new slide appears on screen. Viewers will likely want to read the new content first and then return to reading the subtitles as nothing new is shown on screen. This pattern is confirmed when we look at slide fixations over the course of time, where we can see a significant drop in crossovers after the first quarter of the time a slide is shown on screen (a more in-depth discussion of slide reading can be found in later paragraphs of this section). A talking head, on the other hand, provides dynamic information. Viewers thus go back and forth between the image and the subtitles at a steady rate throughout the entire lecture. Thirdly, the time to first fixation when a subtitle appears was shorter in a talking head lecture than it was in a PowerPoint or composite lecture. The processing of moving images can be finished or aborted faster than the processing of verbal information. In the composite lecture, there is too much information on screen to dwell too long on the subtitles so despite the delayed fixation on the subtitles, the exit time out of the subtitles is similar to that in a talking head lecture. In the PowerPoint lecture, however, there is more time available and viewers compensate for the delayed fixation on the subtitles with a significantly later exit, meaning the total time spent on the subtitles is similar to that in a talking head lecture.

So far, we can conclude that the subtitle reading process in the composite lecture is distinctly faster and less thorough than it is in the other two lectures. While there are differences between the PowerPoint lecture and the talking head lecture, it seems the reading of the subtitles in both lectures is profound and only slightly different. However, we also looked at local measures of reading, namely the word frequency and word length effect, i.e., less frequent or longer words are generally fixated for longer periods of time. By examining these effects, we gain insight into whether reading is under local control, i.e., modulated by lexical processing, or global control, i.e., influenced by global features and a possible timer. Liao et al. (2021) found that when processing loads increase, these effects become less prevalent and the process of reading transitions from a word-by-word decision process to a more global decision process with paced reading. Our results confirm this assumption as the word frequency and the word length effect are present in all lectures, but stronger in the talking head lecture. This implies that in the composite lecture, but also in the PowerPoint lecture, viewers alter their verbal processing and reading.

To complete the picture, we examined the reading of the content on the slides as well. This confirms the results and conclusions above. In the composite lecture, there are fewer and shorter fixations on the slides and the total dwell time on the on-screen content is shorter than in the PowerPoint lecture. Again, viewers distribute their attentional resources and attempt to attend to all sources of information based on their own needs. Consequently, viewers read content on slides faster to allow more time for other sources of information.

1.4 Students' perceptions of subtitles in recorded lectures

The sections above have already drawn a picture of the students' perceptions regarding subtitles in recorded lectures and different styles of recorded lectures. It has already shown that the opinion of students on subtitles in lectures is very positive, which is a general answer to research question five ("What is the opinion of students on different styles of subtitled recorded lectures?). A more in-depth answer to the question is formulated below.

The perceptions of subtitles are very positive. Most students claim to read them often and believe subtitles help them understand the content and speech better, aids them in following and focusing on the lecture, allows them to process the content better because of the bimodal input and enables them to retain the content better. Some students report this reading is a very conscious choice, though a significant part (25%) mentions they read the subtitles largely automatically. Previous research has shown that subtitles benefit language learning and vocabulary acquisition (Bird & Williams, 2002; Danan, 2015; Montero Perez, 2020; Montero Perez et al., 2015). Most students also perceive this benefit, especially when the subtitles are the only visual-verbal information on screen, i.e., in a talking head lecture.

In terms of future use, 50% of the students reported they would like to have subtitles in live and/or recorded online lectures. Only 25% stated they would not want to have subtitles, mostly because the subtitles would distract them from learning or make the lecture too easy. The remaining part either had no opinion or would like to have subtitles, but only for certain lectures. In terms of language, nearly all students preferred to have the subtitles in the language of instruction provided this was a language they could understand. If the subtitles were in a language the students were less proficient in, the large majority would prefer their native language instead.

Yet, there are two important nuances to these results. First and foremost, students themselves mentioned their behavior when watching lectures at home was distinctly different from the behavior they had during the experiment. They would be taking notes, possibly using their phone as well, look around more or even walk away from the screen to clean their rooms, prepare a meal, etc. If this is the case, the time spent on the subtitles will be significantly less than it was in these studies. How and why the subtitles are used or read and what benefits stand to be gained from having them in online lectures might thus be different when examined in a home context. Secondly, research has shown that the presentation rate of subtitles has a significant impact in the processing and reading of the subtitles (Kruger et al., 2022; Liao et al., 2021). Half of the students reported they sometimes or even always sped up recorded lectures. If subtitles are added to these recorded lectures, the presentation rate of these subtitles will thus also be significantly faster than intended. Further research would be necessary to explore the processing of subtitles in sped-up lectures and investigate how this might impact cognitive load or comprehension.

1.5 The benefit of thorough experimental preparation

Chapter 4 in this manuscript discusses two papers that illustrate the preparation for the studies conducted as part of this project. In doing so, we wish to contribute to the relatively young field of AVT research and attempt to add a well-founded approach to experimental AVT reception studies. The papers themselves illustrate the process, whereas the advantage of using this ten-step approach becomes clearer in the studies that followed. This section briefly discusses this advantage.

Initially, our main concern was the comparability between the materials used in within-subject studies in this project. We acknowledge that we may have gone too far in the preparation of these materials, as differences between conditions can be captured as random effects in mixed modelling statistics, provided these differences are not substantial. However, our thorough preparation did allow us to leave the materials out of the random effects structure in our statistical models as these revealed to barely explain any variance in the results. Essentially, what this means is that because our materials were so comparable our models were less complex and easier to compute. Additionally, we asked participants whether they thought the lectures and tests used were similar throughout the experiment. The large majority reported they were.

Because of this thorough preparation and the confirmed comparability both statistically and by the participants themselves, we strongly believe in the validity of our studies and their results.

We still strongly believe this approach could help in strengthening the methodological foundations of the field of research and allow for more replicable and valid research.

Beyond the benefits of this preparatory process, our research has also highlighted a number of other important methodological considerations for future research. Firstly, this thesis shows the added value of mixed modelling in AVT research. The use of mixed models in AVT research could already be observed in the past years (e.g., Kruger et al., 2022; Liao et al., 2022; Liao et al., 2021; Szarkowska & Bogucka, 2019). We support these studies in their use of mixed modelling and recommend future research to employ similar statistical methods when possible. Secondly, as is also highlighted in Chapter 5, the perception of AVT is an aspect that should not be forgotten in AVT research. Even when it is not the main goal of the study, it can provide valuable insight into how and why AVT impacts a specific audience. Furthermore, it can reveal important considerations for future research, e.g., the fact that students regularly speed up lectures which likely impacts processing if subtitles were added to these lectures.

2 Limitations

Despite the contribution made in this research, there are still some limitations present throughout the project. Firstly, the duration of the stimuli was only seven minutes. While this is similar to other AVT research with eye tracking (e.g., 15 minutes in Perego et al., 2010; 10 minutes in Liao et al., 2021; 6 minutes in Ghia, 2012; 5 minutes in Hefer, 2013b; 4 minutes in Szarkowska et al., 2021), we do acknowledge that viewing behavior and perceptions may change for lectures that are of a length that is likely more representative of what is available at higher education institutions (one to two hours). Further research would be required to examine what the minimal and optimal duration of a clip would be for it to be ecologically valid and adequately represent reality.

Secondly, we thoroughly tested the comprehension tests, but we personally feel the measurement of comprehension is a vague idea at times. We tried to include both insight and memory-oriented questions in the tests to ensure comprehension and/or performance would be measured adequately, but it is something that remains up for discussion. We do have to mention that most participants did think the tests were comparable to what they would expect in an examination of such a course.

Thirdly, as also mentioned by the participants themselves, the use of eye trackers likely has an impact on the viewing behavior for some students. Evidently, this is something that is present in all eye-tracking research, but we still believe it should be mentioned nonetheless. Eye trackers can provide invaluable information on the processing and viewing patterns of students, though they still are an intrusive measure. We fully support their use in future research, but simply wish to underline that care should be taken when analyzing and interpreting the results of such research.

Lastly, we acknowledge that this thesis has not yet exhausted and/or mentioned all there is to learn from the data that was collected in the considerable number of studies that were conducted. The eye-tracking data are rich and can definitely be explored further, focusing on different, specific elements. One example of this would be pupillometry or micro-saccades and how these correlate with perceived cognitive load or other indications of processing. Additionally, the study discussed in Chapter 6 also included a qualitative component in the form of a survey. A preliminary analysis of the results from this survey has been discussed at a conference (Van Hoecke, 2022), but has not yet been published. Further and more thorough analysis would first be required before it can be cast in a publishable format.

3 Conclusion and future research

This research has contributed to a better understanding of the effects of subtitles and lecture styles in asynchronous, online education. It has shown that numerous factors play a role in the processing of subtitled online lectures, such as the proficiency of students, the presence and type of other on-screen content and the interplay between the language of the audio and the language of the on-screen content. It has further unveiled the intricacy of the matter, adds knowledge and highlights nuances to the complex puzzle that is processing of multimodal education. Despite the considerable contribution, it was also revealed that much is still left to be examined and learned about the topic. Some possible future directions are discussed below.

The first future track we elucidate builds on our assumption that different languages in a single channel can cancel the benefit gained from the modality principle. It would be interesting to investigate this concept to either refute the hypothesis or confirm it in multiple settings so that it can be consolidated in theory. If our hypothesis is true, we would expect that, for example, in an English-taught lecture, learners can benefit from subtitles in their L1 or another proficient language as long as the other on-screen content is in that same language. This is a question that remains to be answered. Furthermore, we also wonder how this assumption of multiple languages in a single processing channel holds when viewers with lower proficiencies in one of the languages watch these lectures. This highlights a clear path for future research.

The second track is based on our findings of viewing behavior at home. Students may speed up lectures, walk around, take notes, use their phones, etc. In such cases, processing is expected to change considerably. It would therefore be interesting to look at how processing and reading happens in sped-up lectures or in a home context where students are allowed to behave normally. This is evidently not a simple task as it would trade a significant amount of control that the researcher has over the experiment for a tremendous amount of ecological validity.

The third and last track of research we would like to highlight is that of presentation slide reading. During the project, we have discovered a significant lack of studies that explore the reading and processing of multimedia slides, especially in education. These could, however, provide valuable insight into optimal design of these slides. Furthermore, we also wonder how viewing and reading behavior would change depending on the content that is present on the slides, e.g., full sentences vs. keywords or multiple images and little text vs. only text. The impact of slide density, slide content and slide reading on processing is a topic that is currently still underexplored.

4 References

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Vol. 2, pp. 206–226). Cambridge, England: Cambridge University Press.

Bartsch, R. A., & Cobern, K. M. (2003). Effectiveness of PowerPoint presentations in lectures. *Computers & Education, 41*, 77–86. doi:10.1016/S0360-1315(03)00027-7

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics*, *23*(4), 509–533. doi:10.1017/S0142716402004022

Chan, W. S. (2020). *An investigation of subtitles as learning support in university education* [Unpublished doctoral dissertation]. Macquarie University, Sydney, Australia.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94*(2), 428–434. doi:10.1037//0022-0663.94.2.428

d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist, 12*(3), 196–205. doi:10.1027/1016-9040.12.3.196

Danan, M. (2015). Subtitling as a language learning tool: Past findings, current applications, and future paths. In Y. Gambier, A. Caimi, & C. Mariotti (Eds.), *Subtitles and language learning: Principles, strategies and practical experiences* (pp. 41–63). Bern, Switzerland: Peter Lang.

Diao, Y., Chandler, P., & Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology, 120*(2), 237–261.

Dommett, E. J., Dinu, L. M., Van Tilburg, W., Keightley, S., & Gardner, B. (2022). Effects of captions, transcripts and reminders on learning and perceptions of lecture capture. *International Journal of Educational Technology in Higher Education, 19*.

Fiorella, L., Stull, A. T., Kuhlmann, S., & Mayer, R. E. (2019). Instructor presence in video lectures: The role of dynamic drawings, eye contact, and instructor visibility. *Journal of Educational Psychology, 111*(7), 1162–1171. doi:10.1037/edu0000325

Ghia, E. (2012). The impact of translation strategies on subtitle reading. In E. Perego (Ed.), *Eye tracking in audiovisual translation* (pp. 157–182). Rome, Italy: Aracne.

Hayati, A., & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology, 42*(1), 181–192. doi:10.1111/j.1467-8535.2009.01004.x

Hefer, E. (2013a). Reading first and second language subtitles: Sesotho viewers reading in Sesotho and English. *Southern African Linguistics and Applied Language Studies, 31*(3), 359–373. doi:10.2989/16073614.2013.837610

Hefer, E. (2013b). Reading second language subtitles: A case study of Afrikaans viewers reading in Afrikaans and English. *Perspectives, 21*(1), 22–41. doi:10.1080/0907676x.2012.722652

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Vol. 2, pp. 247–262). Cambridge, England: Cambridge University Press.

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology, 107*(3), 724–739. doi:http://dx.doi.org/10.1037/edu0000013

Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). *Showing face in video instruction: Effects on information retention, visual attention, and affect.* Paper

presented at the SIGCHI Conference on Human Factors in Computing Systems, Toronto, Canada.
Kruger, J.-L., Hefer, E., & Matthew, G. (2013). *Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts.* Paper presented at the 2013 Conference on Eye Tracking South Africa, Cape Town, South Africa.
Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research, 7*(5), 1–15. doi:10.16910/jemr.7.5.4
Kruger, J.-L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly, 49*(1), 105–120. doi:10.1002/rrq.59
Kruger, J.-L., Wisniewska, N., & Liao, S. (2022). Why subtitle speed matters: Evidence from word skipping and rereading. *Applied Psycholinguistics, 43*, 211–236.
Lavaur, J.-M., & Bairstow, D. (2011). Languages on the screen: Is film comprehension related to the viewers' fluency level and to the language in the subtitles? *International Journal of Psychology, 46*(6), 455–462. doi:10.1080/00207594.2011.565343
Leppink, J., & van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education, 4*(3), 119–127. doi:10.1007/s40037-015-0192-x
Liao, S., Kruger, J.-L., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of Specialised Translation*(33), 70–98.
Liao, S., Yu, L., Kruger, J.-L., & Reichle, E. D. (2022). The impact of audio on the reading of intralingual versus interlingual subtitles: Evidence from eye movements. *Applied Psycholinguistics, 43*(1), 237–269. doi:10.1017/S0142716421000527
Liao, S., Yu, L., Reichle, E. D., & Kruger, J.-L. (2021). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading, 25*(5), 417–435.
Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Vol. 2, pp. 227–246). Cambridge, England: Cambridge University Press.
Mayer, R. E. (2011). *Applying the Science of Learning*. London, England: Pearson.
Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 279–315). New York, NY: Cambridge University Press.
Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187–198. doi:10.1037//0022-0663.93.1.18
Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, 1–25. doi:10.1017/s0272263119000706
Montero Perez, M., Peters, E., & Desmet, P. (2013). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL, 26*(1), 21–43. doi:10.1017/s0958344013000256
Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal, 99*(2), 308–328. doi:10.1111/modl.12215
Nouri, H., & Shahid, A. (2005). The effect of PowerPoint presentations on student learning and attitudes. *Global Perspectives on Accounting Education, 2*, 53–73.
Paivio, A. (1986). *Mental Representations: A dual coding approach*. Oxford, England: Oxford University Press.

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology, 13*(3), 243–272. doi:10.1080/15213269.2010.502873

Stull, A. T., Fiorella, L., & Mayer, R. E. (2021). The case of embodied instruction: The instructor as a source of attentional and social cues in video lectures. *Journal of Educational Psychology, 113*(7), 1441–1453. doi:10.1037/edu0000650

Szarkowska, A., & Bogucka, L. (2019). Six-second rule revisited: An eye-tracking study on the impact of speech rate and language proficiency on subtitle reading. *Translation, Cognition & Behavior, 2*(1), 101–124. doi:10.1075/tcb.00022.sza

Szarkowska, A., & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: Evidence from eye movements. *PLoS One*, *13*(6).

Szarkowska, A., Silva, B., & Orrego-Carmona, D. (2021). Effects of subtitle speed on proportional reading time: Re-analysing subtitle reading data with mixed effects models. *Translation, Cognition & Behavior*, *4*(2), 305–330.

Taylor, G. (2008). Perceived processing strategies of students watching captioned video. *Foreign Language Annals*, *38*(3), 422–427. doi:10.1111/j.1944-9720.2005.tb02228.x

van der Zee, T., Admiraal, W., Paas, F., Saab, N., & Giesbers, B. (2017). Effects of subtitles, complexity, and language proficiency on learning from education videos. *Journal of Media Psychology*, *29*(1), 18–30.

van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 263–278). New York, NY: Cambridge University Press.

Van Hoecke, S. M. (2022, December). *Reception of subtitled online lectures: What does the learner want?* Paper presented at the 2022 Conference of the Australian Linguistic Society (ALS), Melbourne

Vulchanova, M., Aurstad, L. M. G., Kvitnes, I. E. N., & Eshuis, H. (2015). As naturalistic as it gets: Subtitles in the English classroom in Norway. *Frontiers in Psychology, 5*(1510). doi:10.3389/fpsyg.2014.01510

Part VI: Appendix

B

1 Lecture texts

Thomas Piketty on (In)equality

In 2013, French economist Thomas Piketty published his magnum opus, entitled Capital in the Twenty-First Century. This book sent shockwaves through the field of economics. It left a deep impression upon academics, politicians, journalists, but even regular mortals had a crack at it. The book has been translated into more than thirty languages in three years' time. It made Piketty become an academic rock star.

At first sight, the success of Capital in the Twenty-First Century seems rather puzzling, because this tome of 700 pages, littered with graphs and charts and tables, is written in a rather dry style, unwieldy and sometimes redundant. How could it become such a bestseller?

The answer is quite simple: Piketty's book sharply criticizes what is felt by many as a threat and a curse in an era of globalization, namely growing inequality. The first line of the book is a firm statement: "The distribution of wealth is one of today's most widely discussed and controversial issues." This observation leads to a fundamental question: does private capital accumulation lead to the concentration of wealth in ever fewer hands, as Karl Marx thought in the nineteenth century? Or do the balancing forces of growth, democracy and technology lead to reduced inequality and greater harmony among the classes?

To answer this question, Piketty and his team draw on a most impressive amount of data about the US and different European countries. Out of this research a clear pattern emerges. In the period 1870-1914 inequality in Europe was extremely high and stable, whereas the twentieth century (for the most part) saw a narrowing of inequality. The main reasons for this decrease were the introduction of a progressive income tax in Great Britain (1909), the United States (1913), France (1914) and India (1922), but also the economic and financial devastation of the Great Depression and two World Wars. Taken together, these elements had a levelling effect upon conditions and relations of inequality, especially in Europe.

However, the rise to power of Ronald Reagan in the US and Margaret Thatcher in the UK marked the end of this process and the beginning of a return to rather nineteenth-century conditions. Both Reagan and Thatcher favored lowering tax rates for the rich. They curtailed trade unions and encouraged deregulation. They believed firmly in the blessings of capitalism's darling myth, which is also known as the free market.

Thatcher, Reagan and their kindred spirits were strengthened in their neoliberal faith by the influential work of US economist Simon Kuznets. Kuznets argued that "a free market economy will deliver both prosperity and equity, and not (as Karl Marx predicted) the accumulation of wealth at one pole and the accumulation of misery at the other." Kuznets believed that growth is a rising tide that lifts all boats, as John F. Kennedy once put it. But his optimism was based on scant evidence. Piketty proves it wrong, since his analysis shows that the temporary reduction of inequality between, say 1914 and 1980 (when Reagan was elected president of the US), was merely an intermezzo within a capitalist system that is naturally inclined towards ever greater inequality.

Inequality increases when "the rate of return on capital significantly exceeds the growth rate of the economy." In Piketty's formula: $r > g$. If $r > g$, then concentrations of wealth will grow faster than the economy does. This is exactly what is going on today in most parts of the world and it may throw us back into nineteenth-century conditions.

The distorted ratio r > g poses social, political and economic danger, since it allows fortunes to "grow and perpetuate themselves beyond all reasonable limits and beyond any possible rational justification in terms of social utility." This "lasting concentration of capital" threatens democratic systems and structures and institutions, especially in the US but also in Europe and elsewhere in the capitalist world. "The risk of a drift toward oligarchy is real," Piketty warns, "and gives little reason for optimism about where the United States is headed." Now that the White House has been captured by what looks increasingly like a conspiracy of billionaires, these words sound more ominous and credible than before.

According to Piketty, the tendency towards ever greater inequality will not slow down spontaneously. The only way to avert total chaos, massive social unrest or even a third world war is to introduce a progressive, global wealth tax. Piketty knows, however, that such a tax is very unlikely to be implemented any time soon. He therefore calls it a "utopian" plan. Per some opponents, such a tax would be disastrous for the world economy.

Whatever its practical difficulties may be, Piketty's solution is driven by a lofty ideal. As he puts it: "If democracy is someday to regain control of capitalism, it must start by recognizing that the concrete institutions in which democracy and capitalism are embodied need to be reinvented again and again."

Jean-Jacques Rousseau on (In)equality

In 1753, the Geneva-born French philosopher Jean-Jacques Rousseau joined an intellectual competition organized by the Academy of Dijon. The Academy invited contributions in answer to the following question: "What is the origin of inequality among mankind and does natural law decree inequality?" Rousseau submitted a long essay, entitled Discourse on the Origins and Foundation of Inequality among Mankind. It failed to win the competition, but it became one of the key-texts in western philosophy. So much for the Academy.

The Second Discourse, as his essay is commonly referred to, advances a radical hypothesis: people in the eighteenth century are miserable, society is corrupted and decadent – and the root of it all is inequality. How is it, Rousseau wonders, that so many people accept or at least condone "outrageous social and economic disparities? How did we arrive at our present condition?"

Rousseau makes it very clear he is not interested in the historical development of man. His treatise reads like an intellectual fancy, a what if narrative. "Let us begin," he noted famously, "by laying aside facts, for they do not affect the question." Instead, Rousseau posits as the origin of human culture a hypothetical state of nature. This state of nature preceded all the potential blessings and horrors of socialization.

What would men have become, Rousseau wonders, if they had been "left to themselves"? Imagine man stripped "of all the supernatural gifts which he may have received, and of all the artificial faculties, which he could not have acquired but by slow degrees." Let us envision man "such as he must have issued from the hands of nature."

In that hypothetical state of nature, savage man lead a solitary life, driven only by the need for self-preservation. He was speechless and innocent. He was happy without knowing it. He lived in a state of blissful ignorance, his existence unaffected by the pangs of inequality.

Rousseau distinguishes between two kinds of inequality. Natural inequality is constituted by "difference of age, health, bodily strength, and the qualities of the mind (…)." Moral, or political inequality “consists in the different privileges, which some men enjoy, to the prejudice of others, such as that of being richer, more honored, more powerful, and even that of exacting obedience from them.” Natural equality is relatively benign. It must never be enlisted to explain or justify the harmful facts of political inequality. In the state of nature inequality is hardly felt and it does not make people unhappy. It is political inequality that makes man dependent, miserable and weak. So how did political inequality come about, and why did we let it happen?

Infant man’s only concern was self-preservation. After a while, however, the solitary savage discovered that in some cases he had a “common interest” with others and that it might be advantageous to cooperate… This is a crucial moment in Rousseau’s speculative (re)construction of the process of civilization. It inaugurates the tragedy of the human condition as Rousseau saw it. What makes our condition tragic is the fact that our misery is due to the very same conditions that allow us to better satisfy our needs and thus to enhance our happiness.

Cooperation and socialization bring out the best in us: “the habit of living together gave birth to the sweetest sentiments the human species is acquainted with, conjugal and paternal love.” However, this very same process also renders us dependent upon one another. Mutual dependency requires manners, rules, laws. It ensnares and tames the individual who will gradually surrender his autonomy to patterns and relations of subservience, power, hierarchy. Rousseau calls this “the first yoke” which man “inadvertently imposed upon himself”. The process that enables us to better satisfy our needs and desires – cooperation and socialization – and thus to make us happier, also brings about the outrageous political and social inequality that makes us so utterly miserable. This is the paradox, indeed the tragedy of the condition humaine as Rousseau saw it: everything we do to enhance our happiness also enhances our wretchedness.

Division of labor and cooperation begot yet another monster: private property. “The first man,” Rousseau wrote famously in the Second Part of the Second Discourse, “who after enclosing a piece of ground, took it into his head to say, this is mine, and found people simple enough to believe him, was the real founder of civil society.” This civil society is characterized by inequality, social division and a growing antagonism between the rich and the poor. It degenerates into exploitation, violence and chaos. People find themselves living in a state of war of all against all. Atomized, solitary and threatened by each other, people will demand a strong ruler. They will be prepared to sacrifice what is left of their freedom in exchange for order, peace, security. Eventually, this state of total disarray will usher in despotism, “the last term of inequality”. Under despotism, Rousseau concludes rather ironically, “all private men return to their primitive equality, because they are nothing.”

Alexis de Tocqueville on (In)equality

In 1831, the 26-year-old French aristocrat Alexis de Tocqueville went to America. The purpose of his trip was to study the American prison system. Upon his return, Tocqueville, did write a report on the penitentiary system in the United States. However, while travelling across the USA, something else had sparked his intellectual curiosity: American democracy.

Democracy confounded him. How is it, Tocqueville asked himself, that the French revolution led to the Terror, while the American revolution resulted in liberal democracy? So, he wrote another book, Democracy in America, which attempted to answer that question. But the book grew into something much vaster and more profound. Published in two volumes in 1835 and 1840 respectively, Democracy in America was instantly hailed as a masterpiece. The American journalist and politician Horace Greeley, for instance, called it “by far the most important book that has been written on the nature and influence of Democracy." Today it is still considered a classic of political philosophy.

In America, Tocqueville was struck by what he called "the general equality of condition among the people." Equality of condition is "the fundamental fact" of American society. Equality is also fundamental to democracy, the 'new' political system Tocqueville saw emerging all around him. He saw the rise of equality/democracy as a revolution taking place "throughout the Christian world". He considered this process inevitable and irreversible. Any effort to block or undo it would appear "as a fight against God Himself."

In studying equality and its influence upon American society, Tocqueville tried to understand what a democratic society looked like. And since he considered American democracy a precursor of what was to happen in Europe as well, studying American democracy became a way of envisaging Europe's future. In a famous phrase: "I admit that I saw in America more than America; it was the shape of democracy itself which I sought, its inclinations, character, prejudices, and passions; I wanted to understand it so as at least to know what we have to fear or hope therefrom."

What was there "to fear or hope" from democracy? As an aristocrat, Tocqueville greeted the rise of equality and advent of democracy with both terror and awe. Democracy in America examines "what blessings and what ills" this new system produces. It offers a balanced account of gains and losses.

In the First Volume Tocqueville studies American democracy as a series of political, structural and institutional facts. However, democracy is not merely a political reality or ideal. To be a citizen in a democracy is not just a matter of, say, voting rights and fair tax rules. Democracy is also a psychological reality. It creates and requires a particular mindset. To be a democratic citizen is also an attitude, an emotional reality. Equality is also a "passion".

In the Second Volume Tocqueville focuses more upon this psychological dimension of democracy. What does democracy do to you as a human being, Tocqueville wonders. How does equality of condition affect your thoughts, your ambitions, your emotions? And the biggest question of all: will people in a democracy be happier than in an aristocratic society?

Tocqueville saw the advantages as well as the disadvantages of both systems. Here is a key difference: aristocratic societies are much less equal and more static than democracies. In aristocratic nations, "families remain for centuries in the same condition, often on the same spot." All citizens "occupy fixed positions, one above another." Individuals are tied to one another by all sorts of interdependencies: patronage, family, subservience, allegiance. Born a butcher, always a butcher. Born a duke, always a duke. Society was stratified, changes were few and slow.

Democracy is a very different system, because family ties, tradition, place, patronage, … all these factors become less and less decisive in deciding the individual's fate and future. However, this is not to say that there are no social-economic classes in a democracy. Of course, there are, but a democratic society is more flexible, more fluid than the old aristocracy. The idea of social mobility, which made no sense to Tocqueville's forefathers, gathers momentum as equality/democracy spreads. Typically, Tocqueville regards this as a mixed blessing. The same equality which makes the citizen of a democracy independent of other citizens, he writes, "leaves him isolated and defenseless in the face of the majority." To be socially mobile means that you can rise above your class, but also sink below it. A term Tocqueville uses frequently to describe this condition is 'flux'. A democracy is a society in flux. "Nobody's position is quite stable."

F

After 700 pages, Tocqueville was still unable or unwilling to embrace democracy wholeheartedly. It was up to the nations of his day, he wrote, to decide "whether equality is to lead to servitude or freedom, knowledge or barbarism, prosperity or wretchedness."

2 Biographical survey

Biographical survey – Belgian Study (Chapter 5)

1. Wat is je geslacht?

o Man
o Vrouw
o Andere

2. In welk jaar ben je geboren?

[___________________]

3. Wat is je moedertaal?

o Nederlands
o Frans
o Engels
o Andere, namelijk... [___________________]

4. Wat is je hoogst behaalde of huidige opleidingsniveau?

o Middelbaar onderwijs
o Eerste Bachelor
o Tweede Bachelor
o Derde Bachelor
o Master
o Andere, namelijk… [___________________]

5. Geniet je van bijzondere faciliteiten aan deze universiteit (door bv. topsportstatuut, ADHD, dyslexie, etc.)? Zo ja, waarom?

o Nee
o Ja, door…[___________________]

---

6. Wat was de onderwijstaal in jouw lager onderwijs?

o Nederlands
o Frans
o Engels
o Andere, namelijk… [___________________]

7. Wat was de onderwijstaal in jouw middelbaar onderwijs?

o Nederlands
o Frans
o Engels
o Andere, namelijk... [___________________]

---

8. Ik beschrijf mijn LUISTERVAARDIGHEID in het ENGELS als volgt:
o I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.
o I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.
o I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radoi or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.
o I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect.
o I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort.
o I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent.

9. Ik beschrijf mijn LEESVAARDIGHEID in het ENGELS als volgt:
o I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.
o I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.
o I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.
o I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.
o I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.
o I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.

10. Gebruik je regelmatig Engels in een huiselijke context? Zo ja, hoeveel uren gemiddeld per week?

o Ja, namelijk … uren/week [___________________]
o Nee

11. Gebruik je regelmatig Engels in een communicatieve context (bv. praten onder vrienden, chatten)? Zo ja, hoeveel uren gemiddeld per week?

o Ja, namelijk … uren/week [___________________]
o Nee

12. Gebruik je regelmatig Engels voor je werk/studentenjob? Zo ja, hoeveel uren gemiddeld per week?

o Ja, namelijk … uren/week [__________________]

o Nee

13. Gebruik je regelmatig Engels in een nog niet genoemde context (bv. online gaming, series kijken, etc.)? Zo ja, hoeveel uren gemiddeld per week? (In geval van opsommingen, telkens wat en aantal uren bij elkaar)

o Ja, namelijk … en dit … uren/week [__________________]

o Nee

---

14. Heb je ooit langdurig (meer dan twee maanden) in een Engelstalig land verbleven? Zo ja, hoeveel maanden?

o Ja, namelijk … maanden [__________________]

o Nee

---

15. Als ik een Engelstalig programma bekijk, kijk ik waar mogelijk

o zonder ondertiteling

o met ondertiteling in het Engels

o met ondertiteling in het Nederlands

o met ondertiteling in een andere taal, namelijk... [__________________]

---

16. Ik beschrijf mijn LEESVAARDIGHEID in het NEDERLANDS als volgt:

o Ik kan vertrouwde namen, woorden en zeer eenvoudige zinnen begrijpen, bijvoorbeeld in mededelingen, op posters en in catalogi.

o Ik kan zeer korte eenvoudige teksten lezen. Ik kan specifieke voorspelbare informatie vinden in eenvoudige, alledaagse teksten zoals advertenties, folders, menu's en dienstregelingen en ik kan korte, eenvoudige, persoonlijke brieven begrijpen.

o Ik kan teksten begrijpen die hoofdzakelijk bestaan uit hoogfrequente, alledaagse of aan mijn werk gerelateerde taal. Ik kan de beschrijving van gebeurtenissen, gevoelens en wensen in persoonlijke brieven begrijpen.

o Ik kan artikelen en verslagen lezen die betrekking hebben op eigentijdse problemen, waarbij de schrijvers een bepaalde houding of standpunt innemen. Ik kan eigentijds literair proza begrijpen.

o Ik kan lange en complexe feitelijke en literaire teksten begrijpen, en het gebruik van verschillende stijlen waarderen. Ik kan gespecialiseerde artikelen en lange technische instructies begrijpen, zelfs wanneer deze geen betrekking hebben op mijn terrein.

o Ik kan moeiteloos vrijwel alle vormen van de geschreven taal lezen, inclusief abstracte, structureel of linguïstisch complexe teksten, zoals handleidingen, specialistische artikelen en literaire werken.

---

17. Studeer je of heb je Engels gestudeerd in het hoger onderwijs?

o Nee

o Ja, maar nu niet meer

o Ja, ik heb enkele Engelse vakken opgenomen (bv. als keuzevak)

o Ja, Engels is een van mijn hoofdtalen

18. Had je in het verleden al of heb je dit jaar een vak over filosofie op de hogeschool of universiteit?

o Ja, namelijk… [__________________]

o Nee

Biographical survey – Australian Study (Chapter 6)

1. What is your gender?

o Male

o Female

o Other

2. What year were you born in?

[__________________]

3. What is your native language?

o English

o Other, namely... [__________________]

4. What is your highest obtained degree?

o High school degree

o Bachelor/Undergraduate degree

o Graduate diploma/Honours degree

o Masters degree

o Other, namely… [__________________]

5. What was your language of instruction in primary school?

o English

o Other, namely… [__________________]

6. What was your language of instruction in high school?

o English

o Other, namely... [__________________]

7. I would describe my listening skills as follows:

o I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.

o I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.

o I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radoi or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.

o I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect.

o I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort.

o I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent.

8. I would describe my reading skills as follows:*

o I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.

o I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.

o I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.

o I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.

o I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.

o I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.

---

9. Are you interested in philosophy?

o Yes

o No

10. Are you interested in the concept of (in)equality?

o Yes

o No

---

11. Have you ever followed a course on philosophy before?

o Yes, at university or college

o Yes, in high school

o No

12. Do you know anything about Thomas Piketty?

- o Yes, I have read his books/books about him.
- o Yes, I have heard of him and have some knowledge of his work/life.
- o Yes, I have heard of him, but do not know a lot about him
- o No

13. Do you know anything about Jean-Jacques Rousseau?

- o Yes, I have read his books/books about him.
- o Yes, I have heard of him and have some knowledge of his work/life.
- o Yes, I have heard of him, but do not know a lot about him
- o No

14. Do you know anything about Alexis de Tocqueville?

- o Yes, I have read his books/books about him.
- o Yes, I have heard of him and have some knowledge of his work/life.
- o Yes, I have heard of him, but do not know a lot about him
- o No

L